Ascaris phylogeny based on multiple whole mtDNA genomes

Peter Nejsum[a,b,*,1], Mohamed B. F. Hawash[c,1], Martha Betson[d], J. Russell Stothard[e], Robin B. Gasser[f], Lee O. Andersen[g]

[a] *Department of Clinical Medicine, Health, Aarhus University, Denmark*
[b] *Department of Veterinary Disease Biology, Faculty of Health Sciences, University of Copenhagen, Denmark*
[c] *Department of Genetics, Centre Hospitalier Universitaire Sainte-Justine Research Center, Montréal, Canada*
[c] *School of Veterinary Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom*
[e] *Department of Parasitology, Liverpool School of Tropical Medicine, Liverpool, United Kingdom*
[f] *Department of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, Victoria, Australia*
[g] *Department of Microbiology and Infection Control, Statens Serum Institut, Copenhagen, Denmark*

*Corresponding author. Peter Nejsum, Department of Clinical Medicine, Health, , Aarhus University, Palle Juul-Jensens Boulevard 99, 8200 Aarhus N, Denmark. Phone: +45 5054 1392. E-mail: pn@clin.au.dk

[1] Joint first authors

ABSTRACT

Ascaris lumbricoides and *A. suum* are two parasitic nematodes infecting humans and pigs, respectively. There has been considerable debate as to whether *Ascaris* in the two hosts should be considered a single or two separate species. Previous studies identified at least three major clusters (A, B and C) of human and pig *Ascaris* based on partial *cox*1 sequences. In the present study, we selected major haplotypes from these different clusters to characterize their whole mitochondrial genomes for phylogenetic analysis. We also undertook coalescent simulations to investigate the evolutionary history of the different *Ascaris* haplotypes. The topology of the phylogenetic tree based on complete mitochondrial genomic sequences was found to be similar to partial *cox*1 sequencing, but the support at internal nodes was higher in the former. Coalescent simulations suggested the presence of at least two divergence events: the first one occurring early in the Neolithic period which resulted in a differentiated population of *Ascaris* in pigs (cluster C), the second occurring more recently (~900 generations ago), resulting in clusters A and B which might have been spread worldwide by human activities.

Keywords: Ascaris, mitochondrial genomes, human, pig, phylogeny, soil transmitted helminth

1. Introduction

Ascaris lumbricoides (Linnaeus, 1758) in humans and *A. suum* (Goeze, 1782) in pigs are prevalent parasitic nematodes. About a billion people are infected, and *A. suum* is found among pigs globally in both intensive and extensive production systems (Bethony et al., 2006; Hotez and Kamath, 2009; Roepstorff et al., 2011, Nissen et al., 2011, Pullan et al., 2014). There has been ongoing debate as to whether *Ascaris* from humans and pigs represents a single or separate species (e.g., Anderson 2001, Leles et al., 2012, Søe et al., 2016). However, there is no doubt that *Ascaris* from the two host species are genetically very closely related and that both host species can be cross infected (reviewed by Nejsum et al., 2012). This may relate to a short evolutionary history of *Ascaris* in humans and pigs as a host switch is expected to have taken place during domestication of pigs ~10,000 years ago (Cox 2004; Leles et al., 2010; Araujo et al., 2008; Brinkkemper and Haaster, 2012; Mitchell, 2013).

In order to assess the genetic relationship of *Ascaris* individuals from the two host species, sequencing of part of the mitochondrial (mt) *cox*1 gene (383 bp) has been used in multiple studies (e.g. Peng et al, 2005; Cavallero et al., 2013; Betson et al., 2014). In these studies, three main haplotype clusters have been identified (A, B and C). Interestingly, worms from the two host species are found in both cluster A and B, whereas worms from pigs (or humans known to be cross-infected with pig worms) are represented in cluster C. Despite the fact that shared haplotypes between worms from the two host species have been observed, most worms from either humans or pigs are found together in either cluster A or B (Cavallero et al., 2013; Betson et al., 2014). It has also been noted that haplotypes belonging to clusters A and B are found worldwide in both host species, but in different proportions. Hence, most worms from humans and pigs in China have been found to represent cluster B (Peng et al., 2005), whereas most worms in these hosts from Uganda were found in cluster A (Betson et al., 2014), but with no clear geographical association. However, it is not known whether the relationships inferred based on *cox*1 reflect that of other genes in the mt genome. In the present study, we conducted whole mitochondrial genomic haplotyping of six *Ascaris* representing five distinct *cox*1 haplotypes recorded in Betson et al. (2014) for a comparative evolutionary analysis. In addition, coalescent simulation were performed on human and pig *Ascaris* population to gain insights into the deep evolutionary history of the parasite.

