FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Master Thesis

Mohamed Bayoumi Fahmy Hawash – TQS650

Population genetic and phylogenetic studies on *Trichuris* spp. recovered from pigs, humans and baboons in different geographical regions

Supervisor: Peter Nejsum

Submitted: 31/08/2014

Section of Parasitology and Aquatic diseases
Department of Veterinary Disease Biology
Copenhagen University
Dyrlægevej 100, 1871 Frederiksberg C,
Denmark

Mohamed Bayoumi Fahmy Hawash

Cover Photo from: www.tvblogs.nationalgeographic.com/blog/big-baboon-house/

Contents

Summary

Trichuris suis and *T. trichiura* are two different whipworms that infect pigs and humans respectively. *Trichuris suis* is found in pigs in almost all over the world while *T. trichiura* is found in humans mainly in areas with poor sanitation and is responsible for nearly 460 million infections. The taxonomic status of whipworms in non-human primates is unsettled but they are normally designated as *T. trichiura* as well. The pattern and nature of transmission of *Trichuris* spp. infecting human, non human primates and pigs is poorly understood. Population genetics can provide inferences about the direction of parasite transmission on different levels (between individuals, populations and geographical locations) beside the natural and demographic history of the parasite. In the first study, I applied different population genetic and phylogenetic approaches to investigate the pattern of transmission and the demographic history of *Truchuris* spp. in pigs, humans and baboons. I tested a hypothesis regarding the transmission and history of *Trichuris* spp. between humans and pigs which is the *Trichuris* sp. was found in either humans or pigs before it was transmitted to the other host in the Neolithic era upon domestication of animals when the two hosts were in sympatry nearly 10,000 years ago. Moreover I investigated the evolutionary and phylogenetic relationship of the *Trichuris* infecting humans and a non human primate (baboon). Coalescent analysis, Analysis of Molecular Variance (AMOVA) and phylogenetic analysis were applied on two mitochondrial markers namely the *nad*1 and *rrn*L genes on *T. suis* populations from domestic pigs from Denmark, USA, Uganda and China; *T. trichiura* from humans from China and Uganda; *Trichuris* from olive and yellow baboons from Denmark (Copenhagen Zoo and Knuttenborg Park) and from USA (Southwest National Primate Research Center (SNPRC), Texas). The demographic history reconstructed through coalescent analysis of the different populations of *Trichuris* spp. from pigs and humans revealed independent demographic history of the parasite in the two hosts indicating that the speciation events happened much older than the Neolithic era 10,000 years ago and hence the hypothesis was rejected. Moreover, the demographic history of *T. suis* in different areas underlined the major role of the human trading activity in transmitting the parasite through pig transport. No genetic differentiation between *Trichuris* obtained from baboons and humans in Uganda emphasizing on the possible zoonotic potential of *Trichuris* from baboons. In the second study, the aim was to sequence full mitochondrial genome of *Trichuris* recovered from baboons and pigs and compare with the other available mitochondrial genomes of *Trichuris* from human, non human and pig hosts. I investigated the whole mitochondrial genome of two different haplotypes of *Trichuris* recovered from olive baboons (*Papio anubis*) and conducted phylogenetic analysis and compared the genetic and protein

sequences distances of the mitochondrial genes with other *Trichuris* derived from human and a non human primate (François' leaf-monkey) to investigate the evolutionary and genetic relatedness between different *Trichuris* spp. in these primates. Also, phylogenetic and comparative analysis were conducted by comparing the full mitochondrial genomes of *T. suis* from pigs from Uganda and Denmark with the avialable *T. suis* mitochondrial genome from China. The genetic distance between the *Trichuris* spp. recovered from the two different non human primates (olive baboon and François' leaf-monkey) was very high and suggests the presence of different *Trichuris* spp. infecting non human primates. Remarkably, I found that the genetic distance between the two *Trichuris* worms derived from baboons were also genetically very distinct which may hint that there could be different *Trichuris* spp. infecting olive baboons. However, one of the baboon *Trichuris* was genetically related to the human *Trichuris* which might be the same species. Likewise, I found high divergence in protein and genetic sequences between the *T. suis* from pigs from Denmark to that from Uganda and China suggesting also possible presence of different species of *Trichuris* infecting domestic pigs in these areas.

Preface

This thesis's studies were carried out from autumn 2013 until summer of 2014 mainly in the laboratories of Section for Parasitology and Aquatic Diseases, Department of Veterinary Disease Biology. This thesis contains general background followed by brief introduction to the studies with the major finding and conclusions while the detailed studies are represented in manuscripts 1 and 2.

My favorite quote

"The only thing that interferes with my learning is my education"

Albert Einstein

Acknowledgment

I would like to thank my main mentor Peter Nejsum for supervising this work. Peter has many things to be an excellent supervisor. Besides his wide knowledge and excellence in the study field, he is very encouraging and committed to the work. The always smile on his face did make the peak hour of the job go easily and less stressful.

I also would like to thank Lee O´Brien Andersen in the Staten Serum Institute (SSI) for helping in the genome assembly and giving very valuable explanation for the Next Generation Sequencing methodologies. The thanks are also extended to my office mates for making working hours much fun and the general nice atmosphere in the office.

Not to forget to mention Professor Søren Rosendahl in the Department of Biology for providing many valuable advices and helping in the running of some the populations genetic analyses used in this thesis.

Special thank to my mother, the one who is always listening and encouraging along the road to the end. Without her magical words that lessen my stress and enlighten my road, I would not have been able to finish this work.

Background

The background is divided mainly into two parts. The first part is a parasitology background where I will present different aspects regarding the parasite biology. The second part is about the molecular biology where I will first give a background on the population genetic and phylogenetic approaches and their importance in the study of the parasitology besides the main principles undelying some of the population genetic and phylogenetic analyses used in the studies. Then a brief introduction to the parasitic genomes and next generation sequencing technologies will be discussed.

Parasitology

Phylogeny

Genus *Trichuris* belongs to class Enoplea, order Trichocephalida and family Trichuridae. Trichuridae means hair-tail (*Trichos* = hair and *oura* = tail). This nomenclature originated from the misunderstanding of the worm morphology. The worm general morphology consists of anterior thin part and thick posterior part resembling a whip (also called whipworms). Therefore, it has been suggested to use the name *Trichocephalus* instead which means hair-head (Trichos = hair and kephale = head) (Roberts and Schemit, 2009).

Trichuris spp. and their hosts

Several species belong to the genus *Trichuris* infect different mammalian hosts. For instance, *T. trichiura* parasitize human and also the *Trichuris* in non human primates is designated as *T. trichiura* (Abee et al, 2012; Cutillas et al, 2009), *T. suis* infects pigs (Liu et al, 2012a; Cutillas et al, 2009), *T. ovis* and *T. discolor* infect sheep and other ruminants (Liu et al, 2012b), *T. vulpis* infects dogs (Cutillas et al, 2007) and *T. muris* infects wild and laboratory mice (Behnke and Wakelin, 1973)

Morphology

The common name of the whipworm hints to its general morphology (Figure 1, B). The anterior part is long slender and it comprises about 2/3 of worm length. The oesophagus in the anterior part is surrounded by glandular-like cells called stichocytes and the whole structure is called stichosome. The posterior end of the body is thick and shorter. There are obvious morphological differences between adult male and female *Trichuris*. Typically, the length of adult female (30-50 mm) is longer than male (20-35mm). Adult male has a curled posterior end with a spicule while adult female has straight posterior end. Fertilized eggs from the female are released through the vulva located near the

connection between oesophagus and intestine. The egg has a characteristic lemon shape with two polar plugs (Bundy and Cooper, 1989; Roberts and Schmidt, 2009).

Life Cycle

The parasite has a direct life cycle, i.e. no intermediate hosts (Figure 1, A). The host becomes infected by ingesting embryonated eggs containing infective L1 larvae. The egg hatches in the distal part of the gut (ceacum and colon). L1 undergoes 3 different molts (L2, L3, and L4) before it molts to the adult L5 stage. At this stage, the anterior part of the worm is embedded in the gut mucosa feeding on blood and cell contents while the posterior part is set free in the gut lumen. The female starts producing eggs in a period that differes between different *Trichuris* species. For instance, *T. suis* starts producing eggs after 6 weeks post infection (p.i.) while *T. trichiura* produces eggs 16 weeks p.i. (Kringle and Roepstorff, 2006; Roberts and Schmidt, 2009). The female worm produces approximately 3000-20,000 eggs per day (Roberts and Schmidt, 2009). Embryonation of the eggs occurs in suitable environmental conditions in terms of oxygen, temperature and humidity (Burden and Hammet, 1979). Higher temperature will decrease the embryonation time as Beer (1976) found the embryonation time is 28 and 13 days in *T. trichiura* and 37 and 19 days in *T. suis* eggs at 25 °C and 34 °C respectively. *Trichuris* egg has a remarkable ability in withstanding severe environmental conditions. For example, *T. suis* eggs could remain viable up to 11 years and still be infective (Burden et al, 1987).

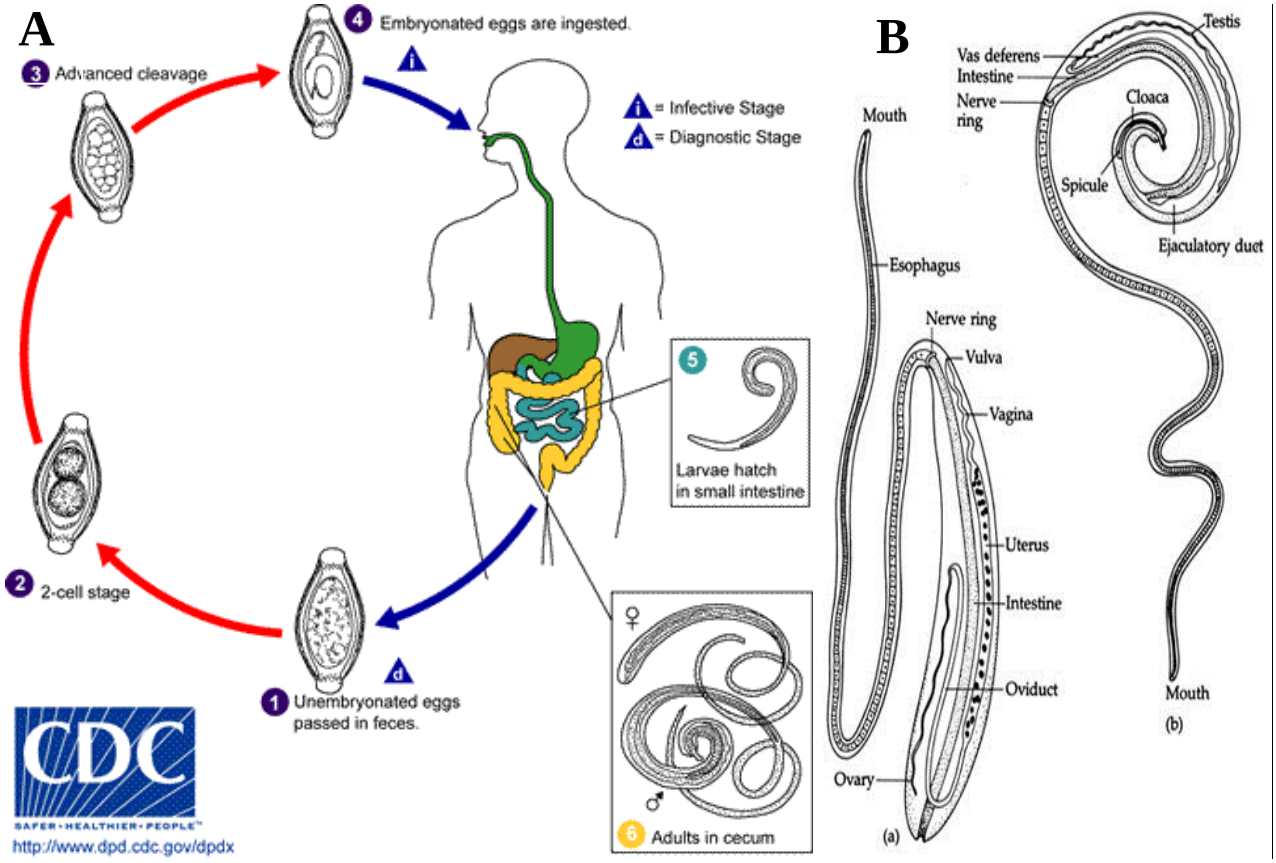

Figure 1. A. Life cycle of *Trichuris trichiura* (Source: http://www.cdc.gov/parasites/whipworm/) B. Adult worm of *T. trichiura* (a) Female (b) Male (Source: http://rowdy.msudenver.edu/~churchcy/BIO3270/Images/Nematodes/Trichuris_trichiura.jpg)

Pathology and clinical symptoms

The pathology due to *T. suis* infections in pigs depends on the worm load of the infection and concurrent bacterial infections and the clinical symptoms could be diarrhea, anorexia, anemia, emaciation and growth retardation (Batte et al, 1977; Pittman et al, 2010). Experimental infection of low worm doses (14,000 infective eggs/pig) did not produce severe clinical symptoms while higher doses (40,000 eggs/pig) initiated severe clinical symptoms of bloody diarrhea (Hale and Stewart, 1979). Moreover, high levels of experimental infections in pigs (25,000-100,000 infective eggs/pig) resulted in high mortality rate up to 66% (Beer and Lean, 1973; Batte et al, 1977). Acute morbidity and mortality were also reported in heavy infection in gilts (Pittman et al, 2010).

In humans, children are more prone to develop clinical symptoms of trichuriasis (Pullan et al, 2014). Heavy infections in children can lead to serious symptoms of *Trichuris* dysentery syndrome, chronic dysentery, anemia and rectal prolapse (Bethony et al, 2006; Roberts and Schemit, 2009). Moreover, chronic trichurisis infections were found to cause congnitive impairment in children (Sternberg et al, 1997).

Similar subclinical of *Trichuris* infections are also found in non human primates. However, in juveniles, intestinal symptoms have been described such as diarrhea and anorexia with few reported death cases in heavy infection (Abee, 2012).

Epidemiology of *T. suis* and *T. trichiura*

T. trichiura infections in humans are generally prevalent in impoverished places with poor sanitation (Bethony et al, 2006) mainly in tropical and subtropical areas where the temperature and humidity conditions are favorable for embryonation of the eggs as well. Figure 2 represents the worldwide distribution of *T. trichiura* with the highest prevalence was found in South and South East Asia and in Sub-Saharan Africa (Pullan et al, 2014).

Trichuris infections are also prevalent in non human primates especially in tropical and subtropical areas where it has been reported in many species ranging from New World species (e.g. howler monkeys and woolly monkeys), Old World species (e.g. rhesus monkeys, Japanese macaques and baboons) and apes (gibbons and chimpanzees) (Abee et al, 2012).

T. suis infections have been found in pigs in different geographical areas such as temperate (Roepstorff et al, 1998), subtropical (Solaymani-Mohammadi et al, 2003; Lai et al, 2011) and tropical regions (Tamboura et al, 2006; Carter et al, 2013). The prevalence dependes on pig age and production systems where highest prevalence is found in young animals and pigs reared in outdoor systems (Nansen and Roepstorff, 1999; Roepstorff et al, 2011).

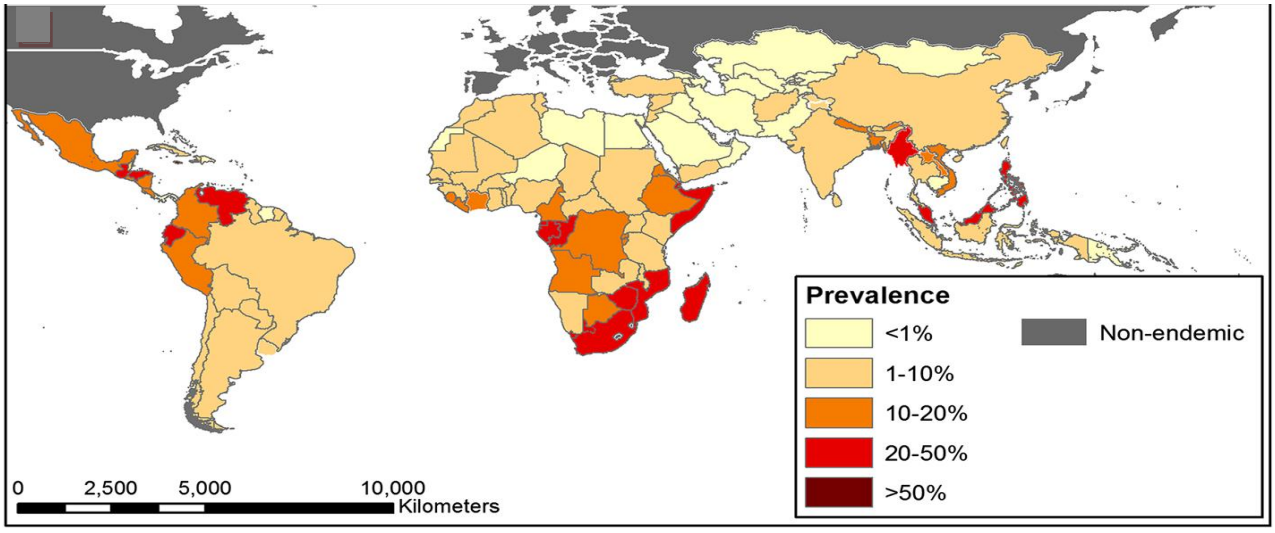

Figure 2: Worldwide distribution of *T. trichiura* based on 2010 surveys (Modified from (Pullan et al, 2014))

Economical and public health importance of *T. trichiura* and *T. suis*

T. trichiura is one of the soil transmitted helminths (STH) together with *Ascaris* and hookworms. STH are one of the neglected tropical diseases (NTD) that hamper prosperity in many developing countries in sub Saharan Africa, Latin America and Asia (Hotez and Kamath, 2009; Pullan et al, 2014). Recently, it has been estimated that the total disability adjusted life years (DALYs) for *T. trichiura* is 0.64 million years and a total of 464.6 millions infected people worldwide (Pullan et al, 2014).

Although *T. suis* rarely causes clinical symptoms in pigs, it may reduce the food intake, reduce growth of pigs and may lead to changed body composition (i.e. less meat) (reviewed by Roepstorff et al, 2011). Hence it may cause financial loss in pig production. Moreover, Nissen et al, (2012) identified *T. suis* worm recovered from humans in a sympatric area in Uganda which stress the possible zoonotic potential of *T. suis*.

Molecular Biology

Phylogenetics and Population Genetics.

Population genetics is the study of the allele frequency distribution and change influenced by the main four evolutionary forces (natural selection, genetic drift, mutation and gene flow) (Archie et al, 2008). In this context, population genetics enables us to gain information about the population history such as population size changes, migration and gene flow. In other words, exploring the genetic variation in the DNA provides insights into the evolutionary history and ecology of the parasites (De Meeus et al, 2006). The investigation of the population genetics of the infectious disease and their vectors has been termed as molecular epidemiology (De Meeus et al, 2006) and it has been applied to study a number of parasitic nematodes (Gilabert and Wasmuth, 2013). Although that phylogenetics is the branch of science that usually deals with the taxonomic positions of different species while population genetics deals with populations of specific species, there is fine line between the two branches in parasitology (Criscione et al, 2005). Since many parasitic nematodes have similar, in some cases identical, morphology, population genetics is usually used also to delimit different parasite species (Gilabert and Wasmuth, 2013).

Population genetics could provide inferences about 1) the direction of parasite transmission on different levels (between individuals, populations and geographical locations) 2) hybridization events 3) species and population delimitation (Archie et al, 2008; Gilabert and Wasmuth, 2013). Knowing the limits and boundaries of parasite populations is crucial to set the control measures on these populations (Gilabert and Wasmuth, 2013). For example, recently, population genetic approaches were applied using microsatellite and mitochondrial markers to investigate the pattern of transmission of *Ascaris suum* and *A. lumbricoides* between pigs and humans around the globe (Betson et al, 2014). It was found that *Ascaris* infections in humans in Europe are of pig origin which coincides with the potential of pig *Ascaris* to contribute to the zoonotic infections in developing areas (Zhou et al, 2012) hinting to the high zoonotic potential of pig *Ascaris* infections. However, in other sympatric areas the two parasite species were found to have separate life cycles with limited cross transmission and gene flow. Such information is important for the implementation of proper control measures in each region (Anderson, 2001).

Moreover, population genetic tools could be used to reconstruct the epidemiological history of the parasites that enables us to track the parasite movement, identify the parasite origin and to deduce the environmental factors responsible for its spread (Archie et al, 2008). One example is the reconstruction of the epidemiological history of *Wucheraria bancrofti* isolates from three different geographical regions in Africa and Asia (Ramesh et al, 2012). The study revealed that the transmission of the parasite between Africa and Asia underlies complex scenario and cannot be explained directly by the historical human migration out of Africa that other factors must contributed to the spread of the disease.

The inferences about the epidemiological history such as to date the parasite introduction and the inferences on the population size changes are based on the coalescence theory and Bayesian methods (Archie et al, 2008). Coalescence theory was first introduced by Kingman in the early eighties which is a way to look back in time using the existing samples (reviewed by Kingman, 2000). The main principle in coalescent theory is that for a selectively neutral DNA loci (i.e. no selection), samples will appear as they coalesce i.e. 'pick' their parents going back in time until all lineages coalesce in to single point which is the most recent common ancestor MRCA (Figure 3). The merging of two lineages (samples) is called coalescent event (Rosenberg and Nordberg, 2002).

Coalescent models have been applied to reconstruct the epidemiological history of many microparasites (viruses and bacteria) but fewer studies have been conducted on macroparasites

(helminthes) (Archie et al, 2008; Lymbery and Thompson, 2012). Coalescent using the Bayesian approaches is most often used to construct suitable models to infer the epidemiological history and events of the parasites such as those mentioned above (Archie et al, 2008).

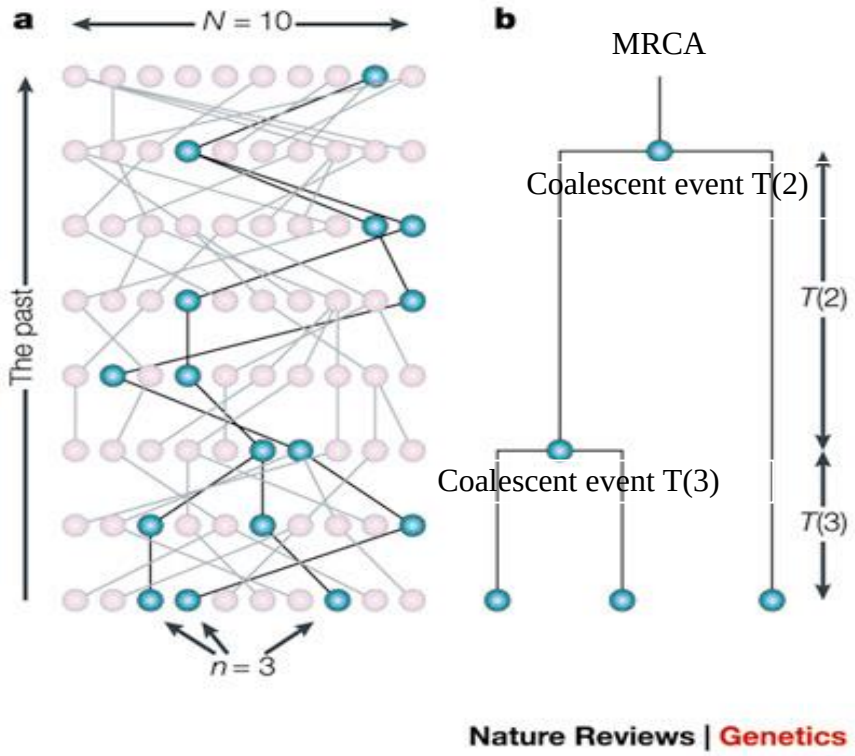

Figure 3. Diagrammatic illustration of the coalescent. A) A population of 10 haploid individuals. All individuals coalesce going back in time until they all merge in to single common ancestor. Bright blue balls represent sampled individuals (n=3) and faint balls represent non-sampled individuals. B) The genealogy of the three sampled individuals with two coalescent events (merging between two lineages) T(3) and T(2) (Modified from Rosenberg and Nordberg, 2002).

For a population genetics and phylogenetic studies, three components are needed: 1) DNA marker, 2) a molecular biological method and 3) computational models to analyze the DNA data (Sunnucks, 2000). Each component will be discussed below.

1. DNA marker

A suitable DNA marker is an informative segment in the parasite genome that shows sufficient variability to differentiate between different species (phylogeny) or different populations of the same species (population genetics). The degree of variability varies between different DNA markers and the most appropriate marker depends on the purpose of the study. If the purpose is to identify the parasite species, the DNA markers should have enough variability between different parasite species and minimum variability within the parasite species itself. On the other hand, if the purpose of the

study is to describe the population genetic structure of a parasite species, the marker should normally contain as much variation as possible (Gasser, 2006). DNA markers could be either nuclear or mitochondrial in origin.

Nuclear DNA markers

A number of targets in the ribosomal DNA complex, 18S-ITS1-5.8S-ITS2-28S, that is present in high copy number in the nuclear genome are commonly used as DNA markers. In this complex, the coding regions (18S, 5.8S, and 28S) are separated by non coding region internal transcribed spacer 1 and 2 (ITS-1 and ITS-2). ITS-1 and ITS-2 are widely used to differentiate between different parasite species 'inter-specific' and, less often, population variation 'intra-specific' (Gasser, 2006). Microsatellite markers are short tandemly repeated sequences of DNA (usually1-4 base pairs) characterized by their high polymorphism (different repeat number between different populations) which have been detected in the ITS regions in several parasitic nematodes (Conole et al, 2001; De Luca et al, 2004). Microsatellites are advantageous as they are abundant, highly polymorphic and could be amplified from small amount DNA template (De Meeus et al, 2007).

Mitochondrial DNA (mtDNA) markers

Mitochondrial (mt) markers are also called cytoplasmic markers and there are a number of advantages of mtDNA that explain its extensive use in population genetics and phylogenetics of parasites. First, the rate of substitution is higher in mtDNA than nuclear DNA and therefore useful for estimation of the variability within and between populations. Secondly, the mtDNA is inherited only from the females and therefore easier to reconstruct the line of evolution of the parasite. In other words, the lack of recombination makes it good marker to reconstruct the genealogy (evolutionary line) between different populations (Nadler and Leon, 2012). Blouin (2002) argues that mitochondrial markers could be more useful than ITS sequences for the identification of new species since ITS might have significant variation within the same individual. However, caution should always be paid whenever using mtDNA since accidental amplification of nuclear mitochondrial pseudogenes (numts) could occur that might lead to inaccurate phylogenies. Numts are mitochondrial homologous pseudogenes that are present in the nuclear genome (Bensasson et al, 2001). Moreover, there are limitations for mitochondrial markers (if used alone) as process such as incomplete lineage sorting and introgression via hybridization cannot be detected which can lead to inaccurate phylogeny (Anderson, 2001).

2. Molecular Biological Methodologies

There are a number of molecular methods that are used to study the population genetics, ecology and phylogenetics of the parasites reviewed by Gasser (2006). Herein I will however only present some of these PCR techniques.

Polymerase Chain Reaction (PCR) and PCR linked techniques

The PCR enables amplification of certain region in the parasite genome i.e. the marker of interest (Gasser, 2006). In a PCR reaction, the double-stranded genomic DNA template is denatured by heating, and the temperature is then decreased to allow artificial synthesized oligonucleotide primers to bind (anneal) to their complementary sequences on the target strands of the DNA template. The template directed DNA synthesis (extension) then proceeds in both directions from the primer sites using a thermostable DNA polymerase that results in double-stranded DNA products. This synthesis is usually repeated 20–40 times in an automated thermal cycler (PCR machine). The amplified DNA can be used for direct sequencing, DNA fingerprinting methods (e.g. restriction fragment length polymorphism (RFLP) or Random amplified polymorphic DNA (RAPD)) or mutation scanning methods such as single strand conformation polymorphism SSCP (reviewed by Gasser, 2006). However, due to the advances of the sequencing technology and PCR amplification, currently, DNA sequencing is the most applied approach for finding and delimiting new species (Nadler and Leon, 2012).

3. Computational analysis

A number of computational analyses are commonly used to explore the genetic data and test the evolutionary relationship of the DNA sequences in a statistical framework. Herein, I will briefly present some of these analysis for phylogenetics and population genetics namely, phylogenetic trees and networks, in addition to the Coalescent analysis.

