RESEARCH ARTICLE

Population Genetics of *Nosema apis* and *Nosema ceranae*: One Host (*Apis mellifera*) and Two Different Histories

Xulio Maside[1,2,3], Tamara Gómez-Moracho[1,2,4], Laura Jara[5], Raquel Martín-Hernández[4,6], Pilar De la Rúa[5], Mariano Higes[4], Carolina Bartolomé[1,2] *

1 Medicina Xenómica, CIMUS, Universidade de Santiago de Compostela, Santiago de Compostela, Galicia, Spain, 2 Xenómica Comparada de Parásitos Humanos, IDIS, Santiago de Compostela, Galicia, Spain, 3 Departamento de Anatomía Patolóxica e Ciencias Forenses, Universidade de Santiago de Compostela, Santiago de Compostela, Galicia, Spain, 4 Laboratorio de Patología Apícola. Centro de Investigación Apícola y Agroambiental (CIAPA), Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal (IRIAF), Consejería de Agricultura de la Junta de Comunidades de Castilla-La Mancha, Marchamalo, Guadalajara, Spain, 5 Departamento de Zoología y Antropología Física, Facultad de Veterinaria, Universidad de Murcia, Murcia, Spain, 6 Instituto de Recursos Humanos para la Ciencia y la Tecnología (INCRECYT-FEDER), Fundación Parque Científico y Tecnológico de Albacete, Albacete, Spain

* carolina.bartolome@usc.es

Abstract

Two microsporidians are known to infect honey bees: *Nosema apis* and *Nosema ceranae*. Whereas population genetics data for the latter have been released in the last few years, such information is still missing for *N. apis*. Here we analyze the patterns of nucleotide polymorphism at three single-copy loci (*PTP2*, *PTP3* and *RPB1*) in a collection of *Apis mellifera* isolates from all over the world, naturally infected either with *N. apis* ($N = 22$) or *N. ceranae* ($N = 23$), to provide new insights into the genetic diversity, demography and evolution of *N. apis*, as well as to compare them with evidence from *N. ceranae*. Neutral variation in *N. apis* and *N. ceranae* is of the order of 1%. This amount of diversity suggests that there is no substantial differentiation between the genetic content of the two nuclei present in these parasites, and evidence for genetic recombination provides a putative mechanism for the flow of genetic information between chromosomes. The analysis of the frequency spectrum of neutral variants reveals a significant surplus of low frequency variants, particularly in *N. ceranae*, and suggests that the populations of the two pathogens are not in mutation-drift equilibrium and that they have experienced a population expansion. Most of the variation in both species occurs within honey bee colonies (between 62%-90% of the total genetic variance), although in *N. apis* there is evidence for differentiation between parasites isolated from distinct *A. mellifera* lineages (20%-34% of the total variance), specifically between those collected from lineages A and C (or M). This scenario is consistent with a long-term host-parasite relationship and contrasts with the lack of differentiation observed among host-lineages in *N. ceranae* (< 4% of the variance), which suggests that the spread of this emergent pathogen throughout the *A. mellifera* worldwide population is a recent event.

Introduction

The genus *Nosema* (Fungi, Microsporidia, Dihaplophasea, Dissociodihaplophasida Nosematidae; Nägeli, 1857) contains over eighty species [1,2] typically found in arthropods. Two species, *N. apis* and *N. ceranae*, parasitize the Western honey bee, *Apis mellifera*. *N. apis* Zander, 1909 is a globally distributed pathogen that was identified in this host more than a hundred years ago [3], whereas *N. ceranae* was described at the end of the twentieth century [4]. This latter species, although initially discovered in the Asian honey bee *Apis cerana* [4], was recently proved to infect *A. mellifera* [5,6], and since then it has been found worldwide in this new host [7,8,9,10,11], as well as in several other *Apis* [12,13] and *Bombus* species [14,15]. Both pathogens are transmitted through the ingestion of spores during feeding, grooming and trophallaxis [16,17]. Once in the gut, they invade the ventricular cells causing disease, but the clinical and epidemiological characteristics of the parasitization by either species are different; the infection by *N. apis* (type A nosemosis) does not usually cause the death of the colonies and is characterized by dysentery, general weakening of the adults, locomotion impairment and crawling [18]. These symptoms are not present in *N. ceranae* infections (type C nosemosis) [19], which produce alterations in the temporal polyethism, foraging activity and life span of infected bees [20,21,22]. Although the same could also be true for *N. apis* [23], the higher prevalence of *N. ceranae* throughout the year in temperate climates [24,25]–in contrast with that of *N. apis*, which usually displays seasonal peaks [26]–, induces a chronic stress on the colony that may eventually lead to its collapse [20,27,28]. This effect, whose potential relationship with the large scale depopulation phenomenon is still matter of debate ([24,28], or [29,30] for a different point of view), is much more dramatic in Mediterranean countries, especially in Spain, where climatic conditions and/or beekeeping practices seem to increase the impact of *N. ceranae* on honey bee colonies reviewed in [31].