2. Methods

2.1. Genomic DNA isolation and cox1 haplotypes

Ascaris DNA samples included in this study have previously been described (Betson et al., 2014). Part of the *cox*1 gene (383 bp) was sequenced and three main haplotype clusters (A, B and C) were defined by phylogenetic analysis (Betson et al., 2014). For the present study, we selected DNA samples from *Ascaris* individuals representing the most common haplotypes H01, H03, H07, H28 and H64 for full mtDNA genome sequencing. Haplotypes H01 and H03 originated from two persons from Tanzania and were identified as *A. lumbricoides* by microsatellite analysis (Betson et al. 2014); the other samples were from pigs from Uganda (H01 and H07), Tanzania (H28) and the UK (H64) and were identified as *A. suum*. The Uganda pig worm H01 was later identified as a cross-infection (Betson et al., 2014) and therefore most likely represents *A. lumbricoides*.

2.2. Long-range PCR amplification of the mtDNA genomes and sequencing

The primers given in Table 1 were designed using Primer 3 (Koressaar and Remm, 2007; Untergasser et al., 2012) and used to amplify the mt genomes in five overlapping fragments by long range PCR. PCR cycling conditions were the same for all primer sets with an initial denaturation at 92°C for 4

min, followed by 35 cycles of denaturation at 92 °C (20 sec), annealing at 55 °C for 30 sec, extension at 62 °C for 5 min and a final extension at 62 °C for 10 min. The long-range PCR was conducted in a 20 µl-volume using a standard buffer, 0.2 mM of each dNTP, 0.4 mM of each primer pair, 2.0 mM $MgCl_2$, and 2.5 U of the Long PCR Enzyme Mix (Thermo Scientific). PCR products were detected by gel electrophoresis (0.8% agarose) using GelRed™ (Biotium) as the stain over ultraviolet light. Aliquots of amplicons (5 µl) were each treated with 1 µl Exonuclease I (Fermentas) and 2 µL FastAP thermosensitive alkaline phosphatase (1 U/µl) (Fermentas) at 37 °C for 15 min, and the enzymes inactivated at 85 °C for 15 min. DNA concentrations were measured spectro-photometrically (NanoDrop 1000, Thermo Fischer Scientific). Individual DNA libraries were constructed and sequenced using Illumina HiSeq 2000 by Macrogen Inc., South Korea.

2.3. Assembly and annotation

Sequence reads (~100 bp) of each genome were assembled using the CLC Genomics Workbench v6.5.1 (CLC Inc, Aarhus, Denmark). Raw data can be provided upon request. Open reading frames were identified using the CLC Genomics Workbench and the BLASTx search tool embedded in the program used to identify the genes. tRNAScan-SE (Schattner et al., 2005) was used to identify tRNAs whereas rRNAs were identified using the BLAST search tools available through NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

The following reference sequences were obtained from the GenBank database for comparative analyses: *A. suum* from USA (NC_001327); *A. suum* from China (HQ704901); *A. lumbricoides* from China (HQ704900); *A. lumbricoides* from unknown location (JN801161); *Ascaris* from Gibbon (KC839987) and a chimpanzee (KC839986) from China.

2.4. Phylogenetic analysis and genetic variation

The 12 protein coding genes and the two ribosomal RNA (rDNA) genes were extracted and aligned using MUSCLE (Edgar, 2004). *Baylisascaris procyonis* from China (NC_016200) was used as outgroup in phylogenetic analyses. A second dataset containing only the 383 bp of the *cox*1 gene used in previous studies was also employed for comparative phylogenetic analysis. Maximum Likelihood (ML) and Maximum parsimony (MP) trees were built using MEGA v6.1 (Tamura et al., 2007) employing 1000 bootstraps to test the stability of the topology. For ML the best-to-fit substitution models were identified using jModelTest2 (Darriba et al., 2012) under Akaike information criterion (AIC) (Akaike, 1974) for each dataset. The Tamura 3-parameter model with gamma distribution and invariant sites was applied to all mt protein-encoding gene sequences, and the Hasegawa-Kishino-Yano model with invariant sites to the partial mt *cox*1 sequences. For MP, a heuristic search using tree bisection-reconnection (TBR) branch swapping was used, with an addition of 10 initial random trees. Bayesian inference (BI) was conducted using BEAST v. 1.6.1. (Drummond and Rambaut, 2007). Log-normal was used as a prior and General Time Reversible (GTR) model with gamma distribution was used as the substitution model. A random starting tree with Yule prior was assumed. Three independent runs with 10 million steps each, with a burn-in of 10,000 steps, were carried out. Tracer v.1.6 (Drummond and Rambaut, 2007) was used to analyze log files of the MCMC chains, and the reliability of parameters was verified by recording effective sample sizes values of > 200. Tree Annotater v.1.6.1 (Drummond and Rambaut, 2007) was used to summarize the tree data, with a posterior probability (pp) limit of 0.5. MEGA was used to estimate the p-distances between the difference clades identified in the phylogenetic analyses.

The genetic variation between the different genomes was characterized in terms of the number of point mutations in the protein-encoding genes. The program SNP sites (Page et al., 2016) was used

to identify the different mutations in each genome using human *Ascaris* from China (HQ704900) as the reference mt genome. The number of mutations per mt genome and per mt gene, and the number of unique mutations (only found in one mt genome) was identified by multiple-wise alignment by MUSCLE (Edgar, 2004). The program DnaSP (Librado and Rozas, 2009) was used to identify synonymous and non-synonymous mutations and the ratio between them, in order to predict the pattern of selection on mt genes.