3.1. Phylogenetic Tree

Phylogenetic trees are a classical way to infer the evolutionary relationship between taxa assuming a bifurcating pattern of evolution. There are two types of phylogenetic tree, distance based and character based.

3.1.1. Distance based Trees

The distance based methods depends on recording the genetic distance (dissimilarity) between the aligned sequences. There are several types of distance based methods such as Minimum Evolution (ME) and Neighbor Joining (NJ). Herein, I will present NJ an example

Neighbor Joining (NJ)

NJ tree construction starts from a tree in which all branches as assumed to be equal forming star shape with one node. Then, the two closest sequences (neighbors) are clustered together forming another internal node as shown in Figure 4. The process is repeated until the whole tree is constructed (Brinkman and Leipe, 2001). The advantage here is that there is no assumption about the distances between nodes since they are directly calculated (Rizzio and Rouchka, 2007). The resulted tree is usually non-ultrametric (i.e. different distances from root to different leaves) unless the species evolve according to molecular clock hypothesis (rate of evolution is constant over time in all lineages). This method is advantageous since it is fast. Moreover, it has been suggested that combining NJ with bootstrap is a the best way to evaluate distance based trees (Van de Peer, 2009)

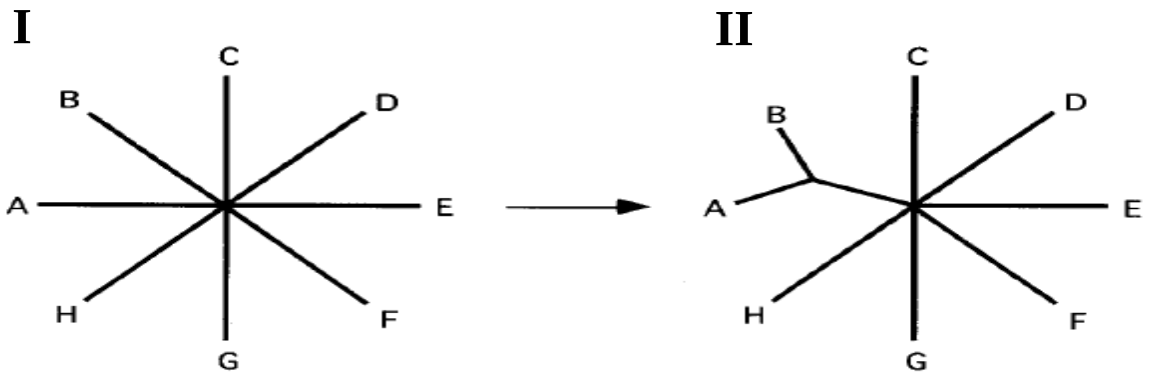

Figure 4. The tree building algorithm for the NJ tree. First, all branches are assumed to be equal making star shape with a single node (I), then the closest two terminals are joined together forming a new node (II). The process is continued until only one terminal is left forming single tree (Modified from (Brinkman and Leipe, 2001)).

3.1.2. Character based Trees

Also known as Discrete Data approaches. These methods compare the columns of the multiple aligned sequences (nucleotides in the aligned DNA or amino acids in protein sequences). The resulted trees are more precise and information rich. However they are computationally heavier and slower than distance based methods (Dowell, 2008).

Maximum Parsimony (MP)

The main criterion in the MP tree is that the correct tree is the one that can explain the differences between sequences sites with the fewest evolutionary changes (substitutions). For construction MP trees, there should be 'parsimony informative sites' which are sites that must have at least two nucleotides represented by at least two sequences in the population under study (Li, 1997). MP performs in a poor way if there is high heterogeneity in the DNA sequences which is a common observation due to backward or parallel substitution with a few possible solutions for the problem

(Nei, 1996; Brinkman and Leipe, 2001). Moreover, it is difficult to assess MP phylogeny in a statistical way since is difficult to compute the means and variances for the minimum number of the substitutions (Nei, 1996).

Maximum Likelihood (ML)

The main criterion for ML tree is for all possible trees, the one that gives the highest likelihood is chosen. For construction of ML tree, the probability (likelihood) of the variation for each site (nucleotide base or amino acid) in the sequence alignment is calculated by a given substitution model, a particular tree and the overall base frequencies. Then the likelihoods of all sites are multiplied to calculate the likelihood of the tree (Brinkman and Leipe, 2001). In other words, this method evaluates competing hypotheses (trees and parameters) by selecting those with the highest likelihood, meaning those that render the observed data most plausible (Liò and Goldman, 1998). Choosing the correct substitution model is crucial to obtain the correct tree as if the substitution model assumes equal rate of mutation while there is, for instance, a transition bias, it may lead to inaccurate results (Brinkman and Leipe, 2001). There are currently wide range of substitution models to choose from and methods to calculate the best to fit substitution model for a given dataset (Nei, 1996; Posada, 2008).

Bayesian Analysis

Bayesian approach is a relatively recent analysis which is based on ML method. It utilizes Markov Chain Monte Carlo (MCMC) simulations and an evolutionary model to produce posterior probability distribution of trees representing the data (Archibald et al, 2003). The posterior probability for a hypothesis is proportional to the likelihood multiplied by the prior probability of that hypothesis. Prior probabilities of a given hypothesis are our beliefs on this hypothesis before having seen the data. In other words, prior probability is set by us before any data are observed. Then priors are modified by the likelihood function to give the posterior probabilities (Holder and Lewis, 2003). Hence, finally, the posterior probability describes the probability of trees given the priors, the model and the data (Archibald et al, 2003). In most applications, prior probability distributions are set in a way that they are believed to be largely uninformative, so that most of the differences in the posterior probability of hypotheses are due to differences in the likelihood. The advantage of the Bayesian over ML is that it could be faster than bootstrapped ML tree. Moreover, Bayesian allows for more complex mutational models that could be more realistic than simple models (Holder and Lewis, 2003).

3.1.3. Bootstrapping

Bootstrapping is a popular method to assess the reliability of the tree. It works by re-sampling new columns from the aligned sequences data set with replacement and makes a new data set then builts the tree for the "new" data sets. This process continues hundreds of times to assess the strength of the data (Brinkman and Leipe, 2001). It has been found that under favorable conditions (equal mutation rates and symmetric branches) bootstrap values >70 indicates highly accurate phylogeny while in less favorable conditions (high rate of heterogeneity between taxa and unequal rate of mutation among taxa) high bootstrap value, can make wrong phylogeny looks accurate (Hillis and Bull, 1993; Brinkman and Leipe, 2001).

3.2. Phylogenetic network

Phylogenetic network is a broad term as defined by Huson and Bryant (2005) "any network in which a taxa are represented by nodes and their evolutionary relationships represented by edges". Given this definition, there are three types of networks. Phylogenetic trees could be regarded as specific type of the phylogenetic network. The other two types are split network and reticulate network.

Split network is used to represent incompatible and ambiguous sites in the data set. Nodes in this network do not necessarily represent ancestral species and they are connected by edges (equivalent to branches in trees). On the other hand, nodes at reticulate network represent ancestral species. A node with more than one parent indicates reticulate events such as hybridization or recombination (Huson and Bryant, 2005). In general, networks are advantageous since they could infer reticulate evolution (recombination, hybridization, horizontal gene transfer) and detect conflicting signals on the data set (Mardulyn, 2012).

3.3. Coalescent analysis

Coalescent as described previously is a mathematical method to describe the genealogical history of selectively neutral alleles sampled from a population (Lymbery and Thompson, 2012). The main parameters that are estimated in the coalescent analysis are the coalescent time (time in terms of number of generations until they all merge into a single most recent common ancestor) and theta which is the scaled product of mutation rate and effective population size (Sigwart, 2009). Effective population size is the ideal population size (with random mating, random variation in reproductive success, equal sex ratio and non-overlapping generation) that would experience the same magnitude of genetic change through random genetic drift as observed in real population (Archie et al, 2008). Mutation rate is the probability of a change of the nucleotides in the genome between parents and the offspring (Barrick and Lenski, 2013).

Coalescent analysis is implemented in many software packages. Each software package has a number of assumptions of which some are different from one to another. However, some general assumptions are hypothesized in all the software such as random sampling, no selection and random mating (reviewed by Kuhner, 2008). In principle, there are two algorithms used by the different coalescent analysis software packages. One algorithm is (IS) that stands for 'importance sampling' implemented in GENETREE (Griffiths and Tavare, 1993). This algorithm relies on the infinite site model of mutation which means that each nucleotide in the DNA sequence mutates only once. The other algorithm is correlated sampling (CS) implemented in a number of software such as BEAST (Drummond and Rambaut, 2007) and LAMARC (Kuhner, 2006). On contrary to IS, the CS algorithm allows a wider range of mutational models including relaxed molecular clock models which allow for different mutation rates between lineages. Generally, IS is appropriate for DNA sequences of low polymorphism while CS is appropriate for highly polymorphic sequences (Kuhner, 2008).

Next generation sequencing (NGS) technologies and parasite genomics

Next generation sequencing (NGS) technology is high-throughput DNA sequencing platforms that enable sequencing of genome (or specific region in the genome) of the living organisms in a greatly faster manner than traditional methods. The main difference between NGS and traditional methods (Sanger) is the ability to process millions of reads (sequences) in parallel unlike the limited number of reads in the traditional methods that enables to finish the experiment in one run or two (Mardis, 2008). Hence, NGS has many applications in a number of fields such as metagenomics and disease genetics (Metzker, 2010). In case of genome sequencing, the reads are then assembled to larger contigs until the whole genome is assembled. Currently, three main NGS platforms are commercially available namely, Roche 454 Genome Sequencer FLX sequencer, illumina (Solexa) genome analyzer and Applied Biosystem (SOLiD) sequencer. The three methods differ in their mode of function and sequencing capacity and Table 1 summaries some of the characteristics of these methods and some pros and cons. For how each of these technologies work, I refer to the review by Ansorge (2009).

Table (1). Brief comparison between the three main NGS platforms (Modified from (Mardis, 2008; Ekbolm and Galindo, 2011)).

	Roche (454)	Illumina	SOLiD
Capacity (MB/run)	100 Mb	1300 Mb	3000 Mb
Read length	250bp	32-40bp	35bp
Time/run	7 h	4 days	5 days
Cost/run	$8439	$8950	$17 447
Pros	Relatively long reads enables assembly of contigs even in the absence of a reference genome	Very deep coverage because of large number of reads	Large number of reads and very deep coverage. Low error rate because of duplicate sequencing of each base pair
Cons	Relatively few reads results in shallower coverage of sequencing.	Short read length means that a reference genome is desirable for assembly	Short read length means that a reference genome is desirable for assembly.

Parasite genomes

Thanks to the NGS technologies, many non-model animal genomes have now been sequenced including 23 parasitic and free living nematodes (Ellegren, 2014). Nematodes are a very diverse group of organisms that comprises free living nematodes and parasites of animal and plant hosts. Comparative genomic studies between parasitic and free living nematodes could provide insights on the key genetic factors for parasite adaptation to its host (Gilabert and Wasmuth, 2013). For example, analysis of the *Brugia malayi* genome indicates that 20% of *B. malayi* genes are specific suggesting the role of this huge pool of genes in the nematode defense against human and insect hosts (Dieterich and Sommer, 2009). Also, Comparative genomic studies could provide valuable insights on identifying potential targets for control as had been already shown for number of nematodes (Mitreva et al, 2007).

Parasite mitochondrial genome

Mitochondrial genome is a small circular genome of a size range of ~ 13-26 kpb that encodes for enzymes required for oxidative phosphorylation (Hu and Gasser, 2006). Mitochondrial genome provides a rich source of markers that could be used in systematic, ecology, molecular epidemiology and population structure of many helminthes (Jex et at, 2010). The high variation in the

mitochondrial DNA enables the investigation of evolutionary relationships between parasite species (phylogenetic analyses) and populations 'strains' of the same species (population structure analysis) (Blouin, 2002; Jex et al, 2010). Moreover, mitochondrial genome studies provide information about molecular biology and the evolutionary processes related to the mitochondrial genome such as overall organization, unusual gene structure and horizontal gene transfer (Burger et al, 2003). For example, gene rearrangement which occurs due to insertions, deletions or recombination in the mitochondrial genome has been observed within different parasitic nematode species (Burger et al, 2003; Jex et al, 2010). Another example is the RNA editing which was observed in trypanosomatids mitochondrial genome (Sloof and Benne, 1997; Burger et al, 2003).

The first NGS platform applied on mitochondrial genome sequencing was Roche 454 platform on the nematode *Heamonchus contortus* (Jex et al, 2008). This experiment evaluated NGS accuracy by comparing mitochondrial genome output of NGS with the publicly available Expressed sequence tags (EST) and genome survey sequence (GSS). The accuracy was estimated to be around 99.8% which lend confidence in the 454 platform (Jex et al, 2010). Currently, many mitochondrial genomes of parasitic nematodes have been sequenced using NGS such as *Hypodontus macropi* (Jabbar et al, 2013a), *Protostrongylus rufescens* (Jabbar et al, 2013b) and *Trichinella murrelli* (Webb and Rosenthal, 2011).

NGS workflow for mitochondrial genome

NGS workflow differs depending on the target genome to be sequenced. Herein, I will focus on mitochondrial genome. Sequencing mitochondrial genomes by NGS can be divided in to two phases. First phase is sample preparation and high throughput sequencing (HTS); the second is the bioinformatics phase (Figure 5) (Jex et al, 2010).

Sample preparation and HTS step involve identification of the parasite species, PCR enrichment of the targeted mitochondrial genome and HTS by one of the NGS platforms. Identification of the parasite species can be done by sequencing a suitable marker e.g. ITS-2. Then, PCR enrichment for the mitochondrial genome can be done by amplification of the mitochondrial genome in fragments, typically two fragments of 5 and 10 kbp using long range PCR (Hu et al, 2007; Jex et al, 2008). The last step is to add equimolar amounts of the fragments before it is going to one of the NGS platform for sequencing.

Second phase is the bioinformatics. It consists of manipulating the raw output data of the NGS to give the final targeted mitochondrial genome. NGS raw output consists of hundreds to millions of reads. These reads are processed with bioinformatic tools until they are assembled to the whole mitochondrial genome. The bioinformatic manipulation involves three steps, quality control (QC), assembly and annotation. QC means to check the quality of the reads that are going to be assembled. Reads that either shorter than certain threshold value, have unidentified base 'N' or have quality value below 20 (Q<20) are excluded (McCormak et al, 2013). Assembly is connecting and combining the short NGS reads into longer sequences (contigs). Assembly could be *de novo* (without reference genome) or using reference genome. Assembly with reference genome is easier as it allows direct alignment with the reference genome (McCormak et al, 2013). The second step after assembly is the annotation which is the detection of the open reading frame (ORF) for each gene in the genome by identifying start, stop and genes' amino acids codons.

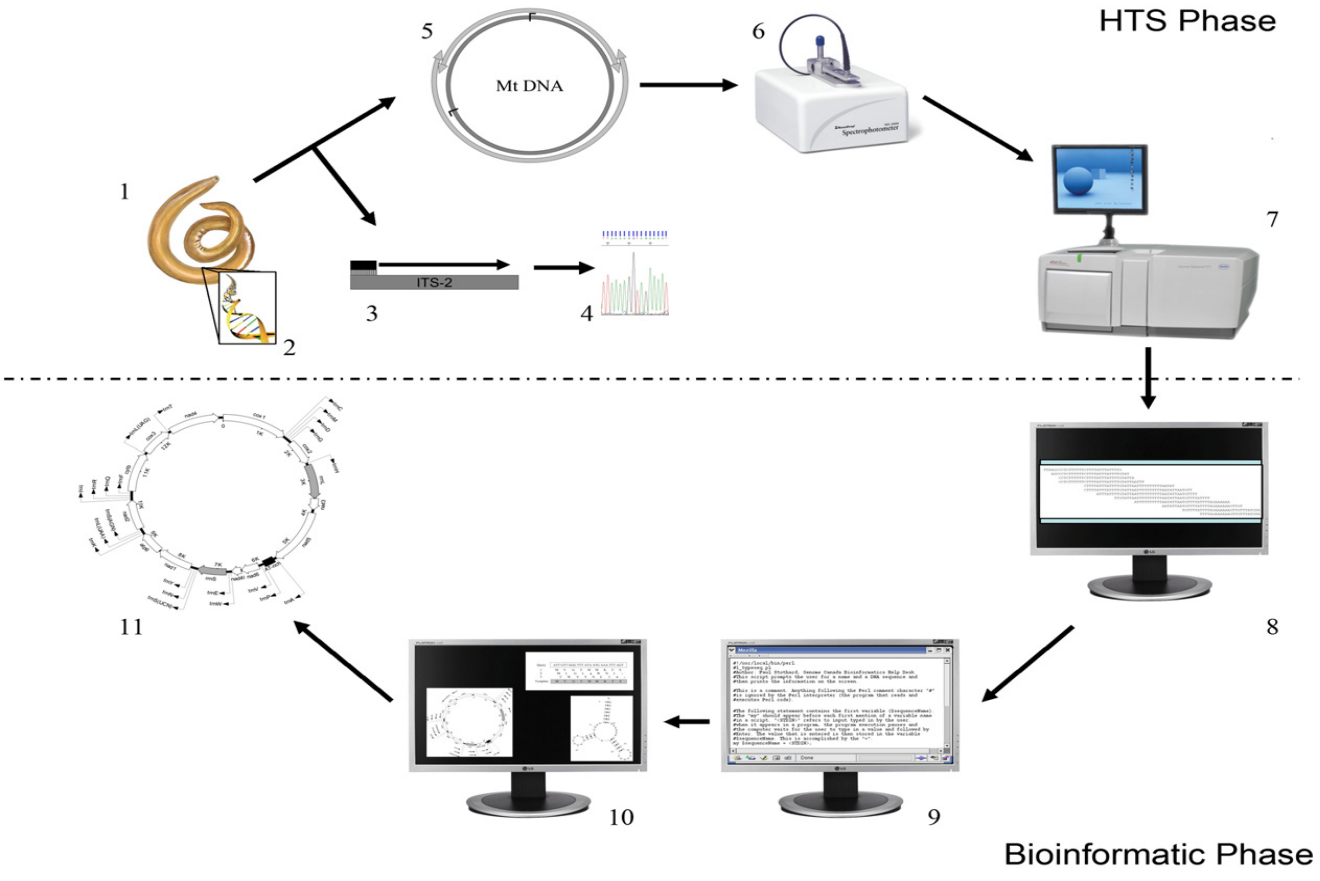

Figure 5. Diagrammatic illustration for the NGS workflow for mitochondrial genome. Steps from 1-7 is the species identification and high throughput sequencing phase and steps from 8-11 is the bioinformatics phase. Step 1-4 is parasite species identification, 5 is PCR amplification, 6 is pooling PCR amplicons, 7 is sequencing, 8 is quality control, 9 is assembly of the reads, 10 is the annotation and 11 is mapping and further analysis of the annotated genome (Modified from (Jex et al, 2010))

Thesis's studies introduction

In this chapter, I will present briefly the incentives, aims and major findings for the studies presented here. Besides, limitations of the studies and conclusions and perspectives on further studies are discussed. The detailed studies are presented in manuscripts 1 and 2.

Brief introduction and major findings

Trichuris trichiura and *T. suis* are two whipworms found in humans and pigs. *T. suis* is widespread in nearly all over the world and may cause economical losses for the pig industry (Roepstorff et al, 1998; Solaymani-Mohammadi et al, 2003; Tamboura et al, 2006; Roepstorff et al, 2011). Moreover, it was found that *T. suis* may cause zoonotic infections in humans (Nissen et al, 2012). *T. trichiura* in humans infects around 460 million people mainly in impoverished areas of poor sanitation causing 0.64 million DALYs (Pullan et al, 2014). Several studies suggest that humans and non human primates may share same *Trichuris* species suggesting a potential risk of zoonotic infections of *Trichuris* from non human primates to humans. Moreover, it is unknown whether the widespread of *T. suis* around the globe is accompanied by historical host switching events from humans due to domestication of animals. Herein, a hypothesis was tested regarding the transmission and natural history of *Trichuris* spp. between pigs and humans which is

- The *Trichuris* sp. was infecting either humans or pigs before it was transmitted from one host to the other in the Neolithic era upon domestication of animals when the two hosts were in sympatry nearly 10,000 years ago.

I also studied different populations of *Trichuris* infecting humans and a non human primate (baboon) to investigate the evolutionary and taxonomic relationship between them.

Different population genetic approaches namely, coalescent analysis, Analysis of Molecular Variance (AMOVA) and phylogenetic analysis were applied using two mitochondrial markers namely, the *nad*1 and *rrn*L genes on *T. suis* populations from domestic pigs from Denmark, USA, Uganda and China; *T. trichiura* from humans from China and Uganda; *Trichuris* from olive and yellow baboons held in captivity from Denmark (Copenhagen Zoo and Knuttenborg Park) and USA (Southwest National Primate Research Center (SNPRC) in Texas). I found strong genetic differentiation between the populations of *T. suis* that cannot be explained by the tested hypothesis. Moreover, the coalescent analysis of the different populations of *Trichuris* spp. from pigs and

humans revealed independent demographic history of the parasite in the two hosts indicating that the speciation events happened much older than the Neolithic era 10,000 years ago and hence the hypothesis was rejected. Moreover, the demographic history of *T. suis* hints to the role played by the human activity in the spreading of *T. suis* associated with the pig transport. No genetic differentiation was found between *Trichuris* obtained from baboons and humans in Uganda suggesting to be the same species (*T. trichiura*). This is in accordance with a previous study by Hansen et al (2013) and emphasize on the potential of *Trichuris* from baboons to cause zoonotic infections. However, remarkably, the genetic distance between *T. trichiura* from humans from China and Uganda was very high considering intra-species variation (Blouin 2002), indicating the possible presence of cryptic species of *Trichuris* infecting humans which might have important implications on the control measures in different regions.

In the second study, the aim is to sequence full mitochondrial genome of *Trichuris* recovered from baboons and pigs and compare with the other available mitochondrial genomes of *Trichuris* from human, non human primate and pigs. Mitochondrial DNA has a number of advantages for detection of cryptic species due its high substitution rate coupled to low effective population size that leads to rapid lineage sorting following speciation (distinctiveness between different species in short time) (Nadler and Leon, 2012). In this study, I did the sequencing and the annotation of the whole mitochondrial genome from two different haplotypes of *Trichuris* recovered from olive baboons (*Papio anubis*) and conducted phylogenetic analysis and compared the genetic and protein sequences distances of mitochondrial genes with other avilable the mitochondrial genome of *Trichuris* derived from human from China and from a non human primate (François' leaf-monkey) to investigate the evolutionary and genetic relatedness between different *Trichuris* spp. in these primates. Similarly, the full mitochondrial genomes of *T. suis* from pigs from Uganda and Denmark were sequenced and annotated then phylogenetic and comparative analysis was conducted with the already available *T. suis* from China. Moreover, to investigate the phylogenetic relationship of the mitochondrial genome haplotypes of the *Trichuris* spp. in this study with other *Trichuris* spp. haplotypes recovered from pigs and non human primates in other regions, I conducted phylogenetic analysis on partial sequences of *cox*1 gene by including *cox*1 sequences from the GenBank namely *Trichuris* from domestic pigs from Spain and China, from wild pigs and mantled guereza (a non human primate) from Spain and from different baboon species and macaque (non human primates) from Czech Republic. The complete mitochondrial genome of *Trichuris* sp. from the two baboon worms were of 14,105 and 14,009 bp and for *Trichuris* from pigs from Denmark and Uganda were 14,586 and 14,410 bp respectively. Annotation of the genomes indentified 13 PCGs, 22 transfer RNA (tRNA)

and 2 ribosomal DNA (rDNA) genes in each genome. The genetic distance between the *Trichuris* spp. recovered from the two different non human primates (olive baboon and François' leaf-monkey) was very high and suggests the presence of different *Trichuris* spp. infecting non human primates. Remarkably, I found that the genetic distance between the two *Trichuris* worms derived from baboons were genetically very distant that suggests there could be different *Trichuris* spp. infecting olive baboons. However, one of the baboon *Trichuris* was genetically related to the human *Trichuris* which might be the same species. The *cox*1 phylogeny suggested there could be five potential different species of *Trichuris* infecting different primates. Likewise, I found high sequences divergence in protein and genetic sequences between the *T. suis* from pigs in Denmark to that from Uganda and China suggesting also possible presence of cryptic species of *Trichuris* infecting domestic pigs in these areas. The *cox*1 phylogeny clustered the Spanish *T. suis* from pigs with Danish *T. suis* from pigs which might suggest the *T. suis* from European pigs is the same species and is different from *T. suis* in pigs from China. Also the overall sequence differences in the *T. suis* from Uganda and China was limited suggesting to be the same species.

Limitations

In the first study, I found some samples of human *Trichuris* from Uganda and China showed duplicated peaks in the chromatogram files indicating co-amplification of non targeted DNA fragments from the parasite genome by PCR. I sequenced these samples twice and the double peaks were observed in the second time as well. This non targeted DNA could be numts which are mitochondrial pseudogenes present in the nuclear genome or other homologous sequences in the nuclear genome. There are two ways to overcome this problem. First solution is bioinformatically by doing BLAST search of the mitochondrial genome against the nuclear genome of the parasite to detect the homologous regions between the two genomes and hence avoid these regions. However, as the nuclear genome draft was not available for the *T. trichiura* and *T. suis* in the time of study, this solution was not available. The second solution is experimentally by performing cloning for the PCR amplicons before sequencing which is a preferable method and could identify precisely the co-amplifying DNA fragments. However, due to lack of funding, this solution was not available as well. Another drawback is that I used only mitochondrial markers to reconstruct the demographic history through coalescent analysis of the *Trichuris* populations from different hosts. Coalescent theory assumes neutrality of the DNA markers used which might not be true for mitochondrial DNA. However, presence of selection in the mitochondrial genes was suggested not to affect the topology of the tree, but it may affect the divergence time and the effective population size estimations (negative selection might make the coalescent more recent whereas the positive selection might

make it appears as older) (Gerber et al, 2001; Archie et al, 2008). In the second study, the aim was to also include *T. trichiura* in humans from Uganda to compare it with the *Trichuris* in human in China to evaluate if they are two distinct species as suggested in the first study. Unfortunately, the long range PCR products produced for the human *Trichuris* could not be sequenced by the Next Generation Sequencing platform due to technical problems during the library construction and hence it was excluded. Other caveats are also discussed in the context of the manuscripts.

Conclusions and Perspectives

I found different demographic history of *Trichuris* spp. infecting humans and pigs with no sign of host switching indicating that speciation of *Trichuris* in both hosts happened independently of the domestication of animals. Consistent with this, recently, the genome and transcriptome of *T. trichiura* and *T. suis* were investigated and there were a major difference between the two parasites indicating high level of adaptation of each parasite to its host (Ghedin, 2014). Moreover, the demographic history of *T. suis* suggests the major role played by the human trading activity in transmitting the parasite into new areas through pig transport. I found also the genetic distance between *T. trichiura* from humans in Uganda and China are very distant suggesting a different species and further studies using nuclear markers (e.g. ITS) could be used to confirm whether there could be different *Trichuris* species infecting humans. Presence of different *T. trichiura* cryptic species infecting humans may have important implications in order to implement proper control measures in different places. On the other hand, the *Trichuris* from baboons was found to be belonging to the same *Trichuris* species that infects humans and therefore can cause zoonotic infections. Further studies on the natural, sympatric habitat of baboons and humans could reveal the nature of cross transmission between the two hosts. The mitochondrial genome analysis of *T. suis* from pigs from Denmark, China and Uganda also revealed high genetic and protein distance between *T. suis* from Denmark and other *T. suis* from Uganda and China suggesting possible presence of different cryptic species of *Trichuris* in pigs in different geographical regions which may have implications on setting proper control measures. The mitochondrial genome of *Trichuris* from different non human primates (baboon and Francois' leaf-monkey) were genetically very distinct emphasizing on that *Trichuris* from different non human primates is a complex multiple species. Further studies on the genetic relatedness using nuclear markers (e.g. ITS) and morphological features are warranted to test if it is concordance with the genetic distinctiveness found in the mitochondrial DNA and hence elucidate the speciation. The mitochondrial genomes for *Trichuris* from non human primates and pigs provide a rich source of markers for further studies on parasite genealogy and molecular ecology.

Referneces

Abee C.R., Mansfield, K., Tardif, S.D. and Morris, T. (2012). Nonhuman primates in biomedical research (2nd Edition, pp 243-244). London: Elsevier/Academic

Anderson, T. J. (2001). "The dangers of using single locus markers in parasite epidemiology: *Ascaris* as a case study." Trends Parasitol 17(4): 183-188.

