Infection, Genetics and Evolution 48 (2017) 4–9

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Ascaris phylogeny based on multiple whole mtDNA genomes

Peter Nejsum [a,b,*,1], Mohamed B.F. Hawash [c,h,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
[d] 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
[h] Zoology Department, Faculty of Science, Cairo University, Giza, Egypt

ARTICLE INFO

Article history:
Received 25 October 2016
Received in revised form 29 November 2016
Accepted 2 December 2016
Available online 7 December 2016

Keywords:
Ascaris
Mitochondrial genomes
Human
Pig
Phylogeny
Soil transmitted helminth

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.

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 van 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

* Corresponding author at: Department of Clinical Medicine, Health, Aarhus University, Palle Juul-Jensens Boulevard 99, 8200, Aarhus N, Denmark.
E-mail address: pn@clin.au.dk (P. Nejsum).
[1] Joint first authors.

http://dx.doi.org/10.1016/j.meegid.2016.12.003
1567-1348/

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 s), annealing at 55 °C for 30 s, 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., 2013) 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 $\Theta = 2\,Ne\mu$, 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,

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