2.5 Demography and history of Ascaris spp.

To identify the most recent common ancestor (TMRCA) of the three major clusters we estimated the effective population size (Ne) as this number is equal to TMRCA for uniparentally inherited DNA. The formula $\Theta=2Ne\mu$, where Θ (theta) is the genetic diversity of a population, μ is the mutation rate per gene/genome and Ne is the effective population size and therefore in this case, TMRCA. Genetree (Bahlo and Griffiths, 2000) was used to estimate Theta (Θ) using the 12 mt protein-encoding genes as described (Hawash et al., 2016).

The mutation rate in the mt genome of free-living nematode *Caenorhabditis elegans* (1.6×10^{-7} per site per generation) (Denver et al., 2000) was used as an estimate for *Ascaris*, as no information is available for parasitic worms. The mutation rate in each mt genome was obtained by multiplying the mutation rate with the length of the 12 mt protein-encoding genes (10,288 bp), giving 0.00165 mutations per generation.

Using an "isolation and migration" model implemented in IMa2 (Hey and Nielsen, 2004), TMRCA and time since divergence were estimated using the 12 protein coding genes. The priors used for the protein coding genes were t = 70 (upper bound of splitting time) and q = 180 (upper bound of population size); no migration after isolation was allowed in the out model (m=0). The Hasegawa-Kishino-Yano substitution model was used for the protein coding genes, while the stepwise mutation model was used for microsatellite markers. 20 Markov chains with geometric heating scheme (first and second heating parameters being 0.96 and 0.90, respectively) and 10^6 burn-in steps with 10^5 sampling genealogies were used. Three independent runs were conducted with different seed numbers to assess the convergence.

3. Results

The mt genomes were sequenced at 80-150-times coverage. The mt genomes of the two human *Ascaris* haplotypes H01 and H03 from Tanzania were 14,280 and 14,274 bp in length, respectively. The mt genomes of the four pig derived *Ascaris* representing H01 (Uganda), H07 (Uganda), H28 (Tanzania) and H64 (United Kingdom) were 14,151, 14,320, 14,210 and 14,187 bp in length, respectively. The genomes contain 12 mt protein-encoding genes, 22 genes for transfer RNAs and two ribosomal RNA genes. Details on individual genomes, including gene annotations, can found in GenBank under the following accession numbers: KY045800-KY045805.

The phylogeny based on the mt genomic sequence data sets (i.e. 12 protein-encoding genes and the two rRNA genes) are depicted in Fig. 1a. All three tree clustering methods gave the same topology. Three main clusters (A, B and C) could be identified, with cluster C being basal to the two other clusters. The p-distances between cluster C and cluster A+B ranged from 0.048-0.051, whereas the p-distance between clusters A and B was in the range of 0.013-0.021. The topology of the phylogenetic tree based on the mt genomic data sets was similar to that obtained using partial *cox*1 sequence data (except for the *A. suum* from China) (Fig. 1b), but the nodes were robust for the former.

The p-distances based on partial *cox*1 sequences were higher between clusters A and C than between clusters B and C (range: 0.047-0.052 and 0.037-0.039, respectively).

Among the 12 *Ascaris* genomes, a total of 791 mutations were identified in the 12 protein-encoding genes (Table 2); 496 of these mutations were found to be unique to a single mt genome with highest numbers in pig *Ascaris* from UK representing cluster C (Table 2). No mutations were found to be specific to *Ascaris* from either human or pig. The number of mutations per gene length was found to be highest in the *nad*4 gene and lowest in the *cox*1 gene (Table 3). The non-synonymous/synonymous (Ka/Ks) mutation ratio indicates that purifying selection is acting on all of the protein coding genes. All SNPs detected are listed in Supplementary file S1.

Based on the 12 protein-encoding genes, theta was estimated at 223 (standard deviation: 82.35) using Genetree (Bahlo and Griffiths. 2000). The effective population size and thereby the TMRCA is thus 63,512 (±23,000) generations. A similar result was obtained when using IMa2 (Hey and Nielsen, 2004), where the most recent common ancestor was estimated to occur 58,238 generations ago, supported by pp estimates (Fig. 2).

The time of divergence between the three main clusters was estimated using IMa2. The first divergence event between cluster A+B and C was ~7500 generations (Fig. 2), whereas the divergence between clusters A and B was 913 generations (Fig. 2). However, the recent divergence was not supported by pp-values, likely due to low genetic structure.