Ansorge, W. J. (2009). "Next-generation DNA sequencing techniques." N Biotechnol 25(4): 195-203.

Archibald J.K, Mort M.E and Crawford D.J. (2003). "Bayesian inference of phylogeny: a non-technical primer." Taxon, 52: 187–191

Archie, E. A., G. Luikart and V. O. Ezenwa (2008). "Infecting epidemiology with genetics: a new frontier in disease ecology." Trends Ecol Evol 24(1): 21-30.

Barrick, J. E. and R. E. Lenski (2013). "Genome dynamics during experimental evolution." Nat Rev Genet 14(12): 827-839.

Batte, E. G., R. D. McLamb, K. E. Muse, S. D. Tally and T. J. Vestal (1977). "Pathophysiology of swine trichuriasis." Am J Vet Res 38(7): 1075-1079.

Beer, R. J. (1976). "The relationship between *Trichuris trichiura* (Linnaeus 1758) of man and *Trichuris suis* (Schrank 1788) of the pig." Res Vet Sci 20(1): 47-54.

Beer, R. J. and I. J. Lean (1973). "Clinical trichuriasis produced experimentally in growing pigs. I. Pathology of infection." Vet Rec 93(7): 189-195.

Behnke, J. M. and D. Wakelin (1973). "The survival of *Trichuris muris* in wild populations of its natural hosts." Parasitology 67(2): 157-164.

Bensasson, D., D. Zhang, D. L. Hartl and G. M. Hewitt (2001). "Mitochondrial pseudogenes: evolution's misplaced witnesses." Trends Ecol Evol 16(6): 314-321.

Bethony, J., S. Brooker, M. Albonico, S. M. Geiger, A. Loukas, D. Diemert and P. J. Hotez (2006). "Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm." Lancet 367(9521): 1521-1532.

Betson, M., P. Nejsum, R. P. Bendall, R. M. Deb and J. R. Stothard (2014). "Molecular Epidemiology of Ascariasis: A Global Perspective on the Transmission Dynamics of *Ascaris* in People and Pigs." J Infect Dis.

Blouin, M. S. (2002). "Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer." Int J Parasitol 32(5): 527-531.

Brinkman, F. S. and D. D. Leipe (2001). "Phylogenetic analysis." Methods Biochem Anal 43: 323-358.

Bundy, D.A.P. and E.S. Cooper (1989). "*Trichuris* and Trichuriasis in Humans." Adv. Parasitol. 28: 107-173.

Burden, D. J. and N. C. Hammet (1979). "The development and survival of *Trichuris suis* ova on pasture plots in the south of England." Res Vet Sci 26(1): 66-70.

Burden, D. J., N. C. Hammet and P. A. Brookes (1987). "Field observations on the longevity of *Trichuris suis* ova." Vet Rec 121(2): 43.

Burger, G., M. W. Gray and B. F. Lang (2003). "Mitochondrial genomes: anything goes." Trends Genet 19(12): 709-716.

Carter, N., C. Dewey, F. Mutua, C. de Lange and D. Grace (2013). "Average daily gain of local pigs on rural and peri-urban smallholder farms in two districts of Western Kenya." Trop Anim Health Prod 45(7): 1533-1538.

Conole, J. C., N. B. Chilton, T. Jarvis and R. B. Gasser (2001). "Mutation scanning analysis of microsatellite variability in the second internal transcribed spacer (precursor ribosomal RNA) for three species of *Metastrongylus* (Strongylida: Metastrongyloidea)." Parasitology 122(Pt 2): 195-206.

Criscione, C. D., R. Poulin and M. S. Blouin (2005). "Molecular ecology of parasites: elucidating ecological and microevolutionary processes." Mol Ecol 14(8): 2247-2257.

Cutillas, C., M. de Rojas, C. Ariza, J. M. Ubeda and D. Guevara (2007). "Molecular identification of *Trichuris vulpis* and *Trichuris suis* isolated from different hosts." Parasitol Res 100(2): 383-389.

Cutillas, C., R. Callejon, M. de Rojas, B. Tewes, J. M. Ubeda, C. Ariza and D. C. Guevara (2009). "*Trichuris suis* and *Trichuris trichiura* are different nematode species." Acta Trop 111(3): 299-307.

De Luca, F., A. Reyes, J. Grunder, P. Kunz, A. Agostinelli, C. De Giorgi and F. Lamberti (2004). "Characterization and Sequence Variation in the rDNA Region of Six Nematode Species of the Genus *Longidorus* (Nematoda)." J Nematol 36(2): 147-152.

De Meeus, T., K. D. McCoy, F. Prugnolle, C. Chevillon, P. Durand, S. Hurtrez-Bousses and F. Renaud (2006). "Population genetics and molecular epidemiology or how to "debusquer la bete"." Infect Genet Evol 7(2): 308-332.

Dieterich, C. and R. J. Sommer (2009). "How to become a parasite - lessons from the genomes of nematodes." Trends Genet 25(5): 203-209.

Dowell, K. (2008). Molecular phylogenetics. Technical report, University of Maine.

Drummond, A.J. and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7: 214

Ekblom, R. and J. Galindo (2011). "Applications of next generation sequencing in molecular ecology of non-model organisms." Heredity (Edinb) 107(1): 1-15.

Ellegren, H. (2014). "Genome sequencing and population genomics in non-model organisms." Trends Ecol Evol 29(1): 51-63.

Gasser, R. B. (2006). "Molecular tools--advances, opportunities and prospects." Vet Parasitol 136(2): 69-89.

Gerber, A. S., R. Loggins, S. Kumar and T. E. Dowling (2001). "Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes?" Annu Rev Genet 35: 539-566.

Ghedin, E. (2014). "Panning for molecular gold in whipworm genomes." Nat Genet 46(7): 661-663.

Gilabert, A. and J. D. Wasmuth (2013). "Unravelling parasitic nematode natural history using population genetics." Trends Parasitol 29(9): 438-448.

Griffiths, R.C. and Tavare´ , S. (1993). "Sampling theory for neutral alleles in a varying environment." Proc. R. Soc. Lond B Biol Sci 344: 403–410

Hale, O. M. and T. B. Stewart (1979). "Influence of an experimental infection of *Trichuris suis* on performance of pigs." J Anim Sci 49(4): 1000-1005.

Hansen, T. V., S. M. Thamsborg, A. Olsen, R. K. Prichard and P. Nejsum (2013). "Genetic variations in the beta-tubulin gene and the internal transcribed spacer 2 region of *Trichuris* species from man and baboons." Parasit Vectors 6(1): 236.

Holder, M. and P. O. Lewis (2003). "Phylogeny estimation: traditional and Bayesian approaches." Nat Rev Genet 4(4): 275-284.

Hotez, P. J. and A. Kamath (2009). "Neglected tropical diseases in sub-saharan Africa: review of their prevalence, distribution, and disease burden." PLoS Negl Trop Dis 3(8): e412.

Hu, M., A. R. Jex, B. E. Campbell and R. B. Gasser (2007). "Long PCR amplification of the entire mitochondrial genome from individual helminths for direct sequencing." Nat Protoc 2(10): 2339-2344.

Hu, M. and R. B. Gasser (2006). "Mitochondrial genomes of parasitic nematodes--progress and perspectives." Trends Parasitol 22(2): 78-84.

Hillis, M.D. and Bull, J. J. (1993)." An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis". Syst. Biol. 4(42): 182-192.

Huson, D. H. and D. Bryant (2005). "Application of phylogenetic networks in evolutionary studies." Mol Biol Evol 23(2): 254-267.

Jabbar, A., I. Beveridge, N. Mohandas, N. B. Chilton, D. T. Littlewood, A. R. Jex and R. B. Gasser (2013a). "Analyses of mitochondrial amino acid sequence datasets support the proposal that specimens of *Hypodontus macropi* from three species of macropodid hosts represent distinct species." BMC Evol Biol 13: 259.

Jabbar, A., N. Mohandas, A. R. Jex and R. B. Gasser (2013b). "The mitochondrial genome of *Protostrongylus rufescens* - implications for population and systematic studies." Parasit Vectors 6(1): 263.

Jex, A. R., D. T. Littlewood and R. B. Gasser (2010). "Toward next-generation sequencing of mitochondrial genomes--focus on parasitic worms of animals and biotechnological implications." Biotechnol Adv 28(1): 151-159.

Jex, A. R., M. Hu, D. T. Littlewood, A. Waeschenbach and R. B. Gasser (2008). "Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda)." BMC Genomics 9: 11.

Kingman, J. F. (2000). "Origins of the coalescent. 1974-1982." Genetics 156(4): 1461-1463.

Kringel, H. and A. Roepstorff (2006). "*Trichuris suis* population dynamics following a primary experimental infection." Vet Parasitol 139(1-3): 132-139.

Kuhner, M. K. (2006). "LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters." Bioinformatics 22(6): 768-770.

Kuhner, M. K. (2008). "Coalescent genealogy samplers: windows into population history." Trends Ecol Evol 24(2): 86-93.

Lai, M., R. Q. Zhou, H. C. Huang and S. J. Hu (2011). "Prevalence and risk factors associated with intestinal parasites in pigs in Chongqing, China." Res Vet Sci 91(3): e121-124.

Li, W. (1997). Molecular evolution (pp 112-116). Massachussets: Sinauer.

Liò, P and Goldman, N (1998) "Models of Molecular Evolution and Phylogeny." Genome Res 8: 1233-1244.

Liu, G. H., R. B. Gasser, A. Su, P. Nejsum, L. Peng, R. Q. Lin, M. W. Li, M. J. Xu and X. Q. Zhu (2012a). "Clear genetic distinctiveness between human- and pig-derived *Trichuris* based on analyses of mitochondrial datasets." PLoS Negl Trop Dis 6(2): e1539.

Liu, G. H., Y. Wang, M. J. Xu, D. H. Zhou, Y. G. Ye, J. Y. Li, H. Q. Song, R. Q. Lin and X. Q. Zhu (2012b). "Characterization of the complete mitochondrial genomes of two whipworms *Trichuris ovis* and *Trichuris discolor* (Nematoda: Trichuridae)." Infect Genet Evol 12(8): 1635-1641.

Lymbery, A. J. and R. C. Thompson (2012). "The molecular epidemiology of parasite infections: tools and applications." Mol Biochem Parasitol 181(2): 102-116.

Mardis, E. R. (2008). "The impact of next-generation sequencing technology on genetics." Trends Genet 24(3): 133-141.

Mardulyn, P. (2012). "Trees and/or networks to display intraspecific DNA sequence variation?" Mol Ecol 21(14): 3385-3390.

McCormack, J. E., S. M. Hird, A. J. Zellmer, B. C. Carstens and R. T. Brumfield (2013). "Applications of next-generation sequencing to phylogeography and phylogenetics." Mol Phylogenet Evol 66(2): 526-538.

Metzker, M. L. (2010). "Sequencing technologies - the next generation." Nat Rev Genet 11(1): 31-46.

Mitreva, M., D. S. Zarlenga, J. P. McCarter and D. P. Jasmer (2007). "Parasitic nematodes - from genomes to control." Vet Parasitol 148(1): 31-42.

Nadler, S. A. and Leon D. E. (2012). "Integrating molecular and morphological approaches for characterizing parasite cryptic species: implications for parasitology." Parasitology 138(13): 1688-1709.

Nansen, P. and A. Roepstorff (1999). "Parasitic helminths of the pig: factors influencing transmission and infection levels." Int J Parasitol 29(6): 877-891.

Nei, M. (1996). "Phylogenetic analysis in molecular evolutionary genetics." Annu Rev Genet 30: 371-403.

Nissen, S., A. Al-Jubury, T. V. Hansen, A. Olsen, H. Christensen, S. M. Thamsborg and P. Nejsum (2012). "Genetic analysis of *Trichuris suis* and *Trichuris trichiura* recovered from humans and pigs in a sympatric setting in Uganda." Vet Parasitol 188(1-2): 68-77.

Pittman JS, Shepherd G, Thacker BJ, Myers GH (2010). "*Trichuris suis* in finishing pigs: Case report and review." J Swine Health Prod. 18(6): 306–313.

Posada, D. (2008). "jModelTest: phylogenetic model averaging." Mol Biol Evol 25(7): 1253-1256.

Pullan, R. L., J. L. Smith, R. Jasrasaria and S. J. Brooker (2014). "Global numbers of infection and disease burden of soil transmitted helminth infections in 2010." Parasit Vectors 7: 37.

Ramesh, A., S. T. Small, Z. A. Kloos, J. W. Kazura, T. B. Nutman, D. Serre and P. A. Zimmerman (2012). "The complete mitochondrial genome sequence of the filarial nematode *Wuchereria bancrofti* from three geographic isolates provides evidence of complex demographic history." Mol Biochem Parasitol 183(1): 32-41.

Rizzo, J. and Rouchka, E.C. (2007). "Review of phylogenetic tree construction." University of Louisville Bioinformatics Laboratory Technical Report Series.

Roberts, L.S., and G.D. Schmidt (2009). Foundations of parasitology. (8th Edition, pp 399-401). Boston: McGraw Hill.

Roepstorff, A., H. Mejer, P. Nejsum and S. M. Thamsborg (2011). "Helminth parasites in pigs: new challenges in pig production and current research highlights." Vet Parasitol 180(1-2): 72-81.

Roepstorff, A., O. Nilsson, A. Oksanen, B. Gjerde, S. H. Richter, E. Ortenberg, D. Christensson, K. B. Martinsson, P. C. Bartlett, P. Nansen, L. Eriksen, O. Helle, S. Nikander and K. Larsen (1998). "Intestinal parasites in swine in the Nordic countries: prevalence and geographical distribution." Vet Parasitol 76(4): 305-319.

Rosenberg, N. A. and M. Nordborg (2002). "Genealogical trees, coalescent theory and the analysis of genetic polymorphisms." Nat Rev Genet 3(5): 380-390.

Sigwart, J. (2009). "Coalescent Theory: An Introduction." Syst. Biol. 58(1): 162-165.

Sloof, P. and R. Benne (1997). "RNA editing in kinetoplastid parasites: what to do with U." Trends Microbiol 5(5): 189-195.

Solaymani-Mohammadi, S., I. Mobedi, M. Rezaian, J. Massoud, M. Mohebali, H. Hooshyar, K. Ashrafi and M. B. Rokni (2003). "Helminth parasites of the wild boar, *Sus scrofa*, in Luristan province, western Iran and their public health significance." J Helminthol 77(3): 263-267.

Sternberg, Robert J., Powell, Christine, McGrane, Patricia, Grantham-McGregor, Sally (1997). "Effects of a parasitic infection on cognitive functioning." J. Exp. Psychol.-Appl 3(1): 67-76.

Sunnucks, P. (2000). "Efficient genetic markers for population biology." Trends Ecol Evol 15(5): 199-203.

Tamboura, H. H., H. Banga-Mboko, D. Maes, I. Youssao, A. Traore, B. Bayala and M. A. Dembele (2006). "Prevalence of common gastrointestinal nematode parasites in scavenging pigs of different ages and sexes in eastern centre province, Burkina Faso." Onderstepoort J Vet Res 73(1): 53-60.

Van de Peer, Y. (2009). Phylogenetic inference based on distance methods, in *The Phylogenetic Handbook* (P. Lemey, M. Salemi and A.-M. Vandamme, Eds.), (pp 142–160). Cambridge: Cambridge University Press.

Webb, K. M. and B. M. Rosenthal (2011). "Next-generation sequencing of the *Trichinella murrelli* mitochondrial genome allows comprehensive comparison of its divergence from the principal agent of human trichinellosis, *Trichinella spiralis*." Infect Genet Evol 11(1): 116-123.

Zhou, C., M. Li, K. Yuan, S. Deng and W. Peng (2012). "Pig *Ascaris*: an important source of human ascariasis in China." Infect Genet Evol 12(6): 1172-1177.

.

Manuscript 1

Mitochondrial DNA shows independent demographic history of *Trichuris trichiura* and *Trichuris suis*.

Mohamed Bayoumi Fahmy Hawash, Department of Veterinary Disease Biology, Faculty of Health and Medical Sciences, Copenhagen University.

Abstract

Trichuris suis and *T. trichiura* are two different whipworms that infect pigs and humans respectively. *Trichuris suis* is found in pigs in almost all over the world while *T. trichiura* is responsible for nearly 460 million infections in people mainly in areas of poor sanitation. The taxonomic status of whipworms in non-human primates is unsettled but they are normally designated as *T. trichiura*. The pattern and nature of transmission of *Trichuris* spp. infecting human, non human primates and pigs is poorly understood. In this study, a hypothesis was investigated regarding the transmission and the natural history of the *Trichuris* in humans and pigs which is that the *Trichuris* sp. infected humans or pigs before it was dropped from one host to the other upon domestication of animals in the Neolithic era when the two hosts were in sympatry. Moreover, the evolutionary and phylogenetic relationship between *Trichuris* infecting human and a non human primate (baboons) was studied. Partial sequencing of the *nad*1 and *rrn*L mitochondrial genes was conducted and AMOVA, phylogenetic and coalescent analyses were applied to test the genetic differentiation, relatedness and to infer the demographic history for the *Trichuris* spp. Populations of *T. suis* were obtained from domesticated pigs from Uganda, China, Denmark and USA; *T. trichiura* from humans from Uganda and China; *Trichuris* from baboons kept in captivity from Denmark and USA. Strong differentiation between populations of *T. suis* was detected that cannot be explained by the tested hypothesis. Moreover, the majority of baboon worms were genetically similar and clustered with the human worms from Africa suggesting high gene flow of *Trichuris* populations between the humans and baboons in Africa. However, the *T. trichiura* from humans from China and Uganda were genetically very distant suggesting possible different species of *T. trichiura* infecting humans which might have important consequences on implementing proper control measures. The coalescent analysis reveals independent demographic history for *T. suis* and *T. trichiura* suggesting that speciation event happened much older than Neolithic era. Moreover, the demographic history of *T. suis* underlies the major role played by human activity in transmitting the parasite to new areas through pig transport.

Key words: *Trichuris trichiura*, *Trichuris suis*, baboon, human, domestic pig, demographic history, mitochondrial DNA.

---

1. Introduction

The genus *Trichuris* (whipworm) comprises a number of species that infect a number of hosts of public health and economical importance. *Trichuris* spp. has a direct fecal-oral life cycle. *T. trichiura*

infects more than 460 million individual worldwide mainly in developing countries with an estimated total disability adjusted life years (DALYs) of 0.64 million years (Pullan et al, 2014). Moreover, traditionally, *T. trichiura* is believed to infect non human primates as well which is supported by a number of recent studies (Ravasi et al, 2012, Hansen et al, 2013). However, other study suggests presence of more than one species of *Trichuris* infects non human primates (Liu et al, 2013; Cutillas et al, 2014).

Trichuris suis infects pigs and although infection rarely causes severe disease in pigs, it is associated with reduced food utilization and weight gain leading to economical loss in the pig industry (reviewed by Roepstorff et al, 2011). In addition to its economic importance, *T. suis* also reported to cause zoonotic infections in sympatric populations with poor sanitation where humans and pigs are in close contact (Nissen et al, 2012).

Population genetic tools provide a valuable opportunity to investigate the epidemiological history and transmission of parasite and they were applied to study many parasitic nematodes (Archie et al, 2008; Gilabert and Wasmuth, 2013). Recently, population genetic approaches were applied to investigate the pattern of transmission of *Ascaris suum* and *A. lumbricoides* between pigs and humans around the globe (Betson et al, 2014). It was found that *Ascaris* infections in humans in Europe are of pig origin and there was evidence of cross-transmission between human and pigs in Africa. Moreover, population genetics can be used to reconstruct the epidemiological and demographic history of number of micro and macro parasites through coalescent analysis coupled with Bayesian approaches (Archie et al, 2008). Reconstruction of the epidemiological and demographic history gives us a window to the past to see which factors facilitated spreading or the introduction of the parasites to new regions (Archie et al, 2008). For example, the demographic history of *W. bancrofti* suggested that introduction of *W. bancrofti* to India could be done during the human migration out of Africa 60,000-70,000 years ago. However, the introduction of the parasite to Papua New Guinea cannot be explained by the past human migration but must have been introduced with more recent migration (Ramesh et al, 2012). Another study estimated the divergence time of the free living nematode *Pristionchus pacificus* from different continents (Molnar et al, 2011). *P. pacificus* is usually associated with beetles and it was suggested that the divergence time between all the populations studied started around 10,000-100,000 years ago that is consistent with the molecular phylogeny of the worm and the divergence time of the host (beetles).

T. suis infection in pigs are widespread in nearly all geographical locations (Roepstorff et al, 1998; Solaymani-Mohammadi et al, 2003; Tamboura et al, 2006). However, the factors responsible for this widespread are not known. Moreover, it is unknown if these widespread is associated with historical host switching events from *Trichuris* infecting humans due to human domestication of animals as has been suggested in *Ascaris* (Betson et al, 2014).

Several studies suggest that *T. trichiura* could infect humans and non-human primates (Ravasi et al, 2012; Hansen et al, 2013). However, these studies relied on markers which are either very conserved such as the beta tubulin gene (Hansen et al, 2013) or markers showing high degree of intra-individual variability and sequence ambiguity such as Internal transcribed spacer (ITS) (Ravasi et al ,2012).

The aim of this study was to test a hypothesis regarding the history and transmission of *Trichuris* between pigs and humans and to investigate the evolutionary and taxonomic relationship between *Trichuris* in humans and baboons. The tested hypothesis is that the *Trichuris* sp. in either humans or pigs was transmitted to the other host species in the Neolithic era 10,000 years ago when the two hosts where in sympatry. Different populations of *Trichuris* spp. recovered from three hosts namely humans, pigs and baboons held in captivity from various geographical localities were included in the study. Phylogenetic analysis was conducted to describe the evolutionary relationships between worms. Coalescent analysis was applied to infer the past demographic history and time of divergence for the different populations. Moreover, I investigated the level of genetic differentiation and variation between the different populations.

2. Material and methods

2.1. Sample collections

A total number of 109 *Trichuris* spp. from pigs, baboons and humans were collected from four different countries (Table 1). All worms were rinsed with tap water and stored in 70% ethanol at 5°C until DNA extraction.

Table 1. A summary of the number of samples, the country of origin, the host from which samples were recovered and the method of sampling.

Host (host numbers)	Country (number of samples)	Sampled localities in each country (number of samples)	Method of sampling
Domesticated pig, *Sus domesticus* (10)	Uganda (18)	Villages ranged 30 Km apart in south west Kabale district (18)	Post mortem
Domesticated pig, *Sus domesticus* (5)	China (14)	Guangdong Province (3), Fujian Province (3), Chongquing Province (4), Hunan Province (4)	Post mortem
Domesticated pig, *Sus domesticus* (2)	Denmark (10)	Experimentally infected pigs with local strains of the parasite (10)	Post mortem
Domesticated pig, *Sus domesticus* (2)	USA (10)	Experimentally infected pigs with local strains of the parasite (10)	Post mortem
Human 'school children' (12)	Uganda (17)	Villages ranged in south west of Kabale district (17)	Antihelmintic treatment
Human (1)	China (2)	Zhanjiang, Guangdong Province (2)	Unknown
baboons (*Papio anubis, P. cynocephalus*) (3)	Denmark (25)	Copenhagen Zoo (12), Knuttenborg Park (13)	Post mortem
baboons (*Papio anubis, P. cynocephalus*) (2)	USA (13)	Southwest National Primate Research Center (SNPRC) Texas (13)	Post mortem

2.2. DNA extraction

Genomic DNA was extracted from the thin, anterior part of the worms. MasterPure DNA Purifications Kit (Epicentre Biotechnologies) was used to extract the DNA according to manufacturer's protocol after the worm material had been homogenized in 300µl of lysis solution (295 µl tissue and cell lysis solution + 5µl proteinase K) in an eppendorf tube with matching plastic pestle.

2.3. Molecular typing of the worms

Worms were typed to confirm worm species by Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP) on two different markers namely internal transcribed spacer-2 (ITS-2) and the beta-tubulin gene. PCR-RFLP on ITS-2 was conducted following the protocol described elsewhere (Nissen et al, 2012). The beta-tubulin gene was amplified by PCR as described by Nissen et al (2012). Restriction sites in the amplified region of the beta-tubulin gene were

identified by Webcutter 2.0 using previously published sequences (Hansen et al, 2013). *Hinc*-II was identified as a potential endonulcease that could differentiate between *Trichuris* from primates and pigs as it produced fragments of 100, 220, 380 and 100, 200, 510, respectively. 5 µl of beta-tubulin PCR products were digested using 2 U of *Hinc*-II endonuclease using Tango buffer in a total volume 10 µl, and incubated at 37 °C for 120 min in a thermo block. Negative water control was included in all runs. PCR amplified fragments were stained by GelRed (Biotium) then visualized under UV light in 1.5% agarose gel. All *Trichuris* from baboons and humans showed banding pattern characteristic for *Trichuris* from primates while all *T. suis* worms showed the banding pattern characteristic for the *T. suis*.

2.4. Genetic markers amplification and sequencing

Two mitochondrial markers were used to estimate the genetic variation and to infer the demographic history of the parasite, namely the large ribosomal subunit (*rrn*L) and NADH dehydrogenase subunit 1 (*nad*1). 562 bp of the *nad*1 gene was amplified using forward SuiND1_F (5′-CGAGCTTATATAGGTATTTCTCAACG-3′) and reverse SuiND1_R (5′-CGTTGTAGCCTCTTACTAATTCTCTTT-3′) primers while 422 bp of the *rrn*L gene was amplified using primers, forward HPrrnL_F (5'-TGTAAWTCTCCTGCCCAATGA) and reverse TSrrnL_R (5'-CGGTTTAAACTCAAATCACGTA). The PCR amplification conditions were identical for both markers and were conducted in a total volume of 20 µl using 1 µl worm DNA as template. Master mix PCR ingredients were: 2 µl 10X PCR buffer, 0.2 mM of each dNTP, 0.4 mM of each primer pair, 2.0 mM MgCl2, and 1 U of Hot Start DNA-polymerase (Ampliqon). PCR conditions were initial denaturation at 95 °C for 15 min followed by 35 cycles consisting of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 1 min and a final extension at 72 °C for 10 min. Gel electrophoresis was used to verify amplification of a single fragment of the expected size as described above. PCR products were enzymatically cleaned prior to sequencing using 10 µl of PCR product, 1 µl Exonuclease I and 2 µl FastAP Thermosensitive Alkaline Phosphotase (1 U/µl) (Fermentas). The samples were incubated for 15 min at 37 °C followed by 15 min at 85 °C. Finally, cleaned amplicons were sequenced in both directions using same primers as given above by Macrogen Inc. in Seoul, South Korea.

2.5. Genetic variation, differentiation and phylogenetic relationships

Forward and reverse sequences from each sample were checked, edited manually and assembled using vector NTI (Lu and Moriyama, 2004) then trimmed using BioEdit (Hall, 1999). Consensus sequences of 410 bp for *nad*1 gene and 397 bp for *rrn*L gene were used for all the analyses

(Appendix of raw data). Two sequences from the gene bank, one for *T. trichiura* (Accession No: GU385218) and one for *T. suis* (Accession No: GU070737) were included in the dataset. Genetic relatedness and evolutionary relationship were analyzed for each of the two markers. Phylogenetic relationship was inferred using neighbor joining (NJ) and maximum likelihood (ML) phylogenetic trees in MEGA v6.1 (Tamura et al, 2013). The best-to-fit substitution model was identified using jModelTest0.1.1 (Posada, 2008) under Akaik information criterion (AIC) (Akaike, 1974). *Trichinella spiralis* was used as an outgroup (Accession No: AF293969). The most parsimonious network was generated by the Neighbor-net method using SplitsTree v.4.13.1 (Huson and Bryant, 2006). Neighbor-net network can reveal ambiguous and incompatible sites which usually appear as a reticulate structure in the network.

The genetic variation was estimated within the different major clades (splits) identified in the phylogenetic analysis by estimating the nucleotide diversity (pi) and haplotype diversity using software DnaSP v.5 (Librado and Rozas, 2009). The degree of genetic differentiation between the populations of *T. suis* populations was estimated by Analysis of Molecular Variance (AMOVA) analysis. Fixation index (F_{st}) was estimated using the software Arlequin v.3.5.1.2 (Excoffier et al, 2010). Since few samples (n=2) of *T. trichiura* were obtained from China, I could not calculate F_{st} between human *T. trichiura* in China and Uganda. Also, as the worm material from baboon were collected from unnatural habitat (i.e. worms of unknown origin), hence F_{st} calculations was not estimated. However, I calculated pairwise F_{st} for *Trichuris* clades of human and baboon identified in the phylogenetic analysis. The p-distance was calculated between the distinct splits (clades) identified in the phylogenetic analysis using MEGA v.6.1. (Tamura et al, 2013).