Genetic data revealed that *N. apis* and *N. ceranae* are highly divergent at the nucleotide level (average nonsynonymous divergence of 10%; [32]) and that there has been considerable gene shuffling since the split from their common ancestor [33], evidencing that they are not very close relatives within the genus *Nosema* [34,35].

The genetic characterization of *N. ceranae* populations in *A. mellifera* has been achieved in the last few years by analyzing different components of the ribosomal DNA (rDNA) [36,37], single-copy genes [32,38,39,40,41] and whole genomes [42]. The most relevant conclusions of these studies are that i) *N. ceranae* isolates obtained from individual honey bees exhibit multiple alleles at single copy loci [38,39], ii) most of the variation resides within honey bee colonies [39,42], iii) there is no differentiation among geographically distant isolates [36,38,39,42], iv) this pathogen has experienced a recent demographic expansion in *A. mellifera* [38,39,42], and v) there is evidence for low, but significant, levels of recombination [36,38,40,41,42].

In contrast, there is little information about the population genetics of *N. apis*. Although a few sequence data have been released in public databases, most of them remain unpublished (e.g. PopSets 723438493, 698364701, 225055863 from GenBank) and/or involve the analysis of rDNA [43,44,45] that harbors multiple copies in the *N. apis* genome [33]. These are organized as tandemly repeated units, each of them consisting of a small (SSU) and large (LSU) subunits separated by an internal transcribed spacer (ITS), and an intergenic spacer (IGS) [45,46] (see [33] for a slightly different organization). The redundancy of these arrays usually promotes the conservation of rDNA sequences through different mechanisms (concerted evolution [47], and/or birth-and-death processes [44]) that preserve their important role in mRNA translation. However, in the case of *N. apis* and *N. ceranae*, rDNA copies are highly diverse [33,36,48] and, although the reasons for the existence of differently expressed rRNA copies are still to be determined [37], the presence of such heterogeneous units complicates the assessment of the

levels of polymorphism [44,49], as within-genome heterogeneity is hard to disentangle from between-individuals diversity. This limits the use of ribosomal loci to estimate the genome-wide patterns of variability.

Here we report a population genetic analysis conducted to address questions that are central to our understanding of the recent evolutionary history of *N. apis*, such as: what are the levels of genetic variation of this parasite? Is there any genetic evidence for a long-term association between *N. apis* and *A. mellifera*? Is the *N. apis* population panmictic or is there any sign of geographical structure? With this aim, the sequences of three single copy genes were studied in a collection of *N. apis* and *N. ceranae* isolates obtained from *A. mellifera* colonies from all over the world. These loci had been previously studied in *N. ceranae* [32,39,40] and their patterns of polymorphism used to yield new insights into this parasite's populations. Our results, along with these previous data, provide the first comparative analysis of the patterns of genetic variation of both pathogens in the same host species.

Material and Methods

Samples

N. apis was isolated from twenty two naturally infected *A. mellifera* colonies from eleven countries worldwide: Algeria, Argentina, Canada, Chile, Germany, Hungary, the Netherlands, Poland, Slovenia, Spain, and Turkey (S1 Table). A similar number of *N. ceranae* isolates (N = 23) were collected from *A. mellifera* colonies from 17 countries: Algeria, Argentina, Australia, Brazil, Canada, Chile, Croatia, Greece, Hawaii (USA), Hungary, Japan, the Netherlands, Slovenia, Solomon Islands, Spain, Taiwan, and continental United States of America (S2 Table).

Ethics statement

No specific permits were required for the described studies, which did not involve endangered or protected species.

DNA extraction, PCR amplification, cloning and sequencing

DNA was extracted from homogenized pools of 15–20 honey bees from each colony. This was carried out as in [50], using the BioSprintTM 96 DNA Blood Kit (QIAgen, Izasa, Barcelona, Spain). The reagents used in this process were tested by PCR to check for the presence of potential contamination with *N. apis*, *N. ceranae* or honey bee DNA in each round of extractions. The identity of *Nosema* species was determined by specific PCR amplification of the *16S rDNA*, as in [51]. No co-infections were detected in these samples.

Specific primers were designed with Primer Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome) using sequences of each species as query (KE647054.1: 13328–14116—locus tag NAPIS_ORF00435—for *PTP2*, KE647278.1: 2294–4285—locus tag NAPIS_ORF01922—for *PTP3* and DQ996230.1 for *RPB1* in *N. apis*, and XM_002995356.1 for *RPB1* in *N. ceranae*, respectively). Primer pairs for amplification of *PTP2*, *PTP3* and *RPB1* in *N. apis* were: PTP2 Na-F (CTGCTACAGCACCGCCATTA) and PTP2 Na-R (TGGGGTTTAATCTTGCTTTTTCCA), PTP3 Na-F (AGACAAGGTGTTTCTCCAAAAGA) and PTP3 Na-R (GCAAGGTTTTCTTCTGTTGCC) and RPB1 Na-F (GTTAAGAGCAGAAGATGATCTAAC) and RPB1 Na-R (CTGATAATTTGTTTTCCTGTCCAATA), respectively. Primer pairs to amplify *RPB1* in *N. ceranae* were those published in [32].