4. Discussion

In this study, we used a long range PCR-coupled approach to directly sequence 6 human and pig *Ascaris* mt genomes representing five major haplotypes identified previously among worms from worldwide locations (cf. Betson et al., 2014). The genes and gene order, including the 22 tRNAs, were as described previously for *Ascaris* (cf. Okimoto et al., 1994; Liu et al., 2012). We found that the phylogeny based on mt genomic data also identified three major clusters, in accordance with that based on partial *cox*1 sequence data (Cavallero et al., 2013; Betson et al., 2014). The topology of the phylogenetic trees are similar with cluster C, basal to the two other clusters, but where *A. suum* from China (HQ704901) are found in cluster A when based on all mtDNA genes, it is found between cluster A and B in the *cox*1 tree (Fig. 1). The genetic distance between cluster A+B and cluster C was similar using data for all mt genes, whereas partial *cox*1 sequencing suggest that cluster B is more basal than cluster A and more closely related to cluster C (Fig. 1).

Although mt *cox*1 is used in many population genetic and phylogenetic studies of *Ascaris* (e.g., Peng et al, 2005; Cavallero et al, 2013; Betson et al, 2014), it was less variable in sequence compared to *nad*4, suggesting that the latter gene may therefore be useful for future studies. We observed a negative selection on the protein-encoding genes (Table 3), which has also been reported for other parasitic nematodes (Ramesh et al., 2012; Hawash et al., 2015). In accordance with previous studies (e.g. Betson et al., 2014), no mutations were found to be characteristic to *Ascaris* from humans or *Ascaris* from pigs; thus there were no mutations of value to distinguish *Ascaris* from the two host species. However, mt genomic data might be informative in delineating the genetic structure based on the 3 clusters (A, B and C) which could reflect a complex demography and evolutionary history as discussed below.

Ascaris representing clusters A and B have been identified in both humans and pigs from worldwide locations, including both sympatric areas as well as regions where transmission among individuals of only one of the host species is expected. In this way, worms belonging to both cluster A and B have been found in pigs in Denmark and United Kingdom, where humans not are expected to contribute to transmission, and in humans living in areas with few or no pigs, including Zanzibar and Bangladesh (cf. Betson et al, 2014). However, the frequencies of haplotypes vary among location.

In this way, 82.9% and 57.8% of the human (n=392) and pig (n=331) worms from China belonged to cluster B whereas the majority of the Uganda worms from humans and pig belonged to cluster A (82.5% and 64.5%, respectively). Cluster C is interesting, as this haplotype only has been identified in pigs from Europe and Africa using *cox*1 sequencing. In this way, it was not identified among 723 human and pig worms in China (Peng et al., 2005), 51 human and pig samples in Ecuador (Iñiguez et al., 2012) or 255 humans in Uganda, Kenya, Zambia or on Zanzibar (Betson et al., 2014). Using a combination of restriction mapping and sequencing of the mtDNA, Anderson and Jaenike (1997) also only identified pig *Ascaris* from Europe in cluster C, except for a single human worm from Guatemala.

In contrast to the mtDNA, where no clear geographic or host separation is observed, host or geographic differentiation between *Ascaris* populations is observed when microsatellite markers are used, but with contrasting results. In the study by Betson et al. (2014), *Ascaris* were first segregated by host and then by geographical origin, suggesting a single host shift, followed by geographical separation. This finding implies that the different haplotypes belonging to clusters A and B were present before the host shift and were subsequently distributed worldwide. In contrast, Criscione et al. (2007) found evidence for a geographical separation of *Ascaris* populations, followed by multiple host shifts. As there is genetic evidence that domestication may have taken place multiple times (Larson et al., 2005), this information supports the multiple host shift model. In this situation, the haplotypes were first distributed worldwide and then transmitted to the new host species. This proposal may explain why *Ascaris* representing cluster A predominates in Uganda, whereas *Ascaris* representing cluster B predominates in China.

The difference in the population genetic structures given by mtDNA or microsatellite analysis may be attributed to the difference in their mutation rates. As estimated in free-living nematodes (Molnar et al., 2011; Molnar et al., 2012), the mutation rate of microsatellites is faster than mtDNA. Hence, mtDNA could be useful for capturing ancient historical events while microsatellite markers are informative for relatively recent evolutionary events. This may be the reason why worms from all three mtDNA clusters can be found among Danish pigs but still belong to same population based on microsatellite analysis (Betson et al., 2014). Moreover, the finding of shared haplotypes and worms from both host belonging to cluster A and B may also be explained by recent or current introgression supported the observation of cross-infection and hybrids in sympatric areas (Criscione et al., 2007; Zhou et al., 2012; Betson et al., 2014). Incomplete lineage sorting may therefore explain the observed mtDNA pattern most likely due to very short evolutionary history of *Ascaris* in the two host species as discussed below.