2.6. Genetree, effective population size

The genetic diversity of a population Θ (theta) and the ancestral history were estimated for *T. suis* and *T. trichiura* populations from the Genetree (Griffiths and Tavare´ 1994). Theta is the product of mutation rate and effective population size which is an estimate for the genetic variation in a population. This analysis was conducted on concatenated dataset of the two markers. First, sequences were aligned then imported to Map modules in the SNAP workbench (Aylor et al. 2006) to collapse the sequences to haplotypes excluding sites which are indels and infinite site violations. Then, compatibility analysis revealed the incompatible sites which were removed using CladeEx (Aylor et al. 2006). The simulations were repeated 5 times with 10 million runs with different random seeds to enhance the reliability of the inference of the ancestral history.

For mitochondrial DNA, theta (Θ) equals $2\mu N_{eff}$ where N_{eff} is the effective population size for females and μ is is the mutation rate per gene per generation. Effective population size was calculated using this formula as $N_{eff}=\mu/2\Theta$ where Θ was calculated from the Genetree and for μ, the mutation rate of *Coenorhabditis elegans* was used which is 1.6×10^{-7}per site per generation (Denver et al, 2000). To obtain the mutation rate per gene, the mutation rate per site was multiplied by the number of nucleotides used (805 nt for both markers). Different values of Θ were used, namely the median value plus and minus the standard error calculated by Genetree. Genetree gives the time of coalescence between lineages in a coalescent time units (T). Coalescent time was converted to real time (t) in terms of number of generations using formula $t= TN_{eff}$ where N_{eff} is effective population size, T is the coalescent time unit.

2.7. Demographic history and time of divergence.

BEAST v.1.6.1 (Drummond and Rambaut, 2007) was also used to infer the divergence time of *T. suis* and *T. trichiura* populations. BEAST employs Bayesian approaches coupled with Markov Chain Monte Carlo (MCMC) for the coalescent simulations. The algorithms behind BEAST and Genetree are different. Genetree assumes infinite site model (nucleotides mutate only once), while BEAST allows more complicated mutation models (Kuhner, 2008). The concatenated dataset of *nad*1 and *rrn*L markers was used in this analysis as well. The analysis was run with different conditions of mutation patterns namely the strict molecular clock and relaxed molecular clock (log normal and uncorrelated). The final analysis was done using strict molecular clock with a normal distributed substitution rate of $1.6 \times 10^{-7}(\pm 0.3 \times 10^{-7})$ based on value of *C. elegans* estimated by Denver et al (2000). The substitution model used here is Hasegawa-Kishino-Yano (HKY) model with gamma distribution as best to fit model based on Bayesian Inferences Criterion (Kass and Raftery, 1995) in jModelTest0.1.1 (Posada, 2008). Yule prior which is suitable for dataset that combine different species was used as a tree prior with a random starting tree. MCMC chains were run by 10^7 iteration with a burn in value of 1000. Tracer v.1.6 was used to analyze log files and to check whether the MCMC chains were sufficient by recording effective sample size values to be above 200. The three log files of the three independent runs were combined using log combiner v1.6.1 (Drummond and Rambaut, 2007). Tree Annotater v1.6.1. (Drummond and Rambaut, 2007) was used to summarize samples from the posterior on maximum credibility tree and the posterior probability limit set to 0.5. Figtree v1.3.1 (Drummond and Rambaut, 2007) was used to depict the tree.

2.8. Neutrality tests and population size changes

The population size changes were inferred by two methods, namely neutrality tests and mismatch distribution. Neutrality tests used here are Tajima's D (Tajima, 1989) and Fu's F_s (Fu, 1997) tests. Neutrality tests reveal whether there are deviations from the neutrality (i.e. presence of selection) in the sequences and test whether a population had experienced expansion or decline. The theory behind both tests is that when a population expands, rare mutations (singletons) are expected to flourish and will not extinct. Hence, excess of rare alleles will appear as in the populations. On the other hand, when a population is declining, rare alleles are expected to extinct and hence excess of the intermediate mutation and few singletons will appear in the population. Arlequin v.3.5.1.2 (Excoffier et al, 2010) was used to estimate values for the two tests through 10,000 computer simulations based on observed pair-wise differences in samples. For both tests positive values indicate population decline and negative values indicate population expansion.

Second method is the mismatch distribution (the number of pair-wise differences in the sequences). For the populations that had experienced an expansion, the mismatch distribution appears as a unimodel, bell shaped curve. On the other hand, if the population have not experienced any expansions (i.e. stable population), the mismatch distribution curve will be more steadily slope (non-wave like). As the rate of heterogeneity in the mitochondrial DNA is high which might give false indication of population expansion, mismatch distribution may give more accurate estimation if the population was constant or not. The mismatch distribution curves were inferred for the different clades using DnaSP v.5 (Librado and Rozas, 2009). The Chinese *T. trichiura* population was excluded due to low number of samples (n=2) and the *T. suis* population from Uganda as there was no DNA polymorphism within this population. Also the mismatch distribution curve for the *rrn*L gene in the *T. suis* population from Denmark and USA was not calculated as it has no variation.

3. Result

3.1. Phylogenetic and genetic differentiation analyses

ML and NJ trees depicted similar tree topologies for both genes and Figure 1 shows NJ tree of the *nad*1 gene. For ML tree, HKY with gamma distribution model was used as the best substitution model that fitted *nad*1 gene and Tamura 3-parameter model for the *rrn*L gene. It is noteworthy that two samples (one *T. trichiura* from Human in Uganda and one from China) showed doubled peaks in the chromatogram of *nad*1 gene that may indicate co-amplification of nuclear mitochondrial pseudogenes (numts) or homologous sequences in the nuclear genome. Hence, I excluded these samples from the further analyses.

Phylogeographic distribution in *T. suis* populations was observed. However, *T. suis* from Denmark and USA clustered together in the same clade while *T. suis* from Uganda and from China were found in a separate clade. The clades are indicated by the country name of worm origin as shown in Figure 1. However, two worms from Uganda clustered with worms from Denmark and USA. The phylogeographic distribution was also seen for *T. trichiura* recovered from humans as well (*T. trichiura* Africa and *T. trichiura* China clades). The majority of the baboon worms were clustered with the human *T. trichiura* from Africa. Meanwhile, few baboon samples (7 individuals) were grouped in different clade which was indicated as '*Trichuris* baboon' clade. Only single baboon worm from USA was clustered in the '*T. trichiura* China' clade. The Neighbor-net network (Figure 2) identified splits corresponds to the clades in the NJ tree as shown in Figure 1.

The genetic distance between and within each clade are given in Table 3 and 4 for the *nad*1 and *rrn*L genes respectively. The AMOVA analysis is summarized in Table 5 where the pairwise F_{st} estimations are shown for the two markers between the populations of *T. suis*. The F_{st} between clades *T. trichiura* Africa and *Trichuris* baboon were 0.3625 for *rrn*L and 0.4712 for *nad*1 with statistical significant p values ($P<0.05$).

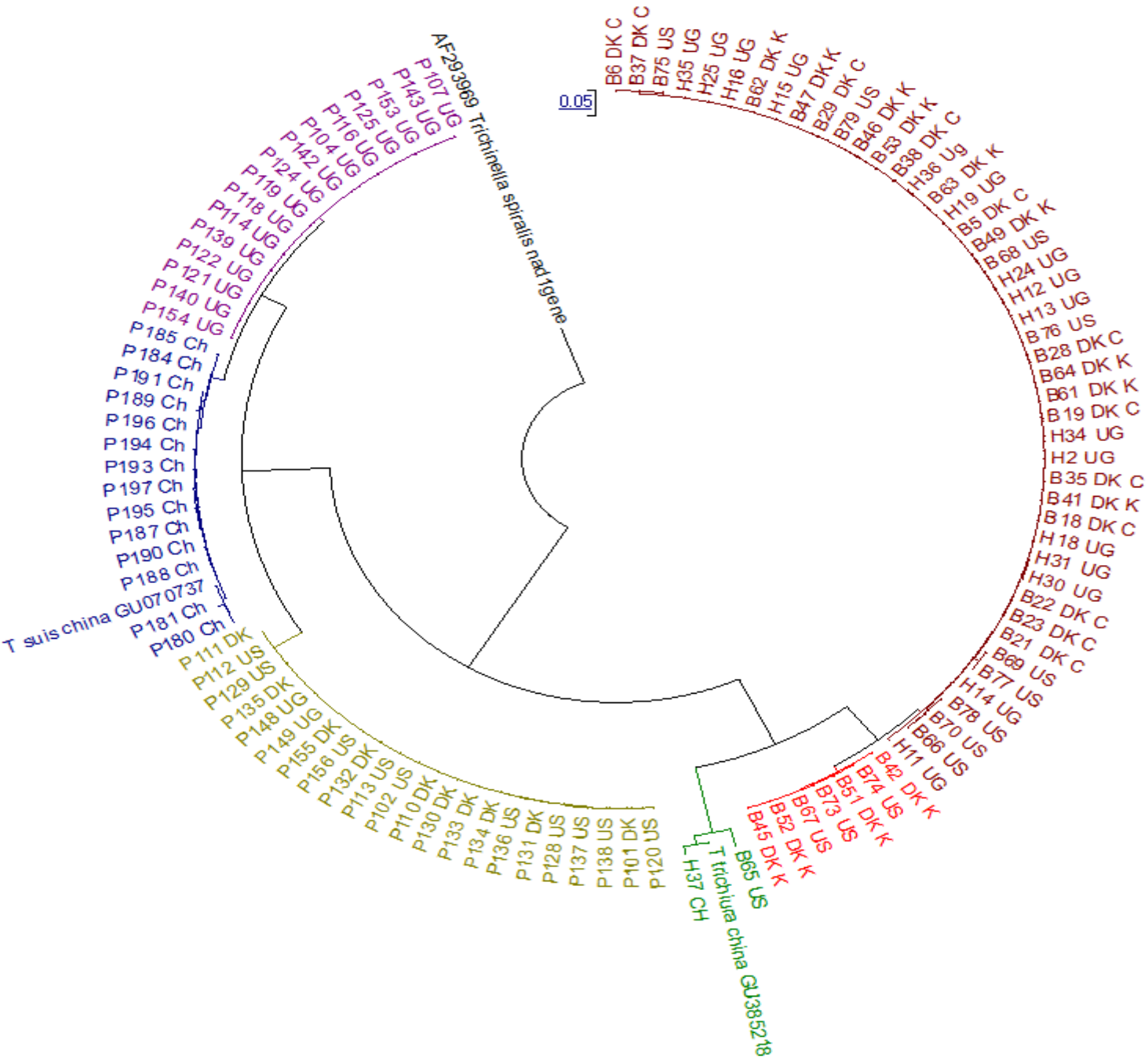

Figure 1. NJ tree for the *nad*1 gene. Six clades were identified and are indicated by different colors. *Trichuris trichiura* from humans from Uganda clustered in one clade where majority of *Trichuris* from baboons clustered in same clade named as '*T.trichiura* Africa' and indicated by maroon color ( ). Only 7 *Trichuris* individuals from baboons clustered in anther clade named as *Trichuris* baboon and indicated by red color ( ). *T. trichiura* from China were distinct and was indicated by green color clade( ). The other three clades are for *T. suis*. *T. suis* populations from USA and Denmark clustered in *'T. suis* DK&US' clade indicated by olive green color ( ), *T. suis* from China *'T. suis* China' clade is indicated by blue ( ) and *T. suis* from Uganda *'T. suis* Uganda' clade is indicated by purple ( ). Samples were named by the host (the first letter) then the number of the worm and then the country of origin (last letters). The Key for the samples name is: B: Baboon, H: Human, P: Pigs; US, USA, DK, Denmark (C for Copenhagen Zoo and K for Knuttenborg), Ch, China, UG, Uganda.

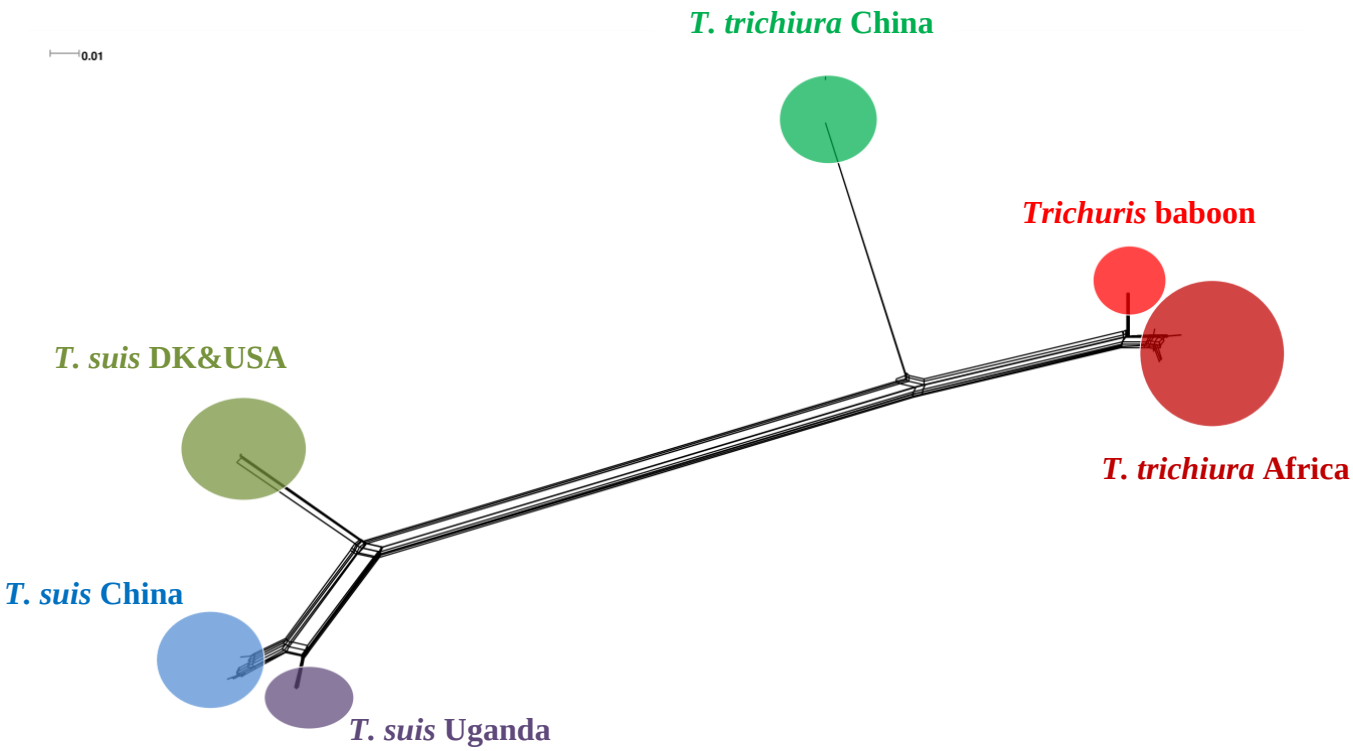

Figure 2. Neighbor-net network for all populations. The colors of the different populations are the same as used for the NL tree above and with the names of clades given in Figure 1.

The nucleotide diversity and haplotype diversity were calculated for the different clades for the two markers identified by phylogenetic tree and network as shown in Table 2. The genetic distances between the different clades were shown in Table 3 for *nad*1 and Table 4 for *rrn*L makers.

Table 2. The nucleotide diversity and haplotype diversity for the *nad*1 and *rrn*L genes for the different clades identified by the phylogenetic tree and network.

***nad*1 gene**	**Number of samples (host)**	**Nucleotide diversity (π)**	**Haplotype diversity (number of haplotypes)**	**GenBank sequences**
***T. trichiura* Africa**	46 (baboons and humans)	0.0121	0.589 (13)	
***Trichuris* baboon**	7 (baboons)	0.0007	0.286 (2)	
***T. trichiura* China**	3 (humans and baboon)	0.0146	1.000(3)	GU385218
***T. suis* DK & US**	22 (pigs)	0.0018	0.579 (5)	
***T. suis* China**	15 (pigs)	0.0100	0.962 (12)	GU070737
***T. suis* Uganda**	16 (pigs)	0.0000	0.000 (1)	
***rrn*L gene**	Number of samples (host)	Nucleotide diversity (π)	Haplotype diversity (number of haplotypes)	
***T. trichiura* Africa**	46 (baboons and humans)	0.0013	0.439 (8)	
***Trichuris* baboon**	7 (baboons)	0.0058	0.900 (4)	
***T. trichiura* China**	3 (humans and baboon)	0.0026	1.000(3)	GU385218
***T. suis* DK & US**	22 (pigs)	0.0000	0.000 (1)	
***T. suis* China**	15 (pigs)	0.0035	0.818 (6)	GU070737
***T. suis* Uganda**	16 (pigs)	0.0000	0.000 (1)	

Table 3. p-distance between the different clades and within each clade for the *nad*1 gene

	***T. trichiura* Africa**	***Trichuris* baboon**	***T. trichiura* China**	***T. suis* DK&USA**	***T. suis* China**	***T. suis* Uganda**
***T. trichiura* Africa**	0.003					
***Trichuris* baboon**	0.028	0.001				
***T. trichiura* China**	0.246	0.249	0.015			
***T. suis* DK & USA**	0.448	0.457	0.482	0.002		
***T. suis* China**	0.481	0.477	0.508	0.135	0.013	
***T. suis* Uganda**	0.446	0.440	0.506	0.126	0.040	0.00

Table 4. p-distance of the *rrn*L gene between and within different clades

	***T. trichiura* Africa**	***Trichuris* baboon**	***T. trichiura* China**	***T. suis* DK&USA**	***T. suis* China**	***T. suis* Uganda**
***T. trichiura* Africa**	0.001					
***Trichuris* baboon**	0.016	0.005				
***T. trichiura* China**	0.083	0.086	0.001			
***T. suis* DK & USA**	0.204	0.208	0.215	0.000		
***T. suis* China**	0.205	0.209	0.215	0.027	0.004	
***T. suis* Uganda**	0.204	0.208	0.215	0.022	0.010	0.00

Table 5. Pairwise F_{st} calculations between different populations of *T. suis* (*nad1* gene below the diagonal and *rrn*L gene above the diagonal). The asterisk (*) indicate statistical significant value (p<0.05).

	T. suis DK	*T. suis* Uganda	*T. suis* China	*T. suis* USA
T. suis DK		0. 9164*	0.5421*	0.0000
T. suis Uganda	0. 9967 *		0. 5422*	0.9164 *
T. suis China	0. 9417*	0.8762 *		0. 5421*
T. suis USA	0.0156	0.9946*	0.9394*	

Genetree, effective population size

The Genetree figures of *T. suis* and *T. trichiura* are given in supplementary data (Figure S2 and S3). For *T. suis*, the estimated Θ was 20.7 ± 6.68. There was no ancestral polymorphism detected between populations of DK&USA with the populations of Uganda and China. However, there was many ancestral polymorphism between population of China and population of Uganda indicating recent divergence (Supplementary data, Figure S2). On the other hand, for *T. trichiura* there was no ancestral polymorphism between Chinese and Ugandan populations. While many shared mutations were detected for the populations of *Trichuris* from baboons and humans from Africa (Supplementary Figure S3). The estimated Θ was 126.4 ± 32.3 for all *Trichuris* populations from humans and baboons.

The effective population size of the ancestral populations of *T. suis* and *T. trichiura* was estimated from the formula $\Theta=2N_{eff}\mu$. The highest possible and lowest possible values of Θ (mean value plus or minus the standard deviation) were used and the mutation rate (μ) is 1.29×10^{-4}per gene per generation was used. The ancestral effective population size (N_{eff}) for *T. suis* females is estimated and approximated to be 80,000 with upper estimate of 110,000 and lower estimate 60,000. Similarly, for *T. trichiura*, N_{eff} for females was estimated to be around 500,000 with an upper estimate of 640,000 and lower estimate of 340,000. Therefore, the coalescent time between all populations of *T. trichiura* (time to the recent common ancestor) was estimated to be 500,000 (640,000-340,000) generations ago. Similarly, the coalescent time for *T. suis* populations is 80,000 (110,000-60,000) generations ago.

Demographic history and time of divergence

The time of divergence was estimated from Genetree and Bayesian analyses. Both tests gave similar results for *T. trichiura*. The divergence time for the *T. trichiura* populations of China and Africa populations is estimated to about 530,000 generation ago in BEAST. However, there was difference in the estimations of *T. suis* between the two analyses. For *T. suis*, there are two divergence events. The ancient one is the split the populations of DK&USA and the populations of Uganda and China that happened around 80,000 (110,000-60,000) generations ago based on the Genetree while the more recent divergence between *T. suis* populations of Uganda and China is around 0.4 in coalescent time unit (Supplmentary data, Figure S2) which is 32,000 (44,000-24,000) generations ago. In BEAST, the estimations were nearly twice as much as Genetree. The ancient divergence between DK&USA population and the Chinese/Ugandan populations is around 240,000 generations ago and 90,000 generations ago for divergence between populations of China and Uganda as shown in Figure 3.

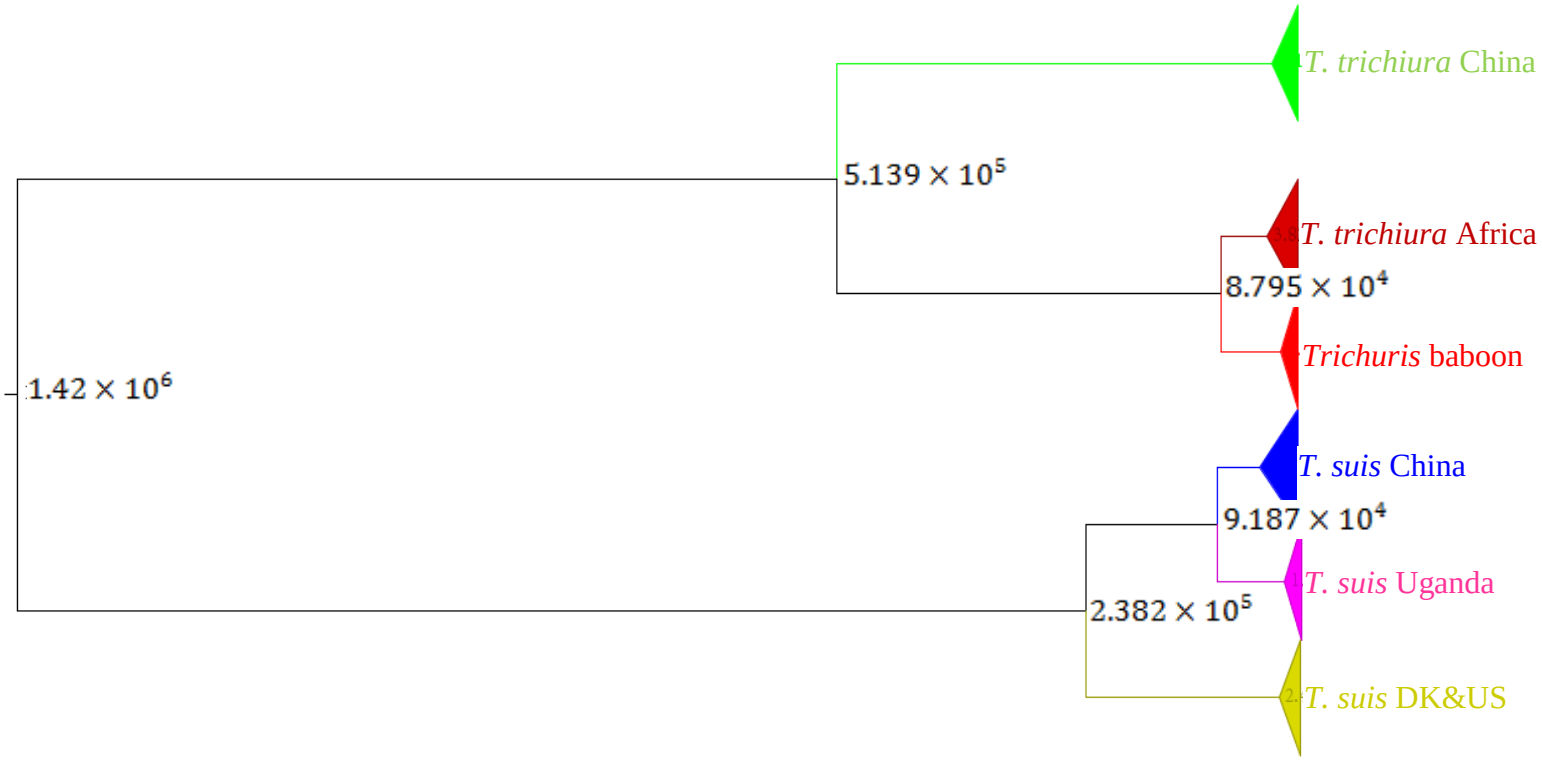

Figure 3. Bayesian analysis on BEAST where the different branches colors are the same as used in the phylogenetic tree (see Figure 1). The time of divergence is given at each node by the number of generations.

Neutrality tests and population size changes

Tajima's D and Fu's F values are shown in Table 6 with their correspondent p-values. Also the pairwise mismatch distribution curves were evaluated. Both Tajima's D and Fu's F tests gave negative results for all the populations tested except for one positive value for Tajima's D in the *T. suis* population of China estimate non-significant value. Also, not all the values were statistically significant (Table 6). The statistically significant negative values may indicate population expansion.

Table 6. A summary values of neutrality tests with statistical significance p values beside. (*) indicates statistically significant p value (p<0.05).

	***nad*1 gene**				***rrn*L gene**			
	Tajima's D	P-value	Fu's F	P-value	Tajima's D	P-value	Fu's F	P-value
***T. trichiura* Africa**	-2.765	0.000*	-25.538	0.000*	-1.804	0.011*	-28.525	0.0000*
***Trichuris* baboon**	-1.006	0.231	-14.742	0.000*	-0.562	0.402	-2.862	0.0098*
***T. suis* DK&USA**	-1.431	0.061	-340	0.000*	No polymorphism found			
***T. suis* China**	-1.423	0.066	-11.658	0.000*	0.367	0.684	-18.248	0.0000*

On the other hand, the mismatch distribution curves did not support stable population in all the tested populations (Supplementary data, Figure S1). The population of *T. suis* from China is the only population that showed expansion for both markers.

4. Discussion

One aim of this study was to investigate a hypothesis about *Trichuris* spp. transmission and demographic history in the humans and pigs which is that the *Trichuris* sp. infecting humans or pigs was transmitted to the other upon domestication in the Neolithic era 10,000 years ago. The strong differentiation of the different *T. suis* populations in Uganda, China and Denmark/USA reveals the incorrectness of this hypothesis. Moreover, the demographic history of *T. suis* was found to be different from *T. trichiura* as will be discussed later.

The coalescent analysis indentified two divergence events of *T. suis* populations. The ancient divergence between DK&USA populations and the Chinese/Ugandan populations is around 80,000-200,000 generations ago while recent divergence between populations of China and Uganda 24,000-90,000 generations ago. Remarkably, the recent divergence of *T. suis* population between China and Uganda is supported by studies done on the genetic relatedness of domesticated pig genetics (Amills et al, 2012). Common alleles between the domesticated pigs of Far East and East African were identified revealing close relationship between both with two possible explanations for this (Ramirez et al, 2009; Amills et al, 2012). First, these pigs were brought from Far East by the European trading routes with Far East during the colonization of Africa (few hundreds of years ago). Secondly, the introduction of domesticated pigs from the Far East could be due to the trading between the ancient civilizations in Africa and Far East or the settlement of Austronesian peoples in East Africa nearly seven thousands of years ago (Amills et al, 2012). Given maximum possible number of generations for *T. suis* per year (assumed as 4 generations/year), divergence between Chinese and Uganda *T. suis* population might have had happened 6,000-22,000 years ago. Hence, it is unlikely that the introduction of *T. suis* done through European intermediary few hundred years ago and it is likely that the introduction of *T. suis* in East Africa done by introduction of domesticated pigs from Far East thousands of years ago. However, this could not rule out that there may be other waves of recent introduction of *T. suis* in domesticated pigs from Far East. Recent study on different populations of *Ascaris* also showed that *A. lumbricoides* infecting humans from Zanzibar and Bangladesh are very closely related (Betson et al, 2014) hinting to the potential of this route from Far East and Indian subcontinent to East Africa in the parasites transmission.

I found no polymorphism in the *T. suis* population in Uganda in both markers which could be due to either founder effect (the establishment of a new population from individuals derived from a much larger population) or selective sweep (spread of beneficial mutation from a single individual to the entire population) or combination of both factors. Such selective sweep might have had happened due to new adaptations of the host physiology in the new environment or due to recent bottleneck in pigs that eliminate the parasite molecular variation (Barret et al, 2008).

The common haplotypes shared between domestic pigs of Far East and East Arica were found to be nearly absent from European domestic pigs (Amills et al, 2012). Additionally, it was found that the domestic pigs in China and Europe are genetically very distant (Giuffra et al, 2000). That could explain the big genetic distinctiveness of *T. suis* populations in Denmark and the populations of China. However, the divergence time was estimated to be roughly a million years ago between the pigs in China and Europe (Groenen et al, 2012) which is much older than the divergence time estimated of Chinese and Danish *T. suis* populations (highest estimate is around 240,000 years ago given 1 generation per year). Since the estimations of the divergence time relies on the calculations to the most recent common ancestor (MRCA) which is actually the time to the most recent bottleneck (Archie et al, 2008). Hence, there could be bottlenecks of the *T. suis* populations in pigs which is also coincided with the bottleneck of the pigs during migration (Groenen et al, 2012) and therefore the divergence time appears more recent than divergence of the host. Another possibility is that it could be that the mitochondrial markers used were under negative selection pressure which make the divergence more recent that it should be (Archie et al, 2008). However, It is noteworthy that the genetic distance between *T. suis* population in China and Denmark is high (around 12% in the *nad*1 gene) considering intra-species differences (Blouin et al, 2002) which suggest possible presense of different cryptic *Trichuris* species infecting pigs in these areas.

No genetic distinction between the *T. suis* populations of Denmark and USA as both populations clustered in same clade. Moreover, the low F_{st} (0.0156) indicated undifferentiated populations. Since the domestic pig in USA are of European and were first brought with the Spanish explorer Hernando de Solo only few hundred years ago, that could explain the lack of differentiation between USA and Denmark *T. suis* populations (More about Pigs, 2009). This was also consistent with another parasite in domestic pigs, *Trichinella spiralis*, which was found to be introduced to Americas by Europeans (Rosenthal et al, 2008). Two *T. suis* worms from Uganda were clustered with the DK&US clade which could be due to recent pig transport of European breeds to Uganda.

Given the previous natural history, it is obvious that the human activity through trade routes had contributed significantly in the transmission of *T. suis* in new areas (such as America and East Africa) through pig transport. The human movements and commercial trade was found to contribute to the translocations of several other parasitic nematodes with their hosts (Gilabert and Wasmuth, 2013). Moreover, the pattern of evolutionary relationships between the *T. suis* populations infecting domesticated pigs may have had coincided with the host genetics and demography. Parasites with high specificity were suggested to act as proxies for their host genealogy due to their dependency of the host and their shorter generation time (Neiberding and olivieri, 2007). It is therefore suggested that *T. suis* may provide valuable inferences about the migration and demography of pigs.

The mismatch distribution curves and the neutrality tests support the presence of a population expansion happened in the *T. suis* population of China which is in concordant with other parasitic nematodes of livestock (Mes et al, 2003; Archie and Ezenwa, 2011). The population expansion might have had happened due to domestication of pigs in China nearly 9000 years ago (Giuffra et al, 2000). However, although there was a domestication of pigs in Europe as well, the mismatch distribution curve did not support population expansion in the DK&US clade in the *nad*1 gene. However, there are several reasons for not detecting demographic expansion in the mitochondrial genes such as the marker might not be neutral or the demographic expansion might follow linear or instantaneous growth rather than exponential growth (Morrison and Hoglund, 2005), thereofore, this finding should be dealt with caution. Other segregated neutral markers of different origin (e.g. nuclear) should be analyzed for better understanding of the changes in population sizes.

The majority of the baboon worms were clustered with the human worms from Africa (Figure 1). Since Olive and yellow baboons (*Papio anubis* and *P. cynocephalus*) are indigenous species in Africa (Cawthon Lang, 2006), this could explain the clustering of many baboon *Trichuris* with the human *T. trichiura* from Africa (*T. trichiura* Africa clade) suggesting a high gene flow between the *Trichuris* populations from baboons and humans in Africa. This is in concordance with Hansen et al (2014) and another study on *Trichuris* from hamadryas baboon (*P. hamadryas*) where worms were found to be closely related to *T. trichiura* from humans in Czech Republic (Callejon et al, 2013). Baboons therefore might act as zoonotic reservoir for *T. trichiura* infections to humans. Seven individuals from baboons from both Denmark and USA clustered in a separate clade (*Trichuris* baboon clade). The *Trichuris* baboon clade is genetically very close to the *T. trichiura* Africa clade (average genetic distance 1.5%) suggesting to be the same species (i.e. *T. trichiura*), however, genetically differentiated (high F_{st}) which may indicate presence of different populations of *T.*

trichiura infecting baboons in their native habitat in Africa. Remarkably, a single baboon worm from USA was clustered with the *T. trichiura* China clade unlike the rest of the baboon *Trichuris* worms. However, similar observation was found in *Trichuris* from Chacma baboon (*P. ursinus*) from South Africa which was closely related to *T. trichiura* from human in China based on ITS markers (Ravasi et al ,2012). One caveat in this study is is the lack of samples from baboons from their natural habitats, hence, the real estimate of the zoonotic potential of *Trichuris* from baboons cannot be properly evaluated. Future studies on native, sympatric habitat of baboons and humans could reveal the nature of cross transmission between the two hosts. Interestingly, the genetic distance between the human *T. trichiura* populations in Africa and China was very high (average 16%) considering intra-species variation (Blouin, 2002) indicating the possible presence of different cryptic (hidden) species of *T. trichiura* infecting humans. The effective population size of *T. trichiura* was nearly 5 times higher than *T. suis* population which could be explained by the potential of *T. trichiura* to infect more than one host species (human and non human primates) or that *T. trichiura* consists of a different cryptic species as referred above or combination of both factors.

The presence of different cryptic species of *T. trichiura* infecting humans migh be very important in order to implement suitable control strategies. For instance, different cryptic species of *Opisthorchis viverrini* (a food born trematode that infects humans) in different localities (Laos and Thailand) were found to have a significant difference in the excretion of eggs/gm/worm (Saijuntha et al, 2007). Moreover, different cryptic species might have different drug resistance potential especially since a single nucleotide polymorphism (SNP) associated with the drug resistance in the beta tubulin gene was detected in *T. trichiura* (Diawara et al, 2009). Hence, the control and treatment in different areas may not be equally effective and therefore detection of the speceis diversity is crucial to compare the pathology, epidemiology and drug resistence potential between different cryptic species (Nadler and Leon, 2012).

The divergence time between the Chinese *T. trichiura* populations and African populations was around 500,000 generations ago in Genetree and BEAST which is 500,000-160,000 years ago given the maximum and minimum number of generations (assumed as 1-3 generations per year). Given the assumption of the African origin of modern human (*Homo sapiens*) and its migration to South East Asia 60,000-80,000 years ago (Beyin, 2011) which is more recent than my lowest estimate 160,000 years ago, two possibilities may explain the migration of *T. trichiura* from Africa to South East Asia. First possibility is that one of the early human ancestor species (e.g. *H. erectus*) is the one responsible for the transmission through the migrations out of Africa (Finlayson, 2005). This might

be true especially with the potential of *T. trichiura* to infect primates generally. The other possibility is that the calculations of the maximum number of generations per year or mutation rate may not accurate since I used the mutation rate of free living nematode (*C. elegance*) that could be different from *Trichuris* sp. Nonetheless, regardless of the exact date of migration of *T. trichiura* from Uganda to China, it is different from the finding of the recent divergence of *T. suis* populations between the two regions revealing different demographic history of *T. trichiura* and *T. suis*.

Recently, Betson et al (2014) investigated the transmission pattern of more than 500 *A. suum* and *A. lumbricoides* worms recovered from pigs and humans, respectively and found very limited genetic distance in the *cox*1 gene between worms from the two host species. Moreover, it was found that the *Ascaris* infections in humans in Europe were of pig origin and in the study by Zhou et al (2012) up to ~14% of the worms in humans were *A. suum* which emphasize the zoonotic potential *A. suum* and its close relationship to *A. lumbricoides*. Hence, in contrast to *Trichuris* in pigs and humans, *Ascaris* might have been transmitted from one host to another upon the domestication of pigs as suggested by Betson et al (2014). Since, *Ascaris* and *Trichuris* in humans and pigs share several characteristics including direct life cycle and mode of infection but having different evolutionary history, comparing their transcriptomes and genomes may give valuable insights on the evolution of parasitism i.e. which genes helped *Ascaris* to adapt to the new host during host switching. A similar study was conducted on the transcriptome of *A. suum*, *Haemonchus contortus* and the free living nematode *C. elegance* which provided valuable insights on the genes involved in the parasitism (Yin et al, 2008). Moreover, the comparative study between *T. trchiura* and *T. suis* showed big difference between total number of gene between the two species indicating high adaptation of the two species for their respective hosts (Foth et al, 2014, Ghedin, 2014, Jex et al 2014). Similar study on *A. suum* and *A. lumbricoides* could also give insights on the host-parasite relationship and the degree of adaptation.

Conclusion and perspectives

In summary, it was found that the speciation between *T. trichiura* and *T. suis* happened many years before the domestication of pigs in the Neolithic era 10,000 years ago. *T. suis* was found to have independent demographic history from *T. trichiura* with anthropogenic factors contributed to the spread of the parasite associated with the pig transport. Remarkably, I found large genetic distance between the two *T. trichiura* infecting humans from Uganda and China indicating possible different species which might have important implications on establishing proper control measures. Further studies using nuclear markers (e.g. ITS) on the *T. trichiura* in humans are warranted to confirm the presense of *T. trichiura* species complex infecting humans. *Trichuris* from baboons was found to be belonging to the same *Trichuris* species complex that infects humans hinting to the possible zoonotic infections from baboons to humans. Further studies including worms from sympatric areas where baboons and humans lives in close contact could further assess the taxonomic relationship between *Trichuris* in the two host species and the zoonotic potential.

References

Akaike H (1974). "A new look at the statistical model identification." IEEE T Automat Contr, (19):16–723.

Amills, M., Ramírez, O., Galman-Omitogun, O., and Clop, A. (2013). "Domestic pigs in Africa." Afr Archaeol Rev 30(1).

Archie, E. A. and V. O. Ezenwa (2011). "Population genetic structure and history of a generalist parasite infecting multiple sympatric host species." Int J Parasitol 41(1): 89-98.

Archie, E. A., G. Luikart and V. O. Ezenwa (2008). "Infecting epidemiology with genetics: a new frontier in disease ecology." Trends Ecol Evol 24(1): 21-30.

Aylor, D. L., E. W. Price and I. Carbone (2006). "SNAP: Combine and Map modules for multilocus population genetic analysis." Bioinformatics 22(11): 1399-1401.

Barrett, L. G., P. H. Thrall, J. J. Burdon and C. C. Linde (2008). "Life history determines genetic structure and evolutionary potential of host-parasite interactions." Trends Ecol Evol 23(12): 678-685.

Betson, M., P. Nejsum, R. P. Bendall, R. M. Deb and J. R. Stothard (2014). "Molecular Epidemiology of Ascariasis: A Global Perspective on the Transmission Dynamics of *Ascaris* in People and Pigs." J Infect Dis.

Beyin, A. (2011). "Upper Pleistocene Human Dispersals out of Africa: A Review of the Current State of the Debate." Int J Evol Biol 2011: 615094.

Blouin, M. S. (2002). "Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer." Int J Parasitol 32(5): 527-531.

Callejon, R., S. Nadler, M. De Rojas, A. Zurita, J. Petrasova and C. Cutillas (2013). "Molecular characterization and phylogeny of whipworm nematodes inferred from DNA sequences of *cox*1 mitochondrialDNA and 18S rDNA." Parasitol Res 112(11): 3933-3949.

Cawthon Lang KA.(2006). Primate Factsheets: Olive baboon (*Papio anubis*) Taxonomy, Morphology, & Ecology. Available from <http://pin.primate.wisc.edu/factsheets/entry/olive_baboon>.

Cutillas, C., M. de Rojas, A. Zurita, R. Oliveros and R. Callejon (2014). "*Trichuris colobae* n. sp. (Nematoda: Trichuridae), a new species of *Trichuris* from *Colobus guereza kikuyensis*." Parasitol Res 113(7): 2725-2732.

Denver, D. R., K. Morris, M. Lynch, L. L. Vassilieva and W. K. Thomas (2000). "High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*." Science 289(5488): 2342-2344.

Diawara, A., L. J. Drake, R. R. Suswillo, J. Kihara, D. A. Bundy, M. E. Scott, C. Halpenny, J. R. Stothard and R. K. Prichard (2009). "Assays to detect beta-tubulin codon 200 polymorphism in *Trichuris trichiura* and *Ascaris lumbricoides*." PLoS Negl Trop Dis 3(3): e397.

Drummond, A. J. and A. Rambaut (2007). "BEAST: Bayesian evolutionary analysis by sampling trees." BMC Evol Biol 7: 214.

Excoffier, L. and H. E. Lischer (2010). "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." Mol Ecol Resour 10(3): 564-567.

Finlayson, C. (2005). "Biogeography and evolution of the genus Homo." Trends Ecol Evol 20(8): 457-463.

Foth, B. J., I. J. Tsai, A. J. Reid, A. J. Bancroft, S. Nichol, A. Tracey, N. Holroyd, J. A. Cotton, E. J. Stanley, M. Zarowiecki, J. Z. Liu, T. Huckvale, P. J. Cooper, R. K. Grencis and M. Berriman (2014). "Whipworm genome and dual-species transcriptome analyses provide molecular insights into an intimate host-parasite interaction." Nat Genet 46(7): 693-700.

Fu, Y. X. (1997). "Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection." Genetics 147(2): 915-925.

Ghedin, E. (2014). "Panning for molecular gold in whipworm genomes." Nat Genet 46(7): 661-663.

Gilabert, A. and J. D. Wasmuth (2013). "Unravelling parasitic nematode natural history using population genetics." Trends Parasitol 29(9): 438-448.

Giuffra, E., J. M. Kijas, V. Amarger, O. Carlborg, J. T. Jeon and L. Andersson (2000). "The origin of the domestic pig: independent domestication and subsequent introgression." Genetics 154(4): 1785-1791.

Griffiths, R. C. and S. Tavare (1994). "Sampling theory for neutral alleles in a varying environment." Philos Trans R Soc Lond B Biol Sci 344(1310): 403-410.

Groenen, M. A., A. L. Archibald, H. Uenishi, C. K. Tuggle, Y. Takeuchi, M. F. Rothschild, C. Rogel-Gaillard, C. Park, D. Milan, H. J. Megens, S. Li, D. M. Larkin, H. Kim, L. A. Frantz, M. Caccamo, H. Ahn, B. L. Aken, A. Anselmo, C. Anthon, L. Auvil, B. Badaoui, C. W. Beattie, C. Bendixen, D. Berman, F. Blecha, J. Blomberg, L. Bolund, M. Bosse, S. Botti, Z. Bujie, M. Bystrom, B. Capitanu, D. Carvalho-Silva, P. Chardon, C. Chen, R. Cheng, S. H. Choi, W. Chow, R. C. Clark, C. Clee, R. P. Crooijmans, H. D. Dawson, P. Dehais, F. De Sapio, B. Dibbits, N. Drou, Z. Q. Du, K. Eversole, J. Fadista, S. Fairley, T. Faraut, G. J. Faulkner, K. E. Fowler, M. Fredholm, E. Fritz, J. G. Gilbert, E. Giuffra, J. Gorodkin, D. K. Griffin, J. L. Harrow, A. Hayward, K. Howe, Z. L. Hu, S. J. Humphray, T. Hunt, H. Hornshoj, J. T. Jeon, P. Jern, M. Jones, J. Jurka, H. Kanamori, R. Kapetanovic, J. Kim, J. H. Kim, K. W. Kim, T. H. Kim, G. Larson, K. Lee, K. T. Lee, R. Leggett, H. A. Lewin, Y. Li, W. Liu, J. E. Loveland, Y. Lu, J. K. Lunney, J. Ma, O. Madsen, K. Mann, L. Matthews, S. McLaren, T. Morozumi, M. P. Murtaugh, J. Narayan, D. T. Nguyen, P. Ni, S. J. Oh, S. Onteru, F. Panitz, E. W. Park, H. S. Park, G. Pascal, Y. Paudel, M. Perez-Enciso, R. Ramirez-Gonzalez, J. M. Reecy, S. Rodriguez-Zas, G. A. Rohrer, L. Rund, Y. Sang, K. Schachtschneider, J. G. Schraiber, J. Schwartz, L. Scobie, C. Scott, S. Searle, B. Servin, B. R. Southey, G. Sperber, P. Stadler, J. V. Sweedler, H. Tafer, B. Thomsen, R. Wali, J. Wang, J. Wang, S. White, X. Xu, M. Yerle, G. Zhang, J. Zhang, J. Zhang, S. Zhao, J. Rogers, C. Churcher and L. B. Schook (2012). "Analyses of pig genomes provide insight into porcine demography and evolution." Nature 491(7424): 393-398.

Hall, T.A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." Nucl. Acids. Symp. Ser. 41: 95-98.

Hansen, T. V., S. M. Thamsborg, A. Olsen, R. K. Prichard and P. Nejsum (2013). "Genetic variations in the beta-tubulin gene and the internal transcribed spacer 2 region of *Trichuris* species from man and baboons." Parasit Vectors 6(1): 236.

Huson, D. H. and D. Bryant (2006). "Application of phylogenetic networks in evolutionary studies." Mol Biol Evol 23(2): 254-267.

Jex, A. R., P. Nejsum, E. M. Schwarz, L. Hu, N. D. Young, R. S. Hall, P. K. Korhonen, S. Liao, S. Thamsborg, J. Xia, P. Xu, S. Wang, J. P. Scheerlinck, A. Hofmann, P. W. Sternberg, J. Wang and R. B. Gasser (2014). "Genome and transcriptome of the porcine whipworm *Trichuris suis*." Nat Genet 46(7): 701-706.

Kuhner, M. K. (2008). "Coalescent genealogy samplers: windows into population history." Trends Ecol Evol 24(2): 86-93.

Kass R, Raftery A. (1995). "Bayes factors." J Am Stat Assoc. 90:773–795.

Librado, P. and J. Rozas (2009). "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data." Bioinformatics 25(11): 1451-1452.

Liu, G. H., R. B. Gasser, P. Nejsum, Y. Wang, Q. Chen, H. Q. Song and X. Q. Zhu (2013). "Mitochondrial and nuclear ribosomal DNA evidence supports the existence of a new *Trichuris* species in the endangered francois' leaf-monkey." PLoS One 8(6): e66249.

Lu, G. and E. N. Moriyama (2004). "Vector NTI, a balanced all-in-one sequence analysis suite." Brief Bioinform 5(4): 378-388.

Mes, T. H. (2003). "Demographic expansion of parasitic nematodes of livestock based on mitochondrial DNA regions that conflict with the infinite-sites model." Mol Ecol 12(6): 1555-1566.

Molnar, R. I., G. Bartelmes, I. Dinkelacker, H. Witte and R. J. Sommer (2011). "Mutation rates and intraspecific divergence of the mitochondrial genome of *Pristionchus pacificus*." Mol Biol Evol 28(8): 2317-2326.

More about Pigs, The underestimated animal (2009). Available from <http://www.humanesociety.org/animals/pigs/pigs_more.html>

Morrison, D. A. and J. Hoglund (2005). "Testing the hypothesis of recent population expansions in nematode parasites of human-associated hosts." Heredity (Edinb) 94(4): 426-434.

Nadler, S. A. and Leon D. E. (2012). "Integrating molecular and morphological approaches for characterizing parasite cryptic species: implications for parasitology." Parasitology 138(13): 1688-1709.

Nieberding, C. M. and I. Olivieri (2007). "Parasites: proxies for host genealogy and ecology?" Trends Ecol Evol 22(3): 156-165.

Nissen, S., A. Al-Jubury, T. V. Hansen, A. Olsen, H. Christensen, S. M. Thamsborg and P. Nejsum (2012). "Genetic analysis of *Trichuris suis* and *Trichuris trichiura* recovered from humans and pigs in a sympatric setting in Uganda." Vet Parasitol 188(1-2): 68-77.

Posada, D. (2008). "jModelTest: phylogenetic model averaging." Mol Biol Evol 25(7): 1253-1256.

Pullan, R. L., J. L. Smith, R. Jasrasaria and S. J. Brooker (2014). "Global numbers of infection and disease burden of soil transmitted helminth infections in 2010." Parasit Vectors 7: 37.

Ramesh, A., S. T. Small, Z. A. Kloos, J. W. Kazura, T. B. Nutman, D. Serre and P. A. Zimmerman (2012). "The complete mitochondrial genome sequence of the filarial nematode *Wuchereria bancrofti* from three geographic isolates provides evidence of complex demographic history." Mol Biochem Parasitol 183(1): 32-41.

Ramirez, O., A. Ojeda, A. Tomas, D. Gallardo, L. S. Huang, J. M. Folch, A. Clop, A. Sanchez, B. Badaoui, O. Hanotte, O. Galman-Omitogun, S. M. Makuza, H. Soto, J. Cadillo, L. Kelly, I. C. Cho, S. Yeghoyan, M. Perez-Enciso and M. Amills (2009). "Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds." Mol Biol Evol 26(9): 2061-2072.

Ravasi, D. F., M. J. O'Riain, F. Davids and N. Illing (2012). "Phylogenetic evidence that two distinct *Trichuris* genotypes infect both humans and non-human primates." PLoS One 7(8): e44187.

Roepstorff, A., H. Mejer, P. Nejsum and S. M. Thamsborg (2011). "Helminth parasites in pigs: new challenges in pig production and current research highlights." Vet Parasitol 180(1-2): 72-81.

Roepstorff, A., O. Nilsson, A. Oksanen, B. Gjerde, S. H. Richter, E. Ortenberg, D. Christensson, K. B. Martinsson, P. C. Bartlett, P. Nansen, L. Eriksen, O. Helle, S. Nikander and K. Larsen (1998). "Intestinal parasites in swine in the Nordic countries: prevalence and geographical distribution." Vet Parasitol 76(4): 305-319.

Rosenthal, B. M., G. LaRosa, D. Zarlenga, D. Dunams, Y. Chunyu, L. Mingyuan and E. Pozio (2008). "Human dispersal of *Trichinella spiralis* in domesticated pigs." Infect Genet Evol 8(6): 799-805.

Saijuntha, W., Sithithaworn, P., Wongkham, S., Laha, T., Pipitgool, V., Tesana, V., Chilton, N. B., Petney, T. N. and Andrews, R. H. (2007). "Evidence of a species complex within the foodborne trematode *Opisthorchis viverrini* and possible co-evolution with their first intermediate hosts." Int J Parasitol 37: 695–703

Solaymani-Mohammadi, S., I. Mobedi, M. Rezaian, J. Massoud, M. Mohebali, H. Hooshyar, K. Ashrafi and M. B. Rokni (2003). "Helminth parasites of the wild boar, *Sus scrofa*, in Luristan province, western Iran and their public health significance." J Helminthol 77(3): 263-267.

Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Genetics 123(3): 585-595.

Tamboura, H. H., H. Banga-Mboko, D. Maes, I. Youssao, A. Traore, B. Bayala and M. A. Dembele (2006). "Prevalence of common gastrointestinal nematode parasites in scavenging pigs of different ages and sexes in eastern centre province, Burkina Faso." Onderstepoort J Vet Res 73(1): 53-60.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0." Mol Biol Evol 30(12): 2725-2729.

Yin, Y., J. Martin, S. Abubucker, A. L. Scott, J. P. McCarter, R. K. Wilson, D. P. Jasmer and M. Mitreva (2008). "Intestinal transcriptomes of nematodes: comparison of the parasites *Ascaris suum* and *Haemonchus contortus* with the free-living *Caenorhabditis elegans*." PLoS Negl Trop Dis 2(8): e269.

Zhou, C., M. Li, K. Yuan, S. Deng and W. Peng (2012). "Pig *Ascaris*: an important source of human ascariasis in China." Infect Genet Evol 12(6): 1172-1177.

Supplementary data of manuscript 1

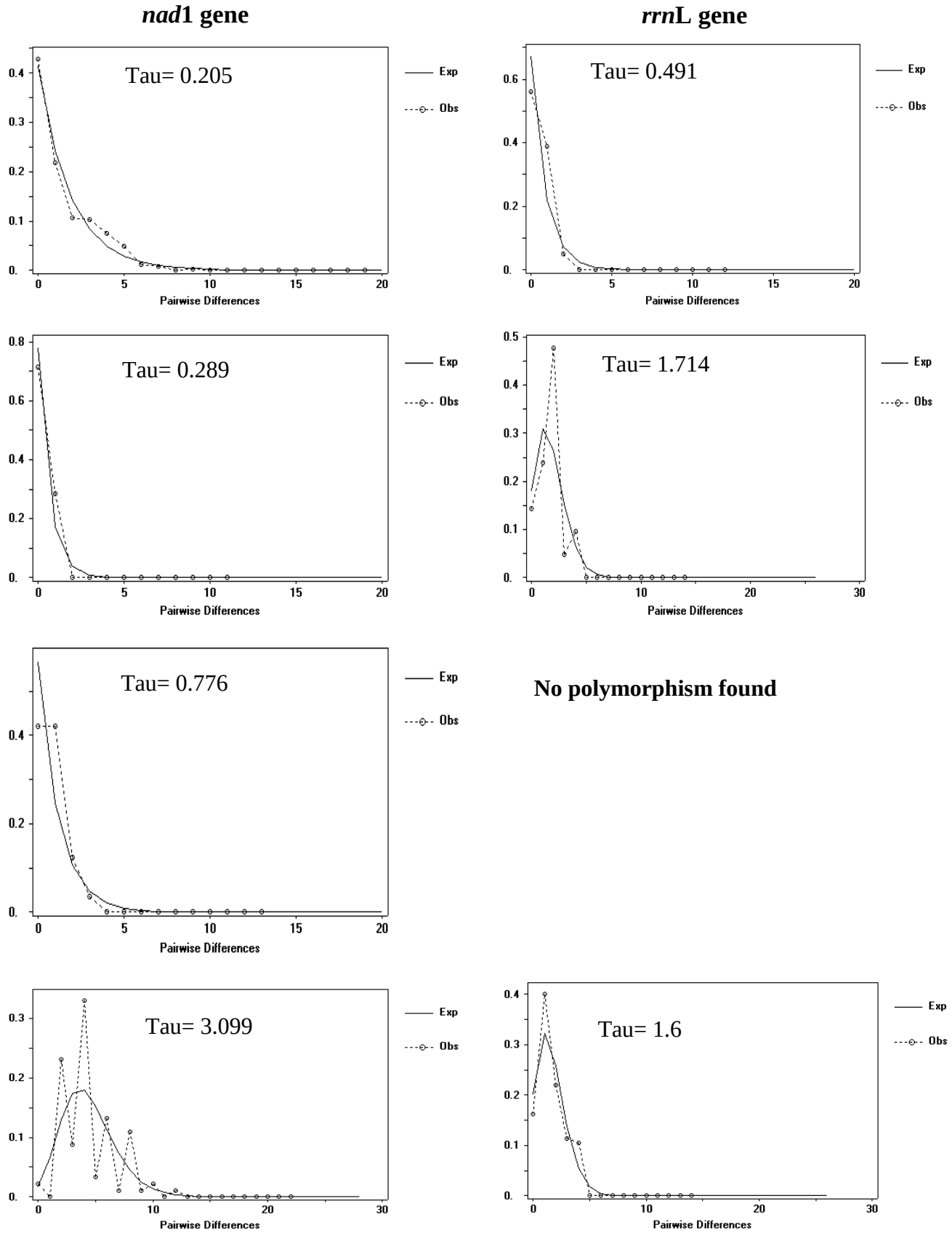

Figure S1. The mismatch distribution curves for the two markers *nad*1 and *rrn*L in the different clades (given in the rightmost side) with the tau values each curve. Tau is an estimate for the time since expansion as it equals $2t\pi$ where (t) is time since expansion in generations and (μ) is the mutation rate.