It is expected that *Ascaris* host shift occurred during domestication about 10,000 years ago but it is unknown whether the direction was from humans to pigs or the other way around. However, as *Ascaris* eggs have been found in human archeological samples from France which dates back before domestication (30,160-24,660) (Loreille and Bouchet, 2003) and as eggs have been identified in human samples in America before introduction of pigs to the continent (reviewed by Goncalves et al., 2003) this suggests a transmission from humans to pigs. The coalescent simulations on the mtDNA suggested that the time to the most recent common ancestor is ~60,000 generations and that at least two divergence events have happened (Fig. 2). The first one happened at the beginning of the Neolithic period (15,000 – 2,500 years ago, given a generation time of ½-3 generations/year) and resulted in cluster C. This may have followed by an isolation event of one pig population that explains why *Ascaris* cluster C is exclusive in European pigs (except for zoonotic infection in humans). Alternatively, cluster C have circulated in specific isolated human and pig populations but later was lost from the human *Ascaris* population due to severe bottleneck. The second divergence has happened more recently (1,800-300 years ago) and resulted in cluster A and B and may have been followed by massive dispersal globally of haplotypes by human activities such as trading,

transportation and colonization. However, these data cannot rule out the alternative hypothesis with a single host shift about ~10,000 years ago followed by a more recent admixture of *Ascaris* populations due to human activities including transport of pigs. This may also explain the complex phylogenetic picture we see based on mtDNA. The mtDNA and microsatellite markers may however also be complementary to each other with mtDNA providing information on more ancient evolutionary events and microsatellite markers on more recent history.

Despite intensive research the taxonomic status of *Ascaris* in humans and pigs still needs to be elucidated. Several studies have shown that *Ascaris* in the two hosts species can cross-infect (Nejsum et al., 2005; Criscione et al., 2007; Peng et al., 2012; Betson et al., 2014) and hybrids have been identified in sympatric areas (Criscione et al., 2007; Zhou et al., 2012). In addition, the phylogeny based on mtDNA as described in this and other studies (e.g. Betson et al. 2014) may suggest that *Ascaris* in humans and pigs represents a single species. However, sympatric populations seem to be genetic differentiated (Anderson and Jaenike, 1997; Criscione et al., 2007; Betson et al., 2014) suggesting that there is no, or very restricted contemporary gene-flow between populations, supporting the idea of two separate species according to the biological species concept. Future studies should apply next-generation sequencing to *Ascaris* populations from both host species from locations worldwide, in order to illuminate the evolution, dispersal and the taxonomic status of this enigmatic parasite.

Competing interests

The authors have declared that no financial and non-financial competing interests exist.

Acknowledgements

This work was supported by a Danish Agency for Science, Technology and Innovation grant awarded to PN. RBG's research is funded mainly through grants from the Australian Research Council (ARC) and the National Health and Medical Research Council (NHMRC) of Australia. MB is supported by University of Surrey. The original analysis of the samples reported in Betson et al. (2014) was supported by the Wellcome Trust (grant 085440) and the Natural History Museum, London. Thanks to Helena Ngowi (Sokoine University of Agriculture, Tanzania), Ida-Hella Poulsen and Sofie Nissen (University of Copenhagen, Denmark) for providing worms.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19, 716–723.

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

Anderson, T.J., Jaenike, J., 1997. Host specificity, evolutionary relationships and macrogeographic differentiation among *Ascaris* populations from humans and pigs. Parasitol. 115, 325-342.

Araujo, A., Reinhard, K.J., Ferreira, L.F., Gardner, S.L., 2008. Parasites as probes for prehistoric human migrations? Trends Parasitol. 24, 112-115.

Bahlo M., Griffiths R.C., 2000. Inference from gene trees in a subdivided population. Theor. Popul. Biol. 57, 79-95.

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

Betson, M., Nejsum, P., Bendall, R.P., Deb, R.M., Stothard, J.R., 2014. Molecular epidemiology of ascariasis: a global perspective on the transmission dynamics of *Ascaris* in people and pigs. J. Infect. Dis. 210, 932-941.

Brinkkemper O., van Haaster H., 2012. Eggs of intestinal parasites whipworm (*Trichuris*) and mawworm (*Ascaris*): Non-pollen palynomorphs in archaeological samples. Rev. Palaeobot. Palynol. 186, 16–21.

Cavallero, S., Snabel, V., Pacella, F., Perrone, V., D'Amelio, S., 2013. Phylogeographical studies of *Ascaris* spp. based on ribosomal and mitochondrial DNA sequences. PLoS Negl. Trop. Dis. 7, e2170.

Cox, F.E., 2004. History of human parasitic diseases. Infect. Dis. Clin. North. Am. 18, 171-188.

Criscione, C.D., Anderson, J.D., Sudimack, D., Peng, W., Jha, B., Williams-Blangero, S., Anderson, T.J.C., 2007. Disentangling hybridization and host colonization in parasitic roundworms of humans and pigs. Proc. Biol. Sci. 274, 2669-2677.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772.

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

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

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

Goncalves, M.L., Araujo, A., Ferreira, L.F., 2003. Human intestinal parasites in the past: new findings and a review. Mem. Inst. Oswaldo. Cruz. 98 Suppl 1, 103-118.

Hawash, M.B., Andersen, L.O., Gasser, R.B., Stensvold, C.R., Nejsum, P., 2015. Mitochondrial genome analyses suggest multiple *Trichuris* species in humans, baboons, and pigs from different geographical regions. PLoS Negl. Trop. Dis. 9, e0004059.