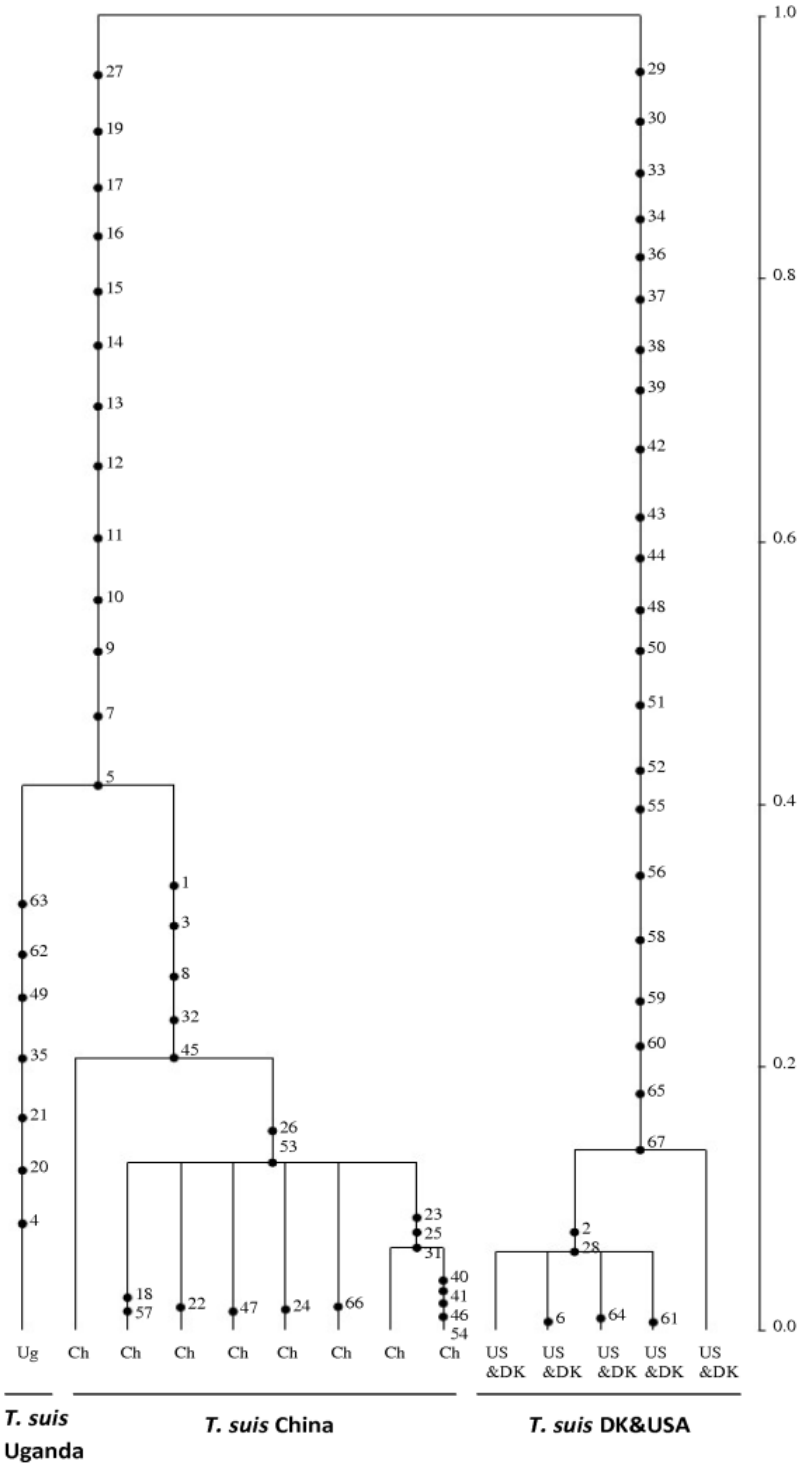

Figure S2. The Genetree of *T. suis* populations. Solid circles indicate mutation according to their positions

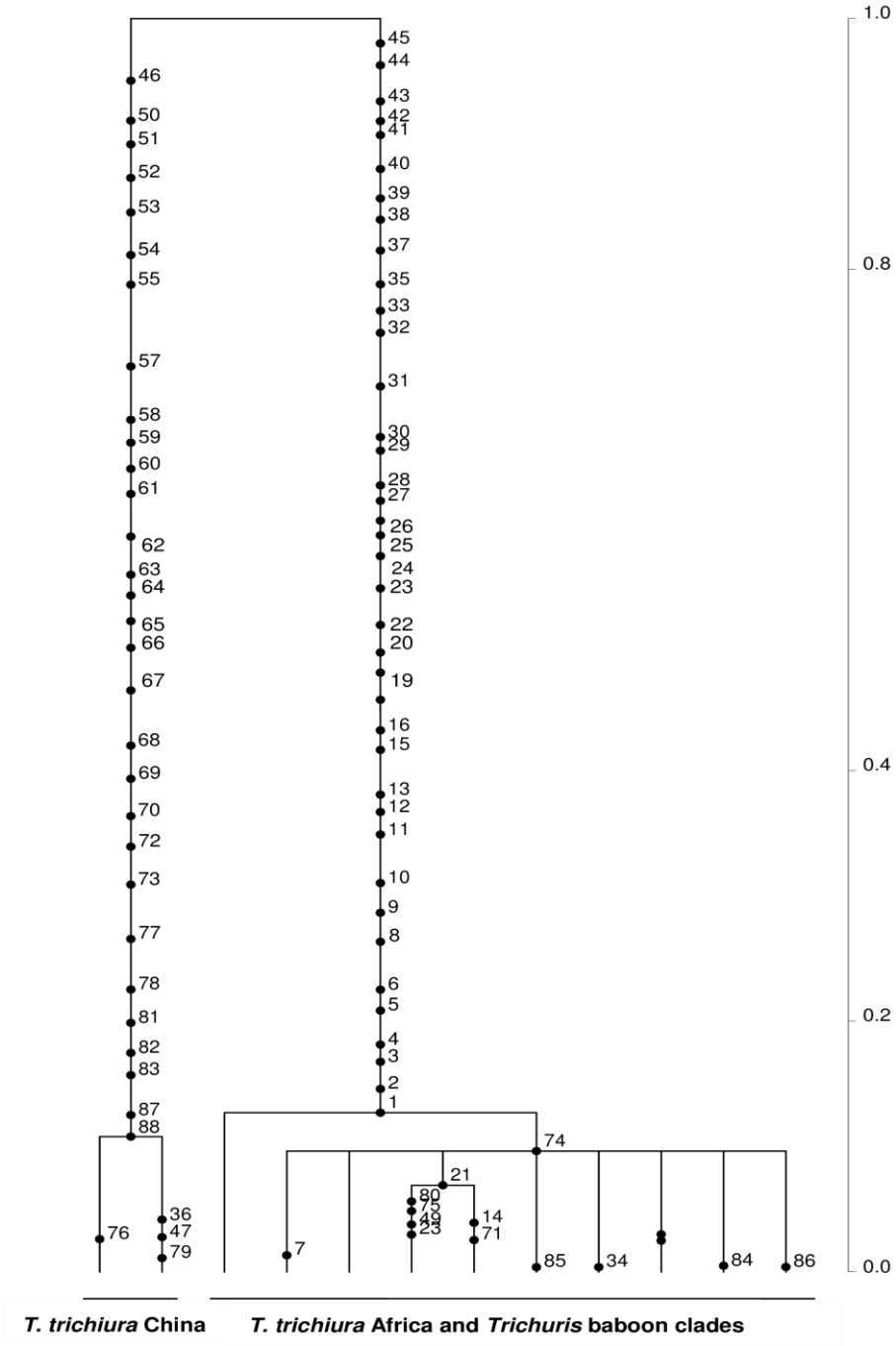

Figure S3. The Genetree of *Trichuris* populations from of humans and baboons. Solid circles indicate mutation according to their positions.

Manuscript 2

Multiple [haplotypic] mitochondrial genomes identified among pig and primate derived *Trichuris*

Mohamed Bayoumi Fahmy Hawash, Department of Veterinary Disease Biology, Faculty of Health and Medical Sciences, Copenhagen University.

Abstract

The whipworms *Trichuris trichiura* and *T. suis* are two nematodes that parasitize humans and pigs respectively. Although historically *T. trichiura* is the species designed to infect both human and non human primates, recent reports support the presence of different *Trichuris* spp. infecting non human primates. Mitochondrial DNA is useful in identifying and delimiting the possible new cryptic (hidden) species in nematodes. In this study, complete mitochondrial genomes were sequenced and annotated for two *Trichuris* worms from olive baboons and two *T. suis* from Denmark and Uganda and were compared to already published genomes. Moreover, the *cox*1 gene phylogeny was conducted to compare the mitochondrial genome haplotypes with other haplotypes of *Trichuris* from pigs and primates. The phylogenetic and sequence analysis suggest the two worms from olive baboons to be different *Trichuris* species. One of two worms from baboons was closely related to human *T. trichiura* mitochondrial genome suggesting to be the same species. The two *Trichuris* from olive baboon and the *T. trichiura* from human were genetically very distinct from *Trichuris* sp. GHL from Francios leaf monkey. Moreover, the *cox*1 phylogeny suggested the presence of five potential different *Trichuris* spp. in primates. The phylogenetic inferences and the genetic distance between *T. suis* in Denmark and China indicated the possible presence of different *T. suis* infecting pigs in these areas which may have important implications for management and control strategies for the parasite.

Key words: *Trichuris* spp., phylogeny, mitochondrial genome, domestic pigs, primates.

1. Introduction

Trichuris trichiura and *T. suis* are two different whipworms infecting humans and pigs respectively. *Trichuris* spp. has direct fecal-oral life cycle starting by ingestion of the infective eggs by the host. After ingestion, larvae undergo several successive molts in host intestin before it becomes adult. *Trichuris* infections rarely cause clinical symptoms. However, infections could be more pronounced in juvenile hosts leading to severe pathological signs in pigs, humans and non human primates (Pittman et al, 2010; Pullan et al, 2014; Abee et al, 2012).

Mitochondrial genome is a small circular genome of a size range of ~ 13-26 kpb that encodes for enzymes required for oxidative phosphorylation (Hu and Gasser, 2006). Mitochondrial DNA has a number of characteristics which make it useful for the study of parasite ecology and genealogy such

as high rate of mutation rate, maternal inheritance and lack of recombination. Besides, phylogenetic analysis of the concatenated amino acid sequences of mitochondrial proteins have shown to identify the phylogenetic relationship between taxa in accurate way (Hu and Gasser, 2006). Mitochondrial genomic studies could also give information about the molecular processes such as gene rearrangements and RNA editing in the parasitic mitochondrial genome (Jex et al, 2010a).

Cryptic 'hidden' species are those that cannot be differentiated by traditional methods e.g. morphology. Mitochondrial DNA has a number of advantages for delimiting closely related species due to its high substitution rate coupled with low effective population size which leads to rapid lineage sorting following speciation (Nadler and Leon, 2012). Blouin (2002) compared the mitochondrial DNA from different nematode taxa and underlined the potential of mitochondrial DNA to detect cryptic species through providing a genetic yardstick for delimiting the closely related nematode species. Considering, until now, the lack of reliable morphological features for differentiation between *T. trichiura* and *T.suis* (Cutillas et al, 2009; Nissen et al, 2012), mitochondrial DNA have been proved to be useful to identify different *Trichuris* species infecting number of hosts (Liu et al, (2012a, b), Liu et al, 2013).

Thanks to the evolutionary advances of the next generation sequencing (NGS) of DNA and the bioinformatic analysis, whole mitochondrial genome sequencing has become cheaper and easier to produce as evident by the number of published genomes in recent years (Jex et al, 2010 (a, b); Hahn et al (2013)). Till now, 5 different mitochondrial genomes from genus *Trichruis* were sequenced from 5 different hosts (Liu et al, 2012 (a, b); Liu et al, 2013). Mitogenomic phylogeny of these genomes revealed clear genetic distinctiveness of the multiple *Tichuris* spp. recovered from the different hosts, namely *T. trichiura, T. suis*, *T. ovis*, *T. discolor* and *Trichuris* sp. GHL from humans, pigs, antelope, yak and François' leaf-monkey respectively (Liu et al, 2012 (a, b); Liu et al, 2013).

Although that *T. trichiura* historically is the species name used for whipworms in human and non human primates, recent studies revealed different results concerning this designation. For example, Ravasi et al (2012) found possible two *Trichuris* species infecting both baboons and humans based on Internal transcribed spacer (ITS) sequences. Another study by Hansen et al (2013) suggested that the same *Trichuris* species infect humans and baboons based on beta-tubulin gene sequencing. On the other hand, Liu et al (2013) identified a potential new species of *Trichuris* in a non human primate (François' leaf-monkey) based on complete mitochondrial genome analysis and nuclear marker ITS-1 and 2.

A study on different *T. suis* populations in different geographical regions identified substantial genetic distinctiveness between them using mitochondrial markers of the NADH dehydrogenase subunit 1 (*nad*1) and large ribosomal subunit (*rrn*L) genes. Moreover, the genetic distance between *T. suis* from Denmark and China was high suggesting presence of different cryptic species of *Trichuris* in pigs in different areas (Hawash, Manuscript 1).

The aim of the current study is to is to sequence full mitochondrial genome of *Trichuris* recovered from olive baboons (*Papio anubis*) and compare with already avilable *Trichuris* genomes from primates (François' leaf-monkey and human) to study the genetic and evolutionary relationship between the different *Trichuris* in these primates. Also, mitochondrial genome was analyzed for *T. suis* worms from pigs from Denmark, Uganda and compared with the available *T. suis* mitochondrial genome from China to analyze the genetic and evolutionary relationship between the different isolates and to test for the presence of complex multiple species of *Trichuris* in pigs.

2. Materials and Methods

2.1. Parasite isolates and total genomic DNA extraction

Adult *Trichuris* worms were recovered from olive baboons *Papio anubis* at Southwest National Primate Research Center (SNPRC), Texas, USA during post mortem examination for reasons not related to this study. Adult *T. suis* worms were collected from domesticated pigs post mortem from Denmark and Uganda as described by Nissen et al. (2012). Worms were rinsed with tap water and transferred to 70% ethanol and stored at 5 °C until DNA extraction.

Total genomic DNA was extracted from the anterior thin part of the worms. MasterPure DNA Purifications Kit (Epicenter Biotechnologies) was used to extract the DNA according to manufacturer's protocol after the worm material had been homogenized in 300µl of lysis solution (295 µl tissue and cell lysis solution + 5µl proteinase K) in an eppendorf tube with matching plastic pestle.

Worms were typed using Restriction Fragment Length Polymorphism-Polymerase Chain Reaction (RFLP-PCR) on two markers namely, internal transcribed spacer-2 (ITS-2) as described previously (Nissen et al, 2012) and the beta tubulin gene as descriped in (Hawash, Manuscript 1) to elucidate worm identity. The Negative water control was included in the runs. PCR amplified products and digested fragments were stained by GelRed (Biotium) then visualized under UV light in 1.5%

agarose gel. Worms from baboon showed banding pattern characteristic to *Trichuris* from primates and worms from pigs showed baning characteristic to *T. suis* in both markers (Nissen et al, 2012; Hawash, Manuscript 1). Partial sequencing of the *rrn*L gene using same PCR conditions described by Hawash (Manuscript 1) on 5 worm samples from baboon and 4 worms from pigs from Denmark and Uganda (two from each country) were chosen based on their distinct haplotypes as found in (Hawash, Manuscript 1). Phylogenetic Neighbor Joining (NJ) tree was depicted for these samples with the available *rrn*L sequences from GenBank to identify and confirm their genetic relatedness (Supplementary data, Figure S1).

2.2. Long range PCR amplification of the mitochondrial genomes

Two baboon worms (TTB1 and TTB2) were chosen for mitochondrial genome sequencing as both showed distinct haplotypes (Supplementary data, Figure S1). The two *T. suis* worms were indicated based on country of origin TSDK from Denmark and TSUG from Uganda. The two mitochondrial genomes of *T. trichiura* and *T. suis* (Accession No: GU385218 and GU070737 respectively) were aligned to identify conserved regions for primer design. Due to the high variation for the two genomes, no suitable conserved regions were found for designing general primers for all samples. Hence, different set of primers were designed for each genome(s). Primers were designed based on the genome of *T. trichiura* from human (GU385218) to amplify the mitochondrial genome on three overlapping fragments (5 Kbp each) (Table 1). However, this primer set could not amplify baboon *Trichuris* sample (TTB1) and another set of primers were designed to amplify this genome in two overlapping fragments (8-6 Kbp each) (Table 1). However, there was unspecific band detected for the fragment *nad*1 – *rrn*L for TTB1. Therefore, in this case, the DNA from targeted band was extracted directly from the agarose gel using spin columns (Millipore) following manufacturers' protocol. For *T. suis*, a primer set was designed based on the mitochondrial genome (GU070737) and the two genomes TSDK and TSUG were amplified in three overlapping fragments (6, 5 and 3 Kbp) (Table 1).

Master mix PCR ingredients were: 2 µl 10X PCR buffer, 0.4 mM of each dNTP, 0.4 mM of each primer pair, 2.0 mM $MgCl_2$, and 2.5 U of Long PCR Enzyme Mix (Thermo Scientific). PCR conditions were initial denaturation at 92 °C for 4 minuts followed by 35 cycles of denaturation at 92 °C (20 sec), annealing at 50 °C for 30 sec, extension at 62-67 °C for 7 minutes and final extension at 60-67 °C for 10 minutes. These conditions were used after testing different temperatures and $MgCl_2$ concentrations. PCR products were stained by Gelred stain (Biotium) and visualized in 0.8% agarose

gel under UV light. Then, fragments were cleaned enzymatically using 10 µl of PCR product, 1 µl Exonuclease I and 2 µl FastAP Thermosensitive Alkaline Phosphotase (1 U/µl) (Fermentas). The samples were incubated for 15 min at 37 °C followed by 15 min at 85 °C. Finally, DNA concentration was measured and equal amounts of fragments of each genome were pooled and sent for Next Generation Sequencing using illumina HiSeq 2000 platform by Macrogen Inc. in Seoul, South Korea.

Table 1. A summary of the primers used for the complete mitochondrial genome sequencing for the *Trichuris* from baboons and *T. suis* from pigs.

	Forward		**Reverse**
TTB2			
TTB2*cox*1F	CAGGAAATCACAAGAAAATTGG	TTB2*nad*5R	AGTGGTTGCAGGAACAATTC
TTB2*nad*5F	AGCAATCTGCGATATTGTTG	TTB2*rrn*LR	TCGCAACGGTTTAAACTCAA
TTB2*rrn*LF	CGCAGTAATCTGACTGTGC	TTB2*cox*1R	AAATTTTCCTGCTATGAATATGA
TTB1			
TTB1*nad*1F	ACAGCCCATCCTAGACGGTA	TTB1*rrn*L	ACCTGTCTCGCAACGGTTTA
TTB1*rrn*LF	TCTGACTGTGCAAAGGTAGCA	TTB1*nad*1R	TTGCGGACCAAAAGGTTATGAAT
T. suis			
TS*rrn*LF	TTAAATGGCCGCAGTAACCT	TS*nad*1R	AGCTCACCCTGTAATAATGATGT
TS*nad*1F	TCTGATCTGTGCTACCCTACAC	TS*nad*5R	CCAACACCCGTGAGTTCTT
TS*nad*5F	CTTTTGCAAGGGCATGATTA	TS*rrn*LR	TCACGTAATGTAGAATCGTCGA

2.3. Assembly, Annotation and genome sequence analysis

The reads of each genome (~ 100 bp each) were assembled using CLC Genomics Workbench v6.0.4 (CLC Inc, Aarhus, Denmark) using reference genomes of *T. trichiura* (GU385218) and *T. suis* (GU070737). The read coverage was >180x for TTB1 and TTB2 and >75x for TSUG and TSDK. After assembly, genome annotation followed the pipeline MITOS (Bernt et al, 2013) and the BLAST search tools available through NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Secondary structures for all tRNA were predicted using tRNAScan-SE (Schattner et al, 2005) and ARWEN (Laslett and Canback, 2008).

The genomes were compared to *T. trichiura* from human in China (TTHCh) (Accession No: GU385218); *Trichuris* sp. GHL from Francois' leaf-monkey (T.GHL) (Accession No: KC461179) and *T. suis* from China (TSCH) (Accession No: GU070737). The genetic distance for all the protein coding genes (PCGs) and the ribosomal DNA (rDNA) was estimated between the genomes beside

the distance for all mitochondrial proteins using MEGA v6.1 (Tamura et al, 2013). The nucleotide diversity (pi) was calculated across the genomes of *Trichuris* from humans and non human primates and *T. suis* from pigs using a sliding window of 100 bp with 25 bp steps implemented in DnaSP v.5 (Librado and Rozas, 2009). Sliding window analysis was also used to estimate pairwise nucleotide diversity between *T. trichiura* from human (Accession No: GU385218) with TTB1, TTB2 and *Trichuris* sp. GHL from Francois' leaf-monkey (Accession No: KC461179) and between *T. suis* from China (Accession No: GU070737) with TSDK and TSUG. Moreover, DnaSP v.5 (Librado and Rozas, 2009) was used to calculate number non-synonymous (Ka) and synonymous substitutions (Ks) for each PCGs and the ratio of Ka/Ks was calculated subsequently. The Ka/Ks ratio gives indication on the pattern of selection acting on the genes. If Ka/Ks >1 this indicates positive (diversifying) selection, Ka/Ks=1 indicates neutral mutation (no selection) and Ka/Ks <1 indicates negative (purifying) selection (Hurst, 2002).

2.4. Phylogenetic analysis

Three different methods were used for phylogenetic analysis namely Neighbor Joining (NJ), Maximum Likelihood (ML) and Bayesian Inferences (BI). Two different data sets were generated for the phylogenetic analyses (one for amino acid sequences and one DNA sequences). Amino acid sequences for the 13 PCGs were aligned by ClustalW for 9 *Trichuris* spp. (the *Trichuris* from baboons TTB1, TTB2; from human from China (Accession No: GU385218); from Francois' leaf-monkey (Accession No: KC461179); from pigs from Uganda (TSUG), Denmark (TSDK) and China (TSCH) (Accession No: GU070737)) besides the *T. discolor* (Accession No: JQ996231) and *T. ovis* (Accession No: JQ996232). Similarly, the data set of DNA sequences consisting of the PCGs and rDNA genes were aligned using ClustalW for analysis. *Trichinella spiralis* (Accession No: AF293969) was used as an outgroup in the phylogeny.

ML and NJ trees were generated using MEGA v6.1 (Tamura et al, 2013). The best-to-fit substitution model was identified using jModelTest0.1.1 (Posada, 2008) under Akaik information criterion (AIC) (Akaike, 1974) for each dataset. BEAST v.1.6.1 (Drummond and Rambaut, 2007) was used for the BI on the two alignments referred above. Uncorrelated log normal was used as prior for the mutation rate with mtRev as a substitution model for protein sequences and General Time Reversible (GTR) model for DNA sequences with gamma distribution and invariant sites assumed in both substitution models. Random starting tree with Yule prior was assumed as well. Three independent runs with 10 millions steps each with a burn in of 10k steps were conducted. Tracer v.1.6 was used to analyze log

files of the MCMC chains and the reliability of parameters checked by recording effective sample size values to be above 200. Tree Annotater v1.6.1 (Drummond and Rambaut, 2007) was used to summarize samples from the posterior on maximum credibility tree and the posterior probability limit set to 0.5. Figtree v1.3.1 (Drummond and Rambaut, 2007) was used to depict the BI tree.

2.5. *Cox*1 phylogeny

In order to investigate the phylogenetic relationship between the mitochondrial genome haplotypes of the *Trichuris* spp. in this study with other *Trichuris* haplotypes from non human primates and pigs, *cox*1 partial sequences from GenBank was obtained (Table 2). ML and NJ trees were analyzed using MEGA v6.1 (Tamura et al, 2013) and the best to fit model was identified using jModelTest0.1.1 (Posada, 2008) under Akaik information criterion (AIC) (Akaike, 1974). *Ascaris lumbricoides* (Accession No: AB591799) was used as outgroup.

Table 2. Different *cox*1 sequences included from the GenBank with the accession numbers, the host and the country from which worms were sampled. All the non human primates represented in the table are in captive conditions.

Host	Country	Accession No. of GenBank
Colobus guereza kikuyensis (mantled guereza)	Spain	HE653116, HE653117, HE653118, HE653119.
Papio anubis (Olive baboons)	Czech Republic	JF690964
Theropithecus gelada (gelada baboon)	Czech Republic	JF690965
Papio hamadryas (hamadryas baboon)	Czech Republic	JF690963
Macaca fascicularis(longtailed macaque)	Czech Republic	JF690967
Sus scrofa domestica (domestic pig)	China	HQ204208, HQ204209, HQ183740, HQ183741
Sus scrofa domestica (domestic pig)	Spain	HE653124, HE653125, HE653126
Sus scrofa scrofa (wild boar)	Spain	HE653127, HE653128, HE653129

3. Results

3.1. Annotation and features of mitochondrial genomes

The complete mitochondrial genome of the two baboon worms TTB1 and TTB2 were of 14,105 and 14,009 bp respectively and for TSDK and TSUG were 14,586 and 14,410 bp respectively (Appendix of raw data). The genomes encode for 13 PCGs, 22 transfer RNA (tRNA) and 2 ribosomal DNA genes as given in Table 3 and 4. The general mitochondrial features and gene order is the same as previously published for *Trichuris* spp. (Liu et al 2012 (a, b)) as all genes are transcriped from heavy chain except 4 PCGs (*nad*2, *nad*5, *nad*4 and *nad*4L) and 10 tRNA motifs (tRNA-Met, tRNA-Phe,

tRNA-His, tRNA-Arg, tRNA-Pro, tRNA-Trp, tRNA-Ile, tRNA-Gly, tRNA-Cys and tRNA-Tyr) are transcriped from the light chain.

The starting and termination codons for some PCGs were different between the genomes of the *Trichuris* spp. recovered from the same host species. For instance, the starting codon for TSDK is ATA for *nad*4 gene while it is ATG in the TSUG genome and for the *atp*6 gene, the starting codon is GTA in TSDK and GTG in TSUG. Also the termination codon was TAG for TSDK and TAA for TSUG in the *cox*1 gene and in the *nad*4 gene TAA is the termination codon in TSDK and TAG for TSUG. Remarkably, I found a incomplete termination codon (T) was in the *nad*4L gene in the TSDK genome will it was TAG in the TSUG genome. Similar observations were found in the *Trichuris* genomes from baboons. For genes *nad*2, *atp*6 and *atp*8 the starting codons are ATA, ATG and ATA respectively in TTB1 while they are GTA, GTG and ATA in TTB2. Likewise, the termination codons are TAG for *cox*2, *nad*1 and *nad*5 genes in TTB1 while it was TAA for the same genes in the TTB2 genome. Also the length of the open reading frame (ORF) for some genes was different between the genomes. *nad*4 and *nad*4L show different ORG between TSDK and TSUG. For TTB1 and TTB2 genomes *nad*1, *nad*2, *nad*5, *nad*4 and *nad*4L, *atp*6 and *atp*8 show different lengths of the ORFs.

Table 3. Mitochondrial genomes of baboon *Trichuris* (TTB1 and TTB2). The protein conding, transfer RNA (tRNA) and the ribosomal DNA (rDNA) genes with the lengths of the nucleotide (nt) and the amino acids (aa) each gene are given. The lengths are given for TTB1 and differences for TTB2 is given between brackets and likewise for the initiation and termination codons.

Gene/Region	Positions and nt lengths		Lengths		Codons		Strand
	TTB2	TTB1	aa	nt	Initiation	Termination	
***cox*1**	1-1545	1-1545	514	1545	ATG	TAA	+
***cox*2**	1559-2234	1557-2232	224	675	ATG	TAG(TAA)	+
tRNA-leu	2248-2308	2254-2317		63(-3)			+
tRNA-glu	2317-2374	2323-2384		61(-4)			+
***nad*1**	2396-3296	2402-3305	300(-1)	903(-3)	ATA	TAG(TAA)	+
tRNA-lys	3423-3484	3447-3513		66(-5)			+
***nad*2**	3486-4371	3510-4407	298(-4)	897(-12)	ATA(GTA)	TAA	-
tRNA-meth	4383-4444	4408-4469		61			-
tRNA-phe	4440-4496	4462-4519		57(-1)			-
***nad*5**	4495-6043	4510-6067	518(-3)	1557(-9)	ATA	TAG(TAA)	-
tRNA-his	6040-6094	6060-6118		58(-4)			-
tRNA-arg	6095-6158	6120-6182		64			-
***nad*4**	6159-7371	6187-7408	406(-3)	1221(-9)	ATG	TAA	-
***nad*4L**	7401-7650	7430-7643	70(+12)	213(249)	ATA	TAA	-
tRNA-thr	7654-7710	7683-7741		58(-2)			+
tRNA-pro	7712-7771	7741-7799		59			-
***nad*6**	7762-8239	7791-8268	158	477	ATT	TAA	+
***cyt*b**	8245-9352	8274-9381	368	1107	ATG	TAG	+
tRNA-ser	9350-9400	9379-9432		53(-3)			+
***rrn*S**	9392-10086	9424-10,130		706(694)			+
tRNA-val	10,088-10,145	10,131-10,188		57			+
***rrn*L**	10,143-11,153	10,189-11,198		1009(+2)			+
***atp*6**	11,123-11,963	11,168-11,981	270(+9)	813(840)	ATG(GTG)	TAA	+
***cox*3**	11,937-12,711	11,986-12,760	257	774	ATG	TAA	+
tRNA-trp	12,717-12,780	12,772-12,835		63			-
tRNA-gln	12,783-12,836	12,838-12,894		56(-3)			+
tRNA-Ile	12,838-12,898	12,896-12,957		61			-
tRNA-gly	12,907-12,964	12,970-13,027		57			-
tRNA-asp	12,969-13034	13,033-13,091		58(+8)			+
***atp*8**	13,015-13,180	13,079-13,247	55(-1)	168(-3)	ATT(ATA)	TAG	+
***nad*3**	13,189-13,531	13,256-13,598	113	342	ATT	TAA	+
tRNA-ser	13,625-13,675	13,691-13,741		50			+
tRNA-asn	13,676-13,731	13,742-13,797		55(-1)			+
tRNA-leu	13,738-13,801	13,804-13,864		60(+3)			+
tRNA-ala	13,807-13,860	13,875-13,933		58(-4)			+
tRNA-cys	13,889-13,942	13,975-14,029		54(-1)			-
tRNA-tyr	13,954-14,001	14,030-14,091		61			-

Table 4. Mitochondrial genome of *T. suis* (TSDK and TSUG). The protein conding, transfer RNA (tRNA) and the ribosomal DNA (rDNA) genes with the lengths of the nucleotide (nt) and the amino acids (aa) each gene are given. The lengths are given for TSDK and the differences for TSUG is given between brackets and likewise for the initiation and termination codons.

Gene/Region	Positions and nt lengths		Lengths		Codon		Strand
	TSDK	TSUG	aa	nt	Initiation	Termination	
***cox*1**	1-1542	1-1542	513	1542	ATG	TAG(TAA)	+
***cox*2**	1578-2259	1576-2257	226	681	ATG	TAA	+
tRNA-leu	2270-2330	2268-2328		60			+
tRNA-glu	2338-2394	2335-2392		57			+
***nad*1**	2416-2315	2413-2313	299	900	ATT	TAA	+
tRNA-lys	3447-3506	3453-3513		60			+
***nad*2**	3521-4403	3527-4409	293	882	ATA	TAG	-
tRNA-meth	4412-4474	4418-4480		62			-
tRNA-phe	4478-4539	4483-4542		61(-2)			-
***nad*5**	4527-6084	4532-6089	518	1557	ATA	TAG	-
tRNA-his	6085-6141	6090-6146		56			-
tRNA-arg	6144-6208	6149-6213		64			-
***nad*4**	6213-7617	6217-7606	467(-5)	1404(-15)	ATA(ATG)	TAA(TAG)	-
***nad*4L**	7765-8033	7627-7891	89(-1)	268(-7)	ATA	T(TAG)	-
tRNA-thr	8035-8090	7883-7938		55			+
tRNA-pro	8082-8153	7943-8001		71(58)			-
***nad*6**	8144-8615	7993-8464	156	471	ATT	TAA	+
***cyt*b**	8627-9740	8475-9588	370	1113	ATG	TAG	+
tRNA-ser	9738-9793	9586-9641		55			+
***rrn*S**	9790-10,500	9638-10,349		710			+
tRNA-val	10,499-10,556	10,348-10,405		57			+
***rrn*L**	10,564-11,575	10,413-11,420		1011			+
***atp*6**	11,564-12,374	11,409-12,219	269	810	GTA(GTG)	TAA	+
***cox*3**	12,382-13,159	12,228-13,005	258	777	ATG	TAA	+
tRNA-trp	13,163-13,230	13,009-13,076		67			-
tRNA-gln	13,233-13,290	13,079-13,136		57			+
tRNA-ile	13,292-13,358	13,138-13,204		66			-
tRNA-gly	13,376-13,432	13,222-13,278		56			-
tRNA-asp	13,446-13,506	13,290-13,350		60			+
***atp*8**	13,485-13656	13,330-13,501	56	171	TTG	TAA	+
***nad*3**	13,680-14,022	13,525-13,867	113	342	ATA	TAA	+
tRNA-ser	14,129-14,183	13,974-14,026		54(-2)			+
tRNA-asn	14,182-14,242	14,025-14,084		60(-1)			+
tRNA-leu	14,257-14,317	14,101-14,164		60(+3)			+
tRNA-ala	14,322-14,377	14,166-14,221		55			+
tRNA-cys	14,408-14,466	14,244-14,299		58(-3)			-
tRNA-tyr	14,465-14,522	14300-14,357		57			-

3.2. Comparative sequences analysis and polymorphism

The two baboon genomes TTB1 and TTB2 were compared with *T. trichiura* from human in China (TTHCh) and *Trichuris* sp. GHL from Francois' leaf-monkey (T.GHL). The genetic distance between each PCGs and rDNA genes between the different genomes of *Trichuris* spp. is shown in Table 5 together with the amino acids differences based on the encoded proteins. The highest genetic variation was found in *atp*8 gene and the most conserved gene was *rrn*S. Among all the mitochondrial proteins, *cox*1 was found to be the most conserved and *atp*8 to be the least conserved. The overall differences for DNA sequences in the PCGs and rDNA and amino acides sequences for mitochondrial proteins for all *Trichuris* spp. genomes are shown in Table 6. The overall differences in the nucleotides and amino acids between the genomes of the baboon worms TTB1 and TTB2 were very high (14.4 in amino acids and 20.4 in nucleotides respectively) with TTB2 to be more genetically related to *T. trichiura* from human China than TTB1. However, they all showed higher variation compared to T.GHL (Table 5). The differences between the TSUG and TSCH was lower compared to TSDK.

Table 5. The genetic distances between the protein coding genes (PCGs) and the ribosomal DNA (rDNA) between the genomes of *Trichuris* spp. from human, non human primates and pigs. The comparison is made between the human *T. trichiura* (GU385218) to the baboons TTB1 and TTB2 and Francois' leaf-monkey *Trichuris* sp. GHL (T.GHL)(KC461179). Similarly, the comparison is between the Chinese *T. suis* (GU070737) and the *T. suis* from Denmark (TSDK) and Uganda (TSUG).

	***T. trichiura* (GU385218) vs.** (TTB1/TTB2/T.GHL)		***T. suis* (GU070737) vs.** (TSDK/TSUG)	
PCGs	Amino acids	Nucleotides	Amino acids	Nucleotides
	TTB1/TTB2/T.GHL	TTB1/TTB2/T.GHL	TSDK/TSUG	TSDK/TSUG
***cox*1**	3.5/2.2/13.1	16.1/5.5/24	1.6/0.4	8.3/3.6
***cox*2**	7.6/2.7/18.8	17.2/6.7/24.3	3.5/0.0	9.0/3.8
***cox*3**	10.9/3.5/25.7	20.9/8.1/29.2	5.8/1.2	10.4/3.2
***cyt*b**	11.1/4.6/20.9	17.7/6.8/26.1	4.6/1.1	9.7/4.0
***nad*1**	11.7/2.0/28.1	19.4/4.6/30.1	4.0/2.7	11.7/4.2
***nad*2**	16.3/5.4/30.6	19.8/9.3/33.6	10.2/3.4	11.0/3.5
***nad*3**	13.4/5.4/26.8	19.2/5.9/29.5	4.4/0.9	8.5/2.6
***nad*4**	17.9/6.5/31.3	20/8.7/31.3	8.3/3.6	10.5/4.4
***nad*4L**	17.1/4.3/30.0	20.7/8/24.9	10.7/0.0	16.1/3.1
***nad*5**	23.9/5.8/37.5	23.3/8/34.6	7.5/4.1	10.3/4.3
***nad*6**	17.4/3.9/34.2	19.2/7.7/31	3.2/1.3	7.9/2.3
***atp*6**	28.1/10.6/39.9	25.6/8.3/34.7	10.8/1.5	11.9/3.5
***atp*8**	40.9/11.4/56.8	27.4/8.5/40.9	11.4/9.1	5.3/2.9
rDNA				
***rrn*S**		13.4/2.5/22.9		5.5/1.8
***rrn*L**		16.2/4.2/23.9		5.0/1.2

Table 6. The overall genetic and protein distances between the *Trichuris* spp. genomes. TTB1 and TTB2 are baboon *Trichuris*, TTHCh is the human *Trichuris* from China and T.GHL is *Trichuris* sp. GHL from Francois' leaf-monkey. TSDK, TSUG and TSCh are the *T. suis* from pigs from Denmark, Uganda and China respectively. The data below the diagonal is for the amino acid distances and above the diagonal is for genetic distances for all the protein coding and rDNA genes.

Trichuris genomes	TTB1	TTB2	TTHCh	T.GHL	TSDK	TSUG	TSCh
TTB1		20.4	19.9	29.4	35.9	35.7	35.8
TTB2	14.4		7.3	30.6	36.5	36.8	36.7
TTHCh	14.8	4.7		29.8	36.1	36.4	36.3
T.GHL	26.7	28	27.6		38	37.6	37.3
TSDK	39.1	38.9	39.2	41		9.3	9
TSUG	33.2	33.4	33.6	35.7	5.5		3.4
TSCh	35.1	35	35.3	37.3	6.1	2.2	

Moreover, the nucleotide diversity among the *Trichuris* spp. genomes was analyzed using sliding window for all the PCGs. Figure 1 shows variation estimated for all the *Trichuris* genomes from human and non human primates and Figure 2 for all the *T. suis* genomes. The pairwise nucleotide diversity across the mitochondrial genome between *T. trichiura* (Accession No: GU385218) with TTB1, TTB2 and *Trichuris* sp. GHL from Francois' leaf-monkey (Accession No: KC461179) is given in supplementary data (Figure S2) while between *T. suis* from China (Accession No: GU070737) with TSDK and TSUG in supplemetary data (Figure S3).

The ratio between the non-synonymous substitutions (Ka) to synonymous substitutions (Ks) is calculated between each pair of *Trichuris* from human and non human primates and *T. suis* genomes for all PCGs (Supplementary data, TableS1, S2). All values were below 1 indicating presence of negative (purifying selection) in the *Trichuris* spp. mitochondrial genes.

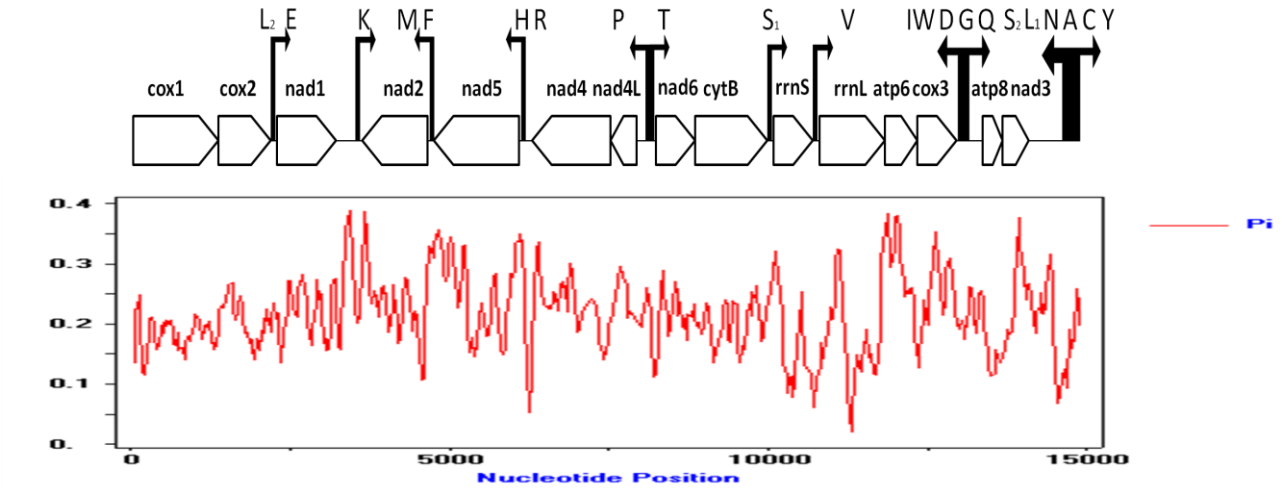

Figure 1. The nucleotide diversity (pi) across the mitochondrial genome for *Trichuris* of TTB1 and TTB2, the human *T. trichiura* from China (GU385218) and the *Trichuris* sp. GHL from Francois leaf monkey (KC461179). Solid arrows indicate the transfer RNA (tRNA).

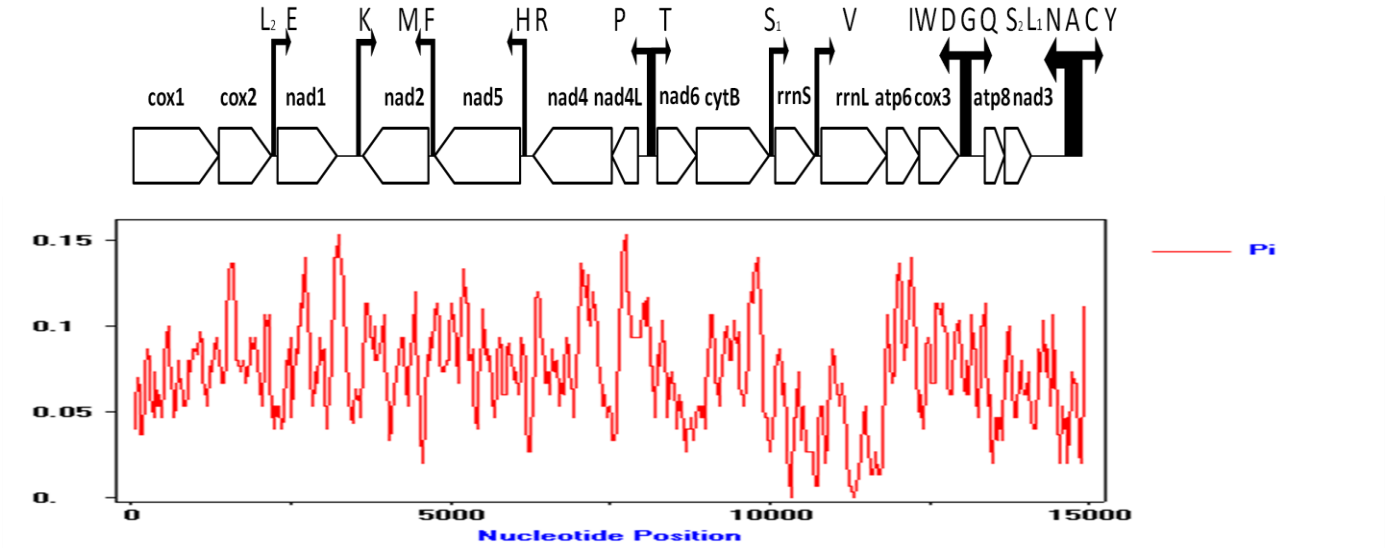

Figure 2. The nucleotide diversity (pi) across the mitochondrial genomes of all *Trichuris suis* genomes from different countries namely, Denmark (TSDK), Uganda (TSUG) and China (TSCH)(GU070737). Solid arrows indicate the transfer RNA (tRNA).

3.3. Phylogenetic analysis

The data set of amino acid and DNA sequences gave similar trees for all the methods applied (NJ, ML and BI). The best to fit model was mtREV+G+I+F for the amino acid sequences and General Time Reversible with gamma distribution and invariant site (GTR+G+I) for the nucleotide sequences. The phylogenetic relationship for all the *Trichuris* spp. was supported by high bootstrap Frequency (BF) value and posterior probability (PP) as shown in Figure 3.

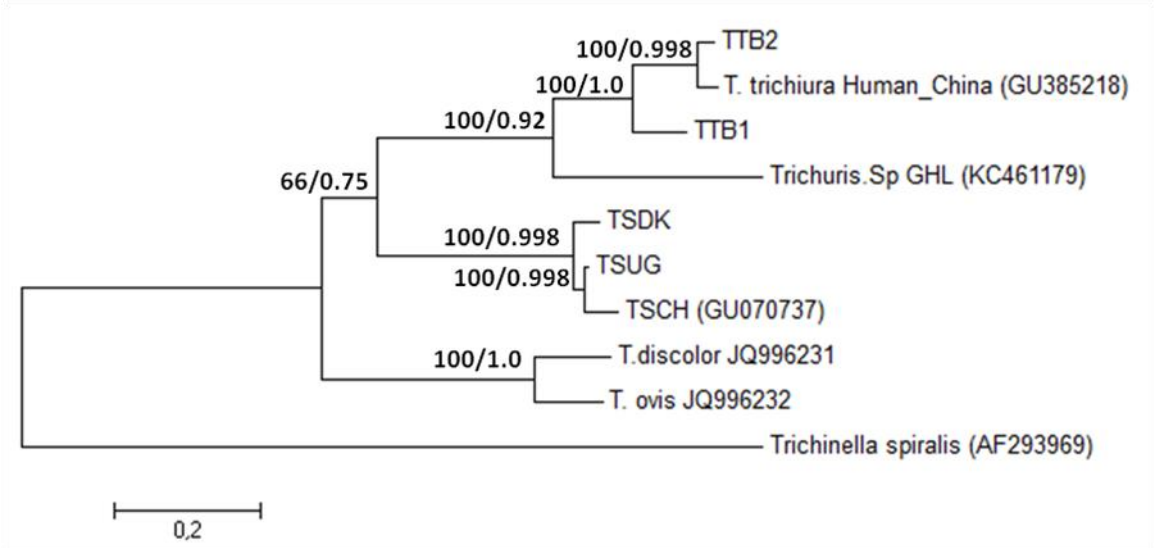

Figure 3. Phylogenetic relationship between the *Trichuris* spp. using concatenated amino acid sequences and Maximum Likelihood and Bayesian Inference. The bootstrap frequency (BF) and posterior probability (PP) are given on the branches (BF/PP).

3.4. *Cox*1 phylogeny

The best fitting model was Tamura 3-parameter model with gamma distribution for *cox*1 sequences. Both NJ and ML depicted similar tree topology, hence NJ tree is represented in Figure 4. TSDK was clustered with Spanish *T. suis* and formed the *'T. suis* Europe' clade and TSCH with *T. suis* from China in the *'T. suis* China' clade while TSUG was closely related to the *T. suis* China clade concordant with the mitochondrial genome phylogeny. The phylogeny of *cox*1 also supported the presence of complex *Trichuris* spp. infecting primates. The *cox*1 phylogeny identified 5 distinct clades which were named after the *Trichuris* spp. recovered from these hosts namely, *T. coloae* is the species in *Colobus guereza kikuyensis* (Cutillas et al, 2014), *Trichuris* sp. GHL in Francois' leaf-monkey (Liu et al, 2013), TTB1 and TTB2 from olive baboon in this study while the last clade comprises *Trichuris* sp. from different non human primates (*Theropithecus gelada*, *Macaca fasciculari* and *P.anubis*) and was named as *'Trichuris* spp. non human primates'. TTB1 was clustered with the *Trichuris* from baboon sample (*P. hamadryas*) while TTB2 clustered with the *T. trichiura* recovered from a human from China in accordance with the mitochondrial genome phylogeny (Figure 4). Remarkably, the *Trichuris* sp. from *P. anubis* in the Czech Republic was genetically very distinct and clustered in a distant clade (*Trichuris* spp. non human primates) than the TTB1 and TTB2 although they are *Trichuris* sp. from the same host species.

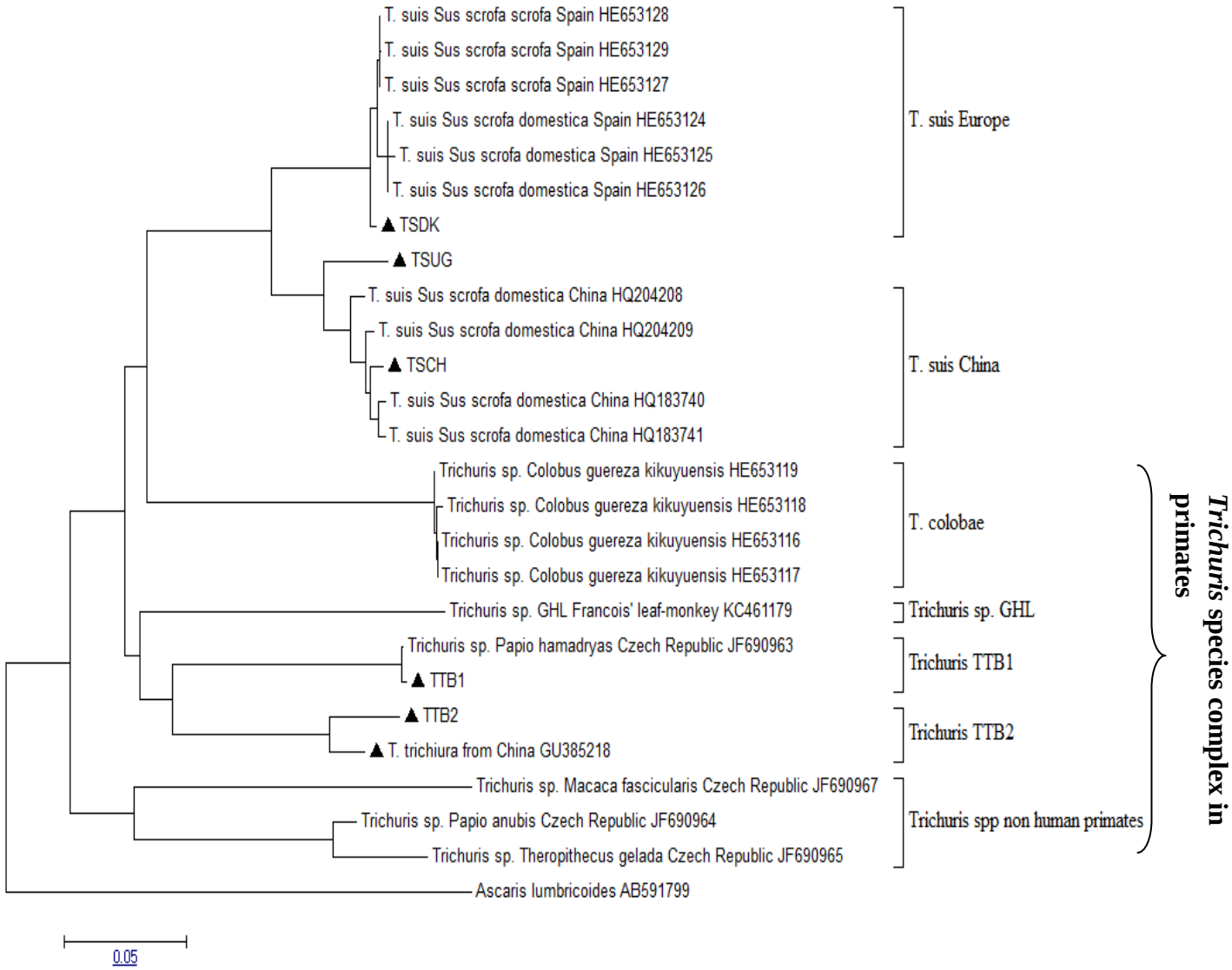

Figure 4. Inference of the *Trichuris* spp. phylogeny of *Trichuris* from pigs and primates based on partial *cox*1 sequences and NJ clustering. The samples that have the whole mitochondrial genome sequenced and included in this tree are indicated by a solid triangle. The phylogeny identified two major clades for *T. suis* in pigs, namely *T. suis* from Europe and *T. suis* from China. Five distinct clades of *Trichuris* spp. recovered from non human primates were identified namely *T. colobae*, *Trichuris* sp. GHL, *Trichuris* TTB1, *Trichuris* TTB2 and *Trichuris* spp. non human primates.

4. Discussion

The aim of this study was to sequence full mitochondrial genomes of *Trichuris* spp. recovered from olive baboons and pigs which carried different haplotypes and compare them to already published genomes of *Trichuris* spp. obtained from pigs, human and a leaf monkey in order infer their genetic relatedness and evolutionary relationship.

Phylogenetic analysis based on the full mitochondrial genomes identified the two olive baboon worms TTB1 and TTB2 in distinct clades supported by the high BF and PP (Figure 3). Moreover, the

difference in amino acid PCGs sequence was around 14.4% which is in the range previously reported between species indicating presence of two *Trichuris* species infecting olive baboons. For example, it was found that differences in amino acid in mitochondrial protein sequences between *T. ovis* and *T. discolor* is 15.4% (Liu et al, 2012b), 11.7% between *Wuchereria bancrofti* and *Brugia malayi* (Ramesh et al ,2012), 10.3% between *Chabertia ovina* and *C. erschowi* (Liu et al, 2014). TTB2 is more genetically related to the human *Trichuris* from China (7.3% in genetic distance). According to the observation by Blouin (2002) to delimit the possible cryptic species, the genetic distance between two closely related nematode species ranges between 10-20% based on the *cox*1 and *nad*4 genes. Hence, TTB2 may be the same species of *Trichuris* infecting the human as average genetic distance is only ~ 5% in *cox*1 and 8.5% in *nad*4 (Table 4). However, previous studies also showed similar diversity between different species based on the mitochondrial protein sequences such as *Ancylostoma duodenale* and *A. caninum* (4%) (Hu et al., 2002; Jex et al., 2009), differences in the range of 5.6-7.2% between different *Toxocara* spp. (Li et al, 2008) and difference in the range of 5.8-18% between different of *Hypodontus* spp. recovered from different hosts (Jabbar et al, 2013). Hence, taxonomic statud of TTB2 is hard to infer and other nuclear markers (e.g. ITS) should be used to investigate its genetic relatedness to other *Trichuris* from humans. The phylogenetic analysis and the genetic distance supported the previous finding suggesting the *Trichuris* sp. GHL from Francios leaf monkey is a different species as it is genetically very distant (26.7 to 28 in protein sequences) from other human and baboon worms.

The *cox*1 phylogeny also supported TTB1 and TTB2 as different species as they were clustered in two distinct clades named here as '*Trichuris* TTB1' and '*Trichuris* TTB2' clades. The previously identified potential new *Trichuris* species from different non human primates (*T. coloae* and *Trichuris* sp. GHL in mantled guereza and Francois' leaf-monkey, respectively) (Liu et al, 2013; Cutillas et al, 2014) were also clustered in distinct clades (Figure 4). Moreover, there was a distinct clade which comprises a *Trichuris* from other non human primates (olive, gelada baboons and longtailed macaque) which also could represent different *Trichuris* species (or species complex) in non human primates. Hence, the *cox*1 phylogeny identified at least five potential *Trichuris* spp. infecting primates. Although that the clade '*Trichuris* spp. non human primates' is distinct from the '*Trichuris* TTB1' and '*Trichuris* TTB2' clades, it contains *Trichuris* sp. infecting the same host species as in the TTB1 and TTB2 clades (*P. anubis* from Czech Republic). This suggests the potential of many *Trichuris* spp. to infect olive baboon (*P. anubis*). In contrast to all *Trichuris* species in primates and pigs, only *T. colobae* is reported to be a morphologically distinct (i.e. not

cryptic). One caveat in this study is that the morphological analysis was not conducted for the different *Trichuris* spp. included which is crucial in order to determine if the suggested different species in primates and pigs are cryptic or pseudo-cryptic (i.e. existence of undetected morphological distinctiveness). Hence, further studies on morphological features combined with analysis of other nuclear markers (e.g. ITS) are warranted to confirm the speciation.