Hawash, M.B., Betson, M., Al-Jubury, A., Ketzis, J., LeeWillingham, A., Bertelsen, M.F., Cooper, P.J., Littlewood, D.T., Zhu, X.Q., Nejsum, P., 2016. Whipworms in humans and pigs: origins and demography. Parasit. Vectors 9, 37.

Hey, J., Nielsen, R., 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167, 747-760.

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

Iñiguez A.M., Leles D., Jaeger L.H., Carvalho-Costa F.A., Araújo A., Amazonas Research Group, 2012. Genetic characterisation and molecular epidemiology of *Ascaris* spp. from humans and pigs in Brazil. Trans. R. Soc. Trop. Med. Hyg. 106, 604– 612.

Koressaar T., Remm M., 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23, 1289-1291

Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., Rowley-Conwy, P., Andersson, L., Cooper, A., 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science 307, 1618-1621.

Leles, D., Gardner, S.L., Reinhard, K., Iniguez, A., Araujo, A., 2012. Are *Ascaris lumbricoides* and *Ascaris suum* a single species? Parasit. Vectors 5, 42.

Leles D., Fugassa M, Ferreira LF, Iniguez AM, Araújo A., 2010. A parasitological paradox: why is ascarid infection so rare in the prehistoric Americas? J. Archaeol. Sci. 37, 1510–1520.

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

Liu, G.H., Wu, C.Y., Song, H.Q., Wei, S.J., Xu, M.J., Lin, R.Q., Zhao, G.H., Huang, S.Y., Zhu, X.Q., 2012. Comparative analyses of the complete mitochondrial genomes of *Ascaris lumbricoides* and *Ascaris suum* from humans and pigs. Gene 492, 110-116.

Loreille, O., Bouchet, F., 2003. Evolution of ascariasis in humans and pigs: a multi-disciplinary approach. Mem. Inst. Oswaldo. Cruz. 98 Suppl 1, 39-46.

Mitchell D.P., 2013. The origins of human parasites: exploring the evidence of endoparasitism throughout human evolution. Int. J. Paleopathol. 3, 191–198.

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

Molnar, R.I., Witte, H., Dinkelacker, I., Villate, L., Sommer, R.J., 2012. Tandem-repeat patterns and mutation rates in microsatellites of the nematode model organism *Pristionchus pacificus*. G3 2, 1027-1034.

Nejsum, P., Betson, M., Bendall, R.P., Thamsborg, S.M., Stothard, J.R., 2012. Assessing the zoonotic potential of *Ascaris suum* and *Trichuris suis*: looking to the future from an analysis of the past. J. Helminthol. 86, 148-155.

Nejsum, P., Parker, E.D., Frydenberg, J., Roepstorff, A., Boes, J., Haque, R., Astrup, I., Prag, J., Sørensen, U.B.S., 2005. Ascariasis is a zoonosis in Denmark. J. Clin. Microbiol. 43, 1142–1148.

Nissen, S., Poulsen, I.H., Nejsum, P., Olsen, A., Roepstorff, A., Rubaire-Akiiki, C., Thamsborg, S.M., 2011. Prevalence of gastrointestinal nematodes in growing pigs in Kabale District in Uganda. Trop. Anim. Health Prod. 43, 567-572.

Okimoto, R., Macfarlane, J.L., Clary, D.O., Wolstenholme, D.R., 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics 130, 471-498.

Page, A.J., Taylor, B., Delaney, A.J., Soares, J., Seemann, T., Keane, J.A., Harris, S.R., 2016. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. bioRxiv, http://dx.doi.org/10.1101/038190.

Peng, W., Yuan, K., Hu, M., Zhou, X., Gasser, R.B., 2005. Mutation scanning-coupled analysis of haplotypic variability in mitochondrial DNA regions reveals low gene flow between human and porcine *Ascaris* in endemic regions of China. Electrophoresis 26, 4317-4326.

412 Pullan, R.L., Smith, J.L., Jasrasaria, R., Brooker, S.J., 2014. Global numbers of infection and disease
413 burden of soil transmitted helminth infections in 2010. Parasit. Vectors 7, 37.
414 Ramesh, A., Small, S.T., Kloos, Z.A., Kazura, J.W., Nutman, T.B., Serre, D., Zimmerman, P.A.,
415 2012. The complete mitochondrial genome sequence of the filarial nematode *Wuchereria*
416 *bancrofti* from three geographic isolates provides evidence of complex demographic history.
417 Mol. Biochem. Parasitol. 183, 32-41.
418 Roepstorff, A., Mejer, H., Nejsum, P., Thamsborg, S.M., 2011. Helminth parasites in pigs: new
419 challenges in pig production and current research highlights. Vet. Par. 180, 72-81.
420 Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers
421 for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33, W686-689.
422 Søe, M.J., Kapel, C.M., Nejsum, P., 2016. *Ascaris* from humans and pigs appear to be reproductively
423 isolated species. PLoS Negl. Trop. Dis. 10, e0004855.
424 Tamura K., Stecher G., Peterson D., Filipski A., Kumar S., 2013. MEGA6: Molecular Evolutionary
425 Genetics Analysis version 6.0. Mol. Biol. Evol. 30, 2725-2729.
426 Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M., Rozen S.G., 2012.
427 Primer3 - new capabilities and interfaces. Nucleic Acids Res. 40, e115.
428 Zhou, C., Li, M., Yuan, K., Deng, S., Peng, W., 2012. Pig *Ascaris*: an important source of human
429 ascariasis in China. Infect. Genet. Evol. 12, 1172-1177.
430