Although that the phylogenetic analysis identified different *Trichuris* spp. from different non human primates, However, since a single human *T. trichiura* was included in the analysis, it is therefore not possible to test whether the *Trichuris* infecting humans is different from those found in non human primates. Further studies on *Trichuris* obtained from humans and non human primates in natural habitat are highly warranted in order to investigate the transmission of *Trichuris* between the different primate spices. Such studies, will have many important implications. First, it will reveal the zoonotic potential of *Trichuris* spp. from non human primates as previous studies supported presence of a common *T. trichiura* species infecting both human and non human primates (Ravasi et al, 2012; Hansen et al, 2013; Hawash, Manuscript 1). Moreover, it could also give insights on the risk of zooanthroponosis (the transmission of pathogens from human to animals) especially since parasites of human origin has been found with higher prevalence in primates under captivity than those in the wild (Rothman et al, 2002; Mbaya and Udendeye, 2011; Messenger et al, 2014). Secondly, it may evaluate the potential of *Trichuris* ova from non human primates in the treatment of Inflammatory Bowel Disease (IBD) in humans. IBD is an autoimmune disease of the gastrointestinal tract of humans which may be cured by treatment with *T. suis* eggs (Reddy and Fried, 2007; Jouvin and Kinet, 2012). Remarkably, a recent clinical trial in rhesus monkeys using *T. trichiura* eggs from humans was successful in the treatment of Idiopathic chronic diarrhea (ICD), which resembles IBD (Broadhurst et al, 2012). Hence, there are suggestion that also *Trichuris* from non human primates could also be used for the treatment of IBD in humans (Callejon et al, 2013) especially since a recent clinical trial using *T. suis* for treatment of IBD failed to show significant effect in the treatment (Garg et al, 2014). However, this approach needs rigorous testing on the different *Trichuris* spp. infecting humans and non human primates to investigate which could be used as a possible treatment.

The phylogenetic relationship between *T. suis* from different regions was also supported by high PP and BF values. Moreover, the nucleotide distance between TSDK compared to TSCH and TSUG were high (around 9%) while TSUG and TSDK were genetically very close (genetic distance 3.4%). Hence, according to the findings of Blouin (2002) the *Trichuris* from pigs in Denmark may therefore be a different species than the ones found in pigs from Uganda and China. The phylogeny based on

*cox*1 sequences (Figure 4) identified the *T. suis* from Spain to be genetically closely related to TSDK. The phylogeny therefore reflects the geographical distribution of worms and it remains unknown whether *T. suis* consist of different populations of worms of the same species or compromise a cryptic species complex. Future studies on the morphological features and other markers (ITS) could elucidate the presence of a cryptic species.

The presence of different cryptic species of might be important in terms of implementing appropriate control measures and strategies (Nadler and Leon, 2012). For example, the parasitic nematode of sheep, *Teladorsagia circumcincta* was found to consist of many cryptic species which might show different pathology and drug resistance potential (Grillo et al, 2007; Nadler and Leon, 2012). Moreover, a previous study suggested possible different cryptic species of *T. trichiura* in humans in different regions which might have important consequences to establish proper control (Hawash, Manuscript 1). Hence, characterization of the species diversity of a parasite is crucial to compare the pathology and epidemiology for the different cryptic species for proper control strategies (Nadler and Leon, 2012).

The ratio between non-synonymous substitutions/ synonymous substitutions (Ka/Ks) was negative for all the mitochondrial genes investigated in *T. suis* and *T. trichiura* indicating negative selection on these genes. Previous studies showed similar finding in order Ascaridida that all mitochondrial genes, except two (*nad*5 and *cyt*b), showed negative Ks/Ks ratio where *nad*5 and *cyt*b genes showed positive values (Xie et al, 2013). Also, similar trend of negative selection was also found in *Wucheraria bancrofti* mitochondrial genomes (Ramesh et al, 2012) suggesting negative selection pressure against mitochondrial genes of parasitic nematodes.

The sliding window analysis showed higher nucleotide diversity among the *Trichuris* spp. from primates than *T. suis* from pigs that reflects the high species diversity in the *Trichuris* from different primates (Figure 1 and 2). Sliding window analysis could be used to detect regions within the mitochondrial genome with high and low nucleotide diversity that could be used for different purposes (Jex et al, 2010a). For instance, the rDNA genes region in the *T. suis* showed the lowest diversity and therefore could be used for diagnostic purposes (i.e. species identification) unlike other loci with higher diversity such as the *nad*4 gene that could be used for population genetic studies (Figure 2) and similar observation was found in *Trichuris* spp. from primates. Also, the pairwise estimations of nucleotide diversity between the different *Trichuris* spp. from pigs and primates could

identify the regions with high and low diversity for the different purposes as mentioned (Supplementary data Figure S2 and S3).

Concluding remarks

Herein, evidences based on mitochondrial DNA are given on the presence of *Trichuris* species complex in primates and probably also for *Trichuris* in pigs. The presence of several *Trichuris* species in human and non human primates may have important implications on the assessment of zoonotic potential of different *Trichuris* spp. in primates. Moreover, *Trichuris* cryptic species complex in different hosts may have implications on the applied management and control strategies.

The produced mitochondrial genomes of different evolutionary haplotypes of *Trichuris* recovered from primates and pigs repesents rich source of markers that can be used for species identification and diagnostics (phylogenetic studies) or for populations investigations (population genetic studies) for further studies about parasite molecular ecology and genealogy. Mitochondrial genome analysis on other *Trichuris* species such as *T. ovis* or *T. discolor* could clarify whether there are complex species of these parasites as well.

References

Abee C.R., Mansfield, K., Tardif, S.D., Morris, T. (2012). Nonhuman primates in biomedical research (2nd Edition, pp 243-244). London: Elsevier/Academic

Akaike H (1974). "A new look at the statistical model identification." IEEE T Automat Contr, (19):16–723.

Bernt, M., A. Donath, F. Juhling, F. Externbrink, C. Florentz, G. Fritzsch, J. Putz, M. Middendorf and P. F. Stadler (2013). "MITOS: improved de novo metazoan mitochondrial genome annotation." Mol Phylogenet Evol 69(2): 313-319.

Blouin, M. S. (2002). "Molecular prospecting for cryptic species of nematodes: mitochondrial DNA versus internal transcribed spacer." Int J Parasitol 32(5): 527-531.

Broadhurst, M. J., A. Ardeshir, B. Kanwar, J. Mirpuri, U. M. Gundra, J. M. Leung, K. E. Wiens, I. Vujkovic-Cvijin, C. C. Kim, F. Yarovinsky, N. W. Lerche, J. M. McCune and P. Loke (2012). "Therapeutic helminth infection of macaques with idiopathic chronic diarrhea alters the inflammatory signature and mucosal microbiota of the colon." PLoS Pathog 8(11): e1003000.

Callejon, R., S. Nadler, M. De Rojas, A. Zurita, J. Petrasova and C. Cutillas (2013). "Molecular characterization and phylogeny of whipworm nematodes inferred from DNA sequences of *cox*1 mitochondrialDNA and 18S rDNA." Parasitol Res 112(11): 3933-3949.

Cutillas, C., M. de Rojas, A. Zurita, R. Oliveros and R. Callejon (2014). "*Trichuris colobae* n. sp. (Nematoda: Trichuridae), a new species of *Trichuris* from *Colobus guereza kikuyensis*." Parasitol Res 113(7): 2725-2732.

Cutillas, C., R. Callejon, M. de Rojas, B. Tewes, J. M. Ubeda, C. Ariza and D. C. Guevara (2009). "*Trichuris suis* and *Trichuris trichiura* are different nematode species." Acta Trop 111(3): 299-307.

Drummond, A. J. and A. Rambaut (2007). "BEAST: Bayesian evolutionary analysis by sampling trees." BMC Evol Biol 7: 214.

Garg, S. K., A. M. Croft and P. Bager (2014). "Helminth therapy (worms) for induction of remission in inflammatory bowel disease." Cochrane Database Syst Rev 1: CD009400.

Grillo, V., Jackson, F., Cabaret, J. and Gilleard, J. S. (2007). "Population genetic analysis of the ovine parasitic nematode *Teladorsagia circumcincta* and evidence for a cryptic species." Int J Parasitol 37: 435–447

Hahn, C., L. Bachmann and B. Chevreux (2013). "Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads--a baiting and iterative mapping approach." Nucleic Acids Res 41(13): e129.

Hansen, T. V., S. M. Thamsborg, A. Olsen, R. K. Prichard and P. Nejsum (2013). "Genetic variations in the beta-tubulin gene and the internal transcribed spacer 2 region of *Trichuris* species from man and baboons." Parasit Vectors 6(1): 236.

Hu, M., N. B. Chilton and R. B. Gasser (2002). "The mitochondrial genomes of the human hookworms, *Ancylostoma duodenale* and *Necator americanus* (Nematoda: Secernentea)." Int J Parasitol 32(2): 145-158.

Hu, M. and R. B. Gasser (2006). "Mitochondrial genomes of parasitic nematodes--progress and perspectives." Trends Parasitol 22(2): 78-84.

Hurst, L. D. (2002). "The Ka/Ks ratio: diagnosing the form of sequence evolution." Trends Genet 18(9): 486.

Jabbar, A., I. Beveridge, N. Mohandas, N. B. Chilton, D. T. Littlewood, A. R. Jex and R. B. Gasser (2013). "Analyses of mitochondrial amino acid sequence datasets support the proposal that specimens of *Hypodontus macropi* from three species of macropodid hosts represent distinct species." BMC Evol Biol 13: 259.

Jex, A. R., D. T. Littlewood and R. B. Gasser (2010a). "Toward next-generation sequencing of mitochondrial genomes--focus on parasitic worms of animals and biotechnological implications." Biotechnol Adv 28(1): 151-159.

Jex, A. R., R. S. Hall, D. T. Littlewood and R. B. Gasser (2010b). "An integrated pipeline for next-generation sequencing and annotation of mitochondrial genomes." Nucleic Acids Res 38(2): 522-533.

Jex, A. R., A. Waeschenbach, M. Hu, J. A. van Wyk, I. Beveridge, D. T. Littlewood and R. B. Gasser (2009). "The mitochondrial genomes *of Ancylostoma caninum* and *Bunostomum phlebotomum*--two hookworms of animal health and zoonotic importance." BMC Genomics 10: 79.

Jouvin, M. H. and J. P. Kinet (2012). "*Trichuris suis* ova: testing a helminth-based therapy as an extension of the hygiene hypothesis." J Allergy Clin Immunol 130(1): 3-10; quiz 11-12.

Laslett, D. and B. Canback (2008). "ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences." Bioinformatics 24(2): 172-175.

Li, M. W., R. Q. Lin, H. Q. Song, X. Y. Wu and X. Q. Zhu (2008). "The complete mitochondrial genomes for three *Toxocara* species of human and animal health significance." BMC Genomics 9: 224.

Librado, P. and J. Rozas (2009). "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data." Bioinformatics 25(11): 1451-1452.

Liu, G. H., L. Zhao, H. Q. Song, G. H. Zhao, J. Z. Cai, Q. Zhao and X. Q. Zhu (2014). "*Chabertia erschowi* (Nematoda) is a distinct species based on nuclear ribosomal DNA sequences and mitochondrial DNA sequences." Parasit Vectors 7: 44.

Liu, G. H., R. B. Gasser, A. Su, P. Nejsum, L. Peng, R. Q. Lin, M. W. Li, M. J. Xu and X. Q. Zhu (2012a). "Clear genetic distinctiveness between human- and pig-derived *Trichuris* based on analyses of mitochondrial datasets." PLoS Negl Trop Dis 6(2): e1539.

Liu, G. H., R. B. Gasser, P. Nejsum, Y. Wang, Q. Chen, H. Q. Song and X. Q. Zhu (2013). "Mitochondrial and nuclear ribosomal DNA evidence supports the existence of a new *Trichuris* species in the endangered francois' leaf-monkey." PLoS One 8(6): e66249.

Liu, G. H., Y. Wang, M. J. Xu, D. H. Zhou, Y. G. Ye, J. Y. Li, H. Q. Song, R. Q. Lin and X. Q. Zhu (2012b). "Characterization of the complete mitochondrial genomes of two whipworms *Trichuris ovis* and *Trichuris discolor* (Nematoda: Trichuridae)." Infect Genet Evol 12(8): 1635-1641.

Mbaya, A. W. and U. J. Udendeye (2011). "Gastrointestinal parasites of captive and free-roaming primates at the Afi Mountain Primate Conservation Area in Calabar, Nigeria and their zoonotic implications." Pak J Biol Sci 14(13): 709-714.

Messenger, A. M., A. N. Barnes and G. C. Gray (2014). "Reverse zoonotic disease transmission (zooanthroponosis): a systematic review of seldom-documented human biological threats to animals." PLoS One 9(2): e89055.

Nadler, S. A. and Leon, D. E. (2012). "Integrating molecular and morphological approaches for characterizing parasite cryptic species: implications for parasitology." Parasitology 138(13): 1688-1709.

Nissen, S., A. Al-Jubury, T. V. Hansen, A. Olsen, H. Christensen, S. M. Thamsborg and P. Nejsum (2012). "Genetic analysis of *Trichuris suis* and *Trichuris trichiura* recovered from humans and pigs in a sympatric setting in Uganda." Vet Parasitol 188(1-2): 68-77.

Pittman JS, Shepherd G, Thacker BJ, Myers GH (2010). "*Trichuris suis* in finishing pigs: Case report and review." J Swine Health Prod.18(6): 306–313.

Posada, D. (2008). "jModelTest: phylogenetic model averaging." Mol Biol Evol 25(7): 1253-1256.

Pullan, R. L., J. L. Smith, R. Jasrasaria and S. J. Brooker (2014). "Global numbers of infection and disease burden of soil transmitted helminth infections in 2010." Parasit Vectors 7: 37.

Ramesh, A., S. T. Small, Z. A. Kloos, J. W. Kazura, T. B. Nutman, D. Serre and P. A. Zimmerman (2012). "The complete mitochondrial genome sequence of the filarial nematode *Wuchereria bancrofti* from three geographic isolates provides evidence of complex demographic history." Mol Biochem Parasitol 183(1): 32-41.

Ravasi, D. F., M. J. O'Riain, F. Davids and N. Illing (2012). "Phylogenetic evidence that two distinct *Trichuris* genotypes infect both humans and non-human primates." PLoS One 7(8): e44187.

Reddy, A. and B. Fried (2007). "The use of *Trichuris suis* and other helminth therapies to treat Crohn's disease." Parasitol Res 100(5): 921-927.

Rothman, J.M., Bowman, D.D., Eberhard, M.L. and Pell, A.N. (2002)."Intestinal parasites found in the research group of mountain gorillas in Bwindi Impenetrable National Park, Uganda: preliminary results." Ann. N.Y. Acad. Sci. 969: 346–349.

Schattner, P., A. N. Brooks and T. M. Lowe (2005). "The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs." Nucleic Acids Res 33(Web Server issue): W686-689.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0." Mol Biol Evol 30(12): 2725-2729.

Xie, Y., L. Niu, B. Zhao, Q. Wang, X. Nong, L. Chen, X. Zhou, X. Gu, S. Wang, X. Peng and G. Yang (2013). "Complete mitochondrial genomes of chimpanzee- and gibbon-derived *Ascaris* isolated from a zoological garden in southwest China." PLoS One 8(12): e82795.

Supplementary data of manuscript 2

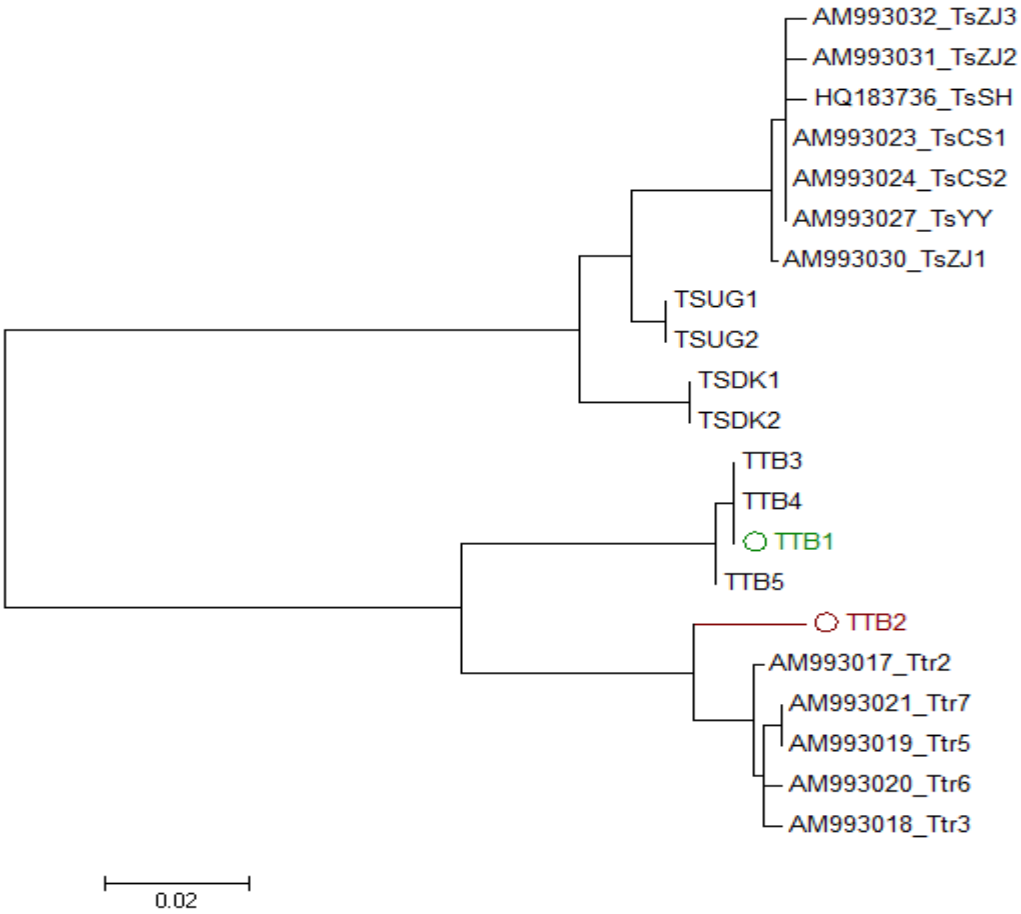

Figure S1. The NJ phylogenetic tree of mitochondrial *rrn*L gene for 5 *Trichuris* worms from baboons indicated as TTB1→TTB5 and 4 *T. suis* worms, two from Uganda (TSUG1 and 2) and 2 from Denmark (TSDK1 and 2). More sequences from China for *T. suis* from pigs and *T. triciura* from humans were included from the GenBank with the their accession numbers given.

Table S1. The Ka/Ks ratio between different *Trichuris* spp from different primates. TTB1 and TTB2 from baboons, TTCh from human in China (GU385218), T.GHL from Francois leaf monkey (KC461179).

Genes	Non-synonymous/synonymous substitutions (Ka/Ks)					
	TTB1/TTB2	TTB1/TTCh	TTB1/T.GHL	TTB2/TTCh	TTB2/T.GHL	T. trChina/T.GHL
***cox*1**	0.422	0.376	0.044	0.11	0.048	0.057
***cox*2**	0.028	0.067	0.039	0.057	0.059	0.066
***cox*3**	0.058	0.047	0.101	0.048	0.096	0.121
***cyt*b**	0.117	0.116	0.084	0.098	0.074	0.094
***nad*1**	0.082	0.08	0.106	0.077	0.108	0.128
***nad*2**	0.100	0.122	0.156	0.125	0.118	0.118
***nad*3**	0.093	0.121	0.161	0.127	0.088	0.114
***nad*4**	0.122	0.135	0.144	0.127	0.121	0.121
***nad*4L**	0.073	0.119	0.125	0.113	0.178	0.182
***nad*5**	0.150	0.161	0.194	0.132	0.179	0.166
***nad*6**	0.104	0.114	0.144	0.103	0.158	0.172
***atp*6**	0.2	0.182	0.265	0.311	0.218	0.170
***atp*8**	0.187	0.252	0.261	0.204	0.243	0.359

Table S2. The Ka/Ks ratio between different *T. suis* from pigs in different countries (TSDK from Denmark, TSUG from Uganda and TSCH from China (GU070737).

Genes	Non-synonymous/synonymous substitutions (Ka/Ks)		
	TSDK/TSUG	TSDK/TSCH	TSUG/TSCH
***cox*1**	0.039	0.035	0.045
***cox*2**	0.038	0.053	0.000
***cox*3**	0.092	0.094	0.078
***cyt*b**	0.065	0.056	0.056
***nad*1**	0.072	0.078	0.089
***nad*2**	0.168	0.170	0.145
***nad*3**	0.042	0.055	0.038
***nad*4**	0.238	0.244	0.550
***nad*4L**	0.082	0.084	0.104
***nad*5**	0.139	0.112	0.186
***nad*6**	0.083	0.064	0.069
***atp*6**	0.188	0.207	0.106
***atp*8**	0.321	0.418	0.508

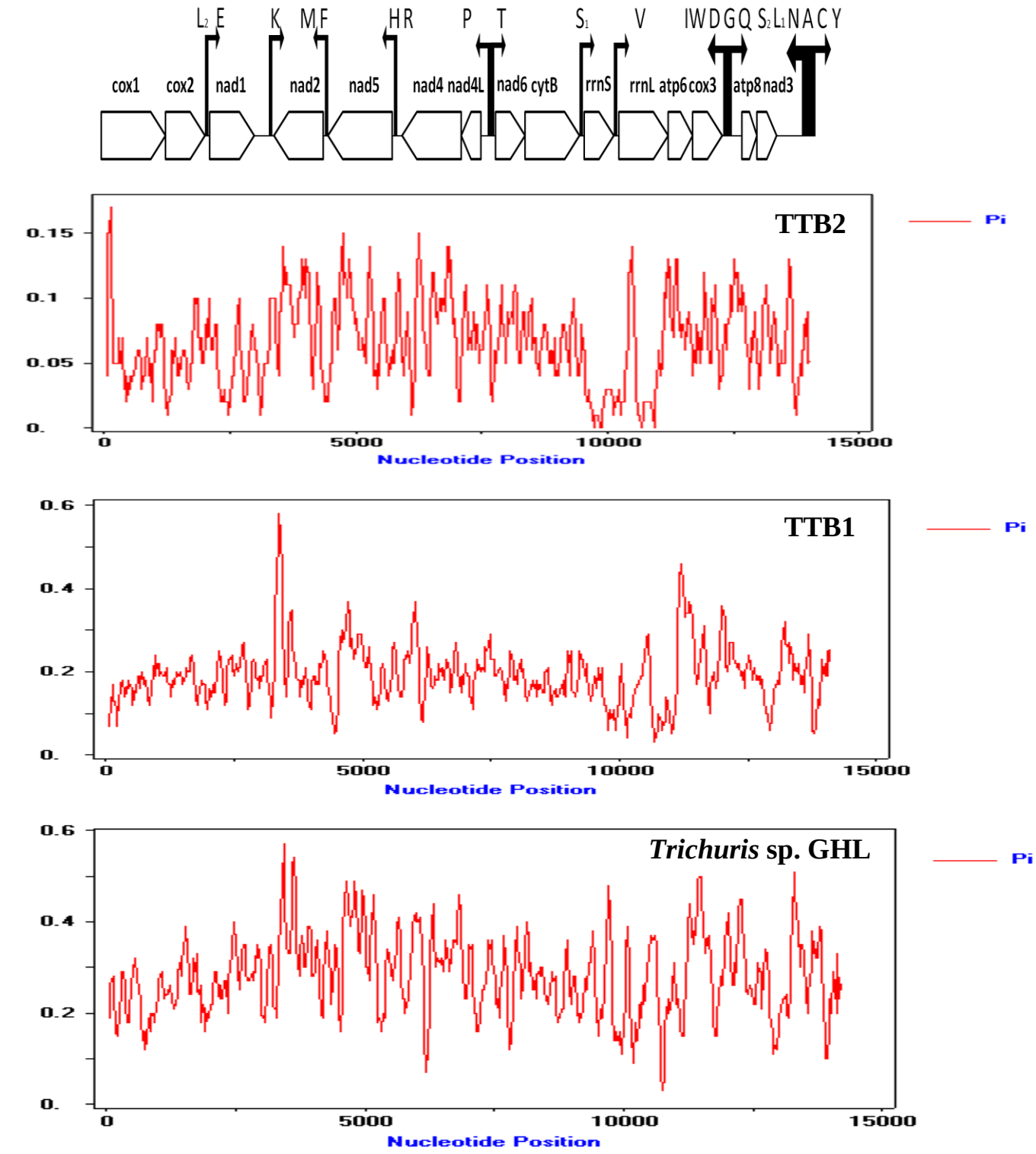

Figure S2. The pairwise nucleotide diversity (pi) across the mitochondrial genome of *T. trichiura* from China (GU385218) with *Trichuris* of TTB1 and TTB2 and the *Trichuris* sp. GHL from Francois leaf monkey (KC461179) as shown in the figure. Solid arrows indicate the tRNA.

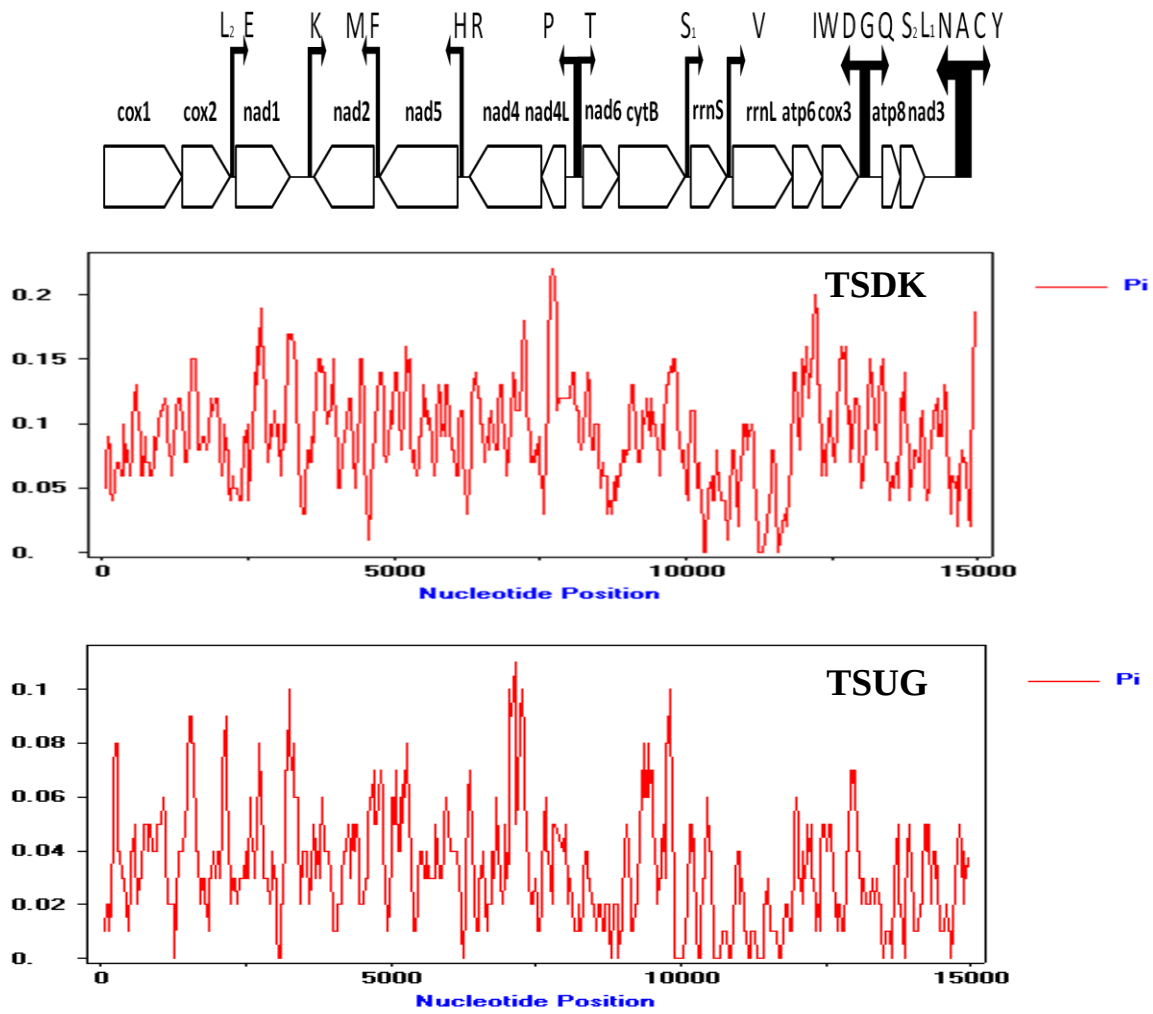

Figure S3. The pairwise nucleotide diversity (pi) across the mitochondrial genome of *T. suis* from from China (TSCH)(GU070737) with *T. suis* from Denmark (TSDK) and Uganda (TSUG). Solid arrows indicate the tRNA.

Appendix of raw data

The CD below contains the sequecnes of the *nad*1 and *rrn*L genes used in the thesis's studies besides the full annotated mitochondrial genomes for the *Trichuris* from baboons TTB1 and TTB2 and from pigs TSUG and TSDK.