Legends to figures

Fig. 1. Inferred evolutionary relationship among the different *Ascaris* mtDNA genomes based on the 12 protein coding genes and the two ribosomal DNA genes (A) and based on 383 bp of the *cox*1 gene (B) using *Baylisascaris procyonis* as outgroup. For comparison, six *Ascaris* genomes were included from GenBank (accession numbers indicated). Maximum likelihood (ML), maximum parsimony (MP) and Bayesian inference (BI) was used and the bootstrap values and posterior probabilities are given at the nodes if >70 in the order: ML/MP/BI. The scale bar: number of base substitutions per site.

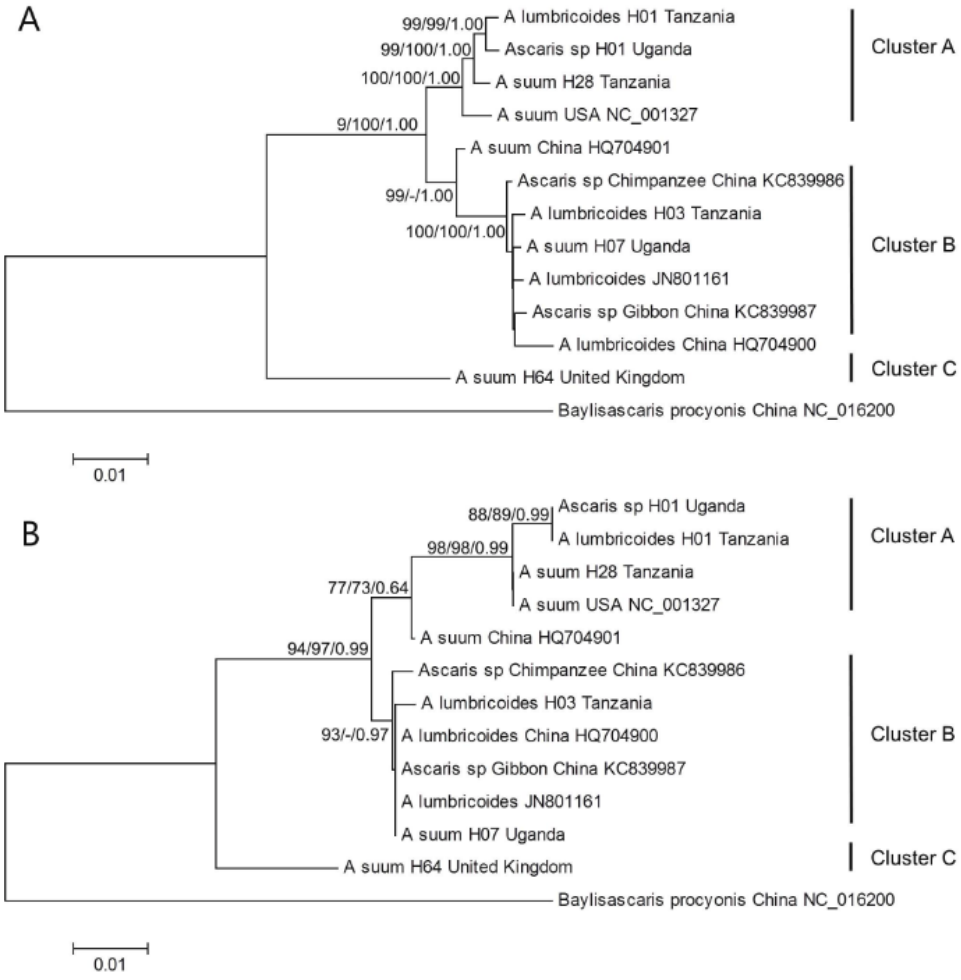
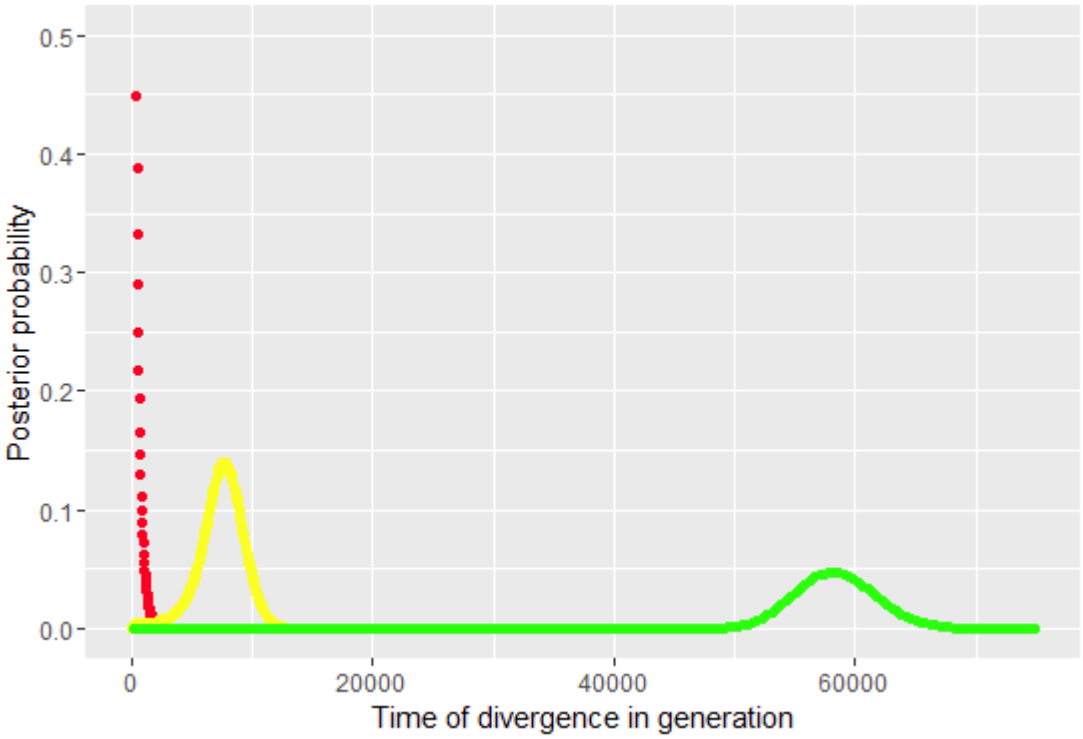


Fig. 2. Splitting time based on the isolation and migration model between *Ascaris* populations of clusters A/B and cluster C (yellow line) and Ascaris populations of cluster A and B (red points) and the time to the most recent common ancestor (TMRCA) (green line). The horizontal axis represents the number of generations since splitting while the vertical axis represents the posterior probability density.



453

Table 1. Primers used for long range PCR of the mtDNA genome of *Ascaris* from humans and pigs

Region	Fragment length	Forward primer	Reverse Primer
CO1F-ND5R	3.4kbp	CO1_F: TGGTTGTGTTGTTTGAGCTCA	ND5_R: ACAAAACTCAAACCAATACCAAC
ND5F-rrnSR	2.8kbp	ND5_F: AGGTGTAGAGGGGCTATGAA	rrnS_R: GGTACTAATCTGATTCATTCACC
rrnSF-ND2R	4.3kbp	rrnS_F: TGTTCCAGAATAATCGGCTAGAC	ND2_R: AAACCAACAAGACTTCCCAA
ND2F-CO3R	3.0kbp	ND2_F: TGTCTAAGGGGTCTGGTTCT	CO3_R: CCAAACTACATCTACAAAATGCC
CO3F-CO1R	3.2kbp	CO3_F: TGGTTTCTTTTGCTTGGGGT	CO1_R: ACCACAAAGTCACACCCGTA

454

455

456

Table 2. Number of SNPs and unique SNPs in the 12 protein coding genes for each of the 12 genomes using *Ascaris lumbricoides* from China (HQ704900) as reference genome

Genome	Total #SNPs	unique #SNPs
Reference genome	791	496
Ascaris suum H64 United Kingdom	576	410
A. suum USA (NC 001327)	240	28
A. suum H28 Tanzania	242	12
Ascaris sp. H01 Uganda	257	11
A. lumbricoides H01 Tanzania	257	12
A. suum China (HQ704901)	141	7
A. suum H07 Uganda	63	7
A. lumbricoides H03 Tanzania	66	6
A. lumbricoides (JN801161)	65	9
Ascaris sp. Chimpanzee China (KC839986)	67	2
Ascaris sp. gibbon China (KC839987)	64	9

457

458

459

Table 3. Number of SNPs, synonymous, non-synonymous mutations and the percentage of variation given as number of SNPs per gene length for each of the 12 protein coding genes for the 12 *Ascaris* genomes

Gene	SNPs per gene	Non-synonymous mutations (Ka)	Synonymous mutations (Ks)	Ka/Ks ratio	% of variation
*cox*1	91	30	61	0.49	5.76
*cox*2	47	11	36	0.31	6.72
*nad*3	24	7	17	0.41	7.14
*nad*5	117	33	84	0.39	7.37
*nad*6	39	10	29	0.34	8.96
*nad*4L	19	3	16	0.19	8.11
*nad*1	65	19	46	0.41	7.45
*atp*6	44	17	27	0.63	7.33
*nad*2	65	25	40	0.63	7.71
*cyt*b	99	33	66	0.50	9.01
*cox*3	57	15	42	0.36	7.34
*nad*4	124	36	88	0.41	10.08

460