PCR amplification, cloning and sequencing procedures were performed as in [39]. PCR annealing temperatures were adjusted for each of the primer pairs. These were 59.0°C, 58.0°C

and 56.5°C for *PTP2*, *PTP3* and *RPB1* in *N. apis* and 55°C for *RPB1* in *N. ceranae*, respectively. Each round of PCR amplification included negative and positive controls (PCR components with no template DNA, and PCR components + DNA extracted from *N. ceranae*–or *N. apis*–positive isolates, respectively).

Sequences were checked for accurate base calling using CodonCode Aligner (CodonCode Corporation, Dedham, MA, USA) and aligned with MUSCLE [52] with their reference sequences to determine the nucleotide positions at each locus. The alignments were manually corrected with BioEdit [53] and the sequences submitted to GenBank (S1 and S2 Tables).

Apis mellifera: lineage assignment

Determination of the *A. mellifera* evolutionary lineage was performed by sequence analysis of the intergenic region between the tRNA[leu] and the cytochrome oxidase II (cox2) gene as described previously [54]. DNA was extracted from a pair of legs using the Chelex® method [55]. The intergenic tRNA[leu]-cox2 region was PCR-amplified in a thermocycler PTC 100 (MJ Research) in a total volume of 12.5 μL with KapaTaq DNA Polymerase (KAPA BIOSYSTEMS), containing 2 μL of DNA template, 200 μM total dNTP, 1 X Reaction Buffer, 0.5 U/rxn Kapa-Taq DNA Polymerase, 1.5 mM $MgCl_2$, 0.4 μM of each primer (E2 and H2, [56]). The thermo-cycler program used was: 94°C (5 min); 35 cycles of a 45 s denaturation at 94°C, a 45 s elongation at 48°C, a 60 s extension at 62°C; and a final extension step at 65°C for 20 min. Amplicons were sequenced with the primer E2 (Secugen S.L., Madrid, Spain). Each sequence was manually checked for base calling and a multiple sequence alignment was performed with the MEGA program, version 6 [57]. Evolutionary lineages were determined by comparison with sequences deposited in GenBank (lineage C: JQ977699, JF723946; lineage M: HQ337441, KF274627; lineage A: EF033650, JQ746693).

Sequence analyses

Nucleotide diversity in *N. apis* and *N. ceranae* was estimated at synonymous and non-synonymous sites with DnaSP v5.10.02 [58]. Sites with alignment gaps were excluded from the analyses. π [59] and θ_W [60] were calculated applying the Jukes-Cantor correction [61]. π, the average number of pairwise differences between sequences, is sensitive to the frequency of polymorphisms and complementary to the estimate of θ_W, which measures the levels of variability by counting the number of segregating sites, independently of their frequency, and thus giving more weight to rare variants. The Tajima´s *D* test [62] compares the two statistics. If the population is in mutation-drift equilibrium, π and θ_W are expected to have same value, and *D* should be equal to zero. Negative *D* values reflect an excess of low frequency variants (greater θ_W), which under neutrality can be interpreted as evidence for a recent population expansion. According to Tajima's considerations on the different distributions followed by *D* at individual or pooled loci [62], the statistical significance of the deviation from neutral expectations for individual genes was determined using DnaSP (which assumes a beta distribution). When several unlinked regions of DNA were combined to describe the patterns of polymorphism of a species (pooled loci) this significance was calculated by applying Tajima´s formulae and assuming a normal distribution [62]. The possibility of a population expansion was further investigated by applying the Fu's F_S [63], as implemented in DnaSP v5.10.02. Its significance was assessed by comparing the observed values with a null distribution generated by 1,000 coalescent simulations. The number of net nucleotide substitutions per site between populations, *D*a [59], was also estimated with DnaSP.

The program MLHKA [64] allows to test for selection at individual loci in a multilocus framework by comparing the relative amounts of within and between-species synonymous

variation across unlinked loci. [65]. The patterns of diversity at the genes used in this study were compared with those observed at other loci with similar data available (*actin*, *Hsp70*, *HSWP4*, and *SWP30*) [32,38,39]. Genes that did not exhibit enough sequence identity between *N. apis* and *N. ceranae* to be confidently assigned as orthologs were discarded from the analysis (*NCER_100064*, *NCER_100070*, *NCER_100533*, *NCER_100768*, *NCER_101165* and *NCER_101600*; [38]), as was *PTP1* [39], which is tightly linked to *PTP2* [48,66]. Rates of synonymous and nonsynonymous divergence between *N. apis* and *N. ceranae* sequences for these loci were estimated using the Yang and Nielsen method [67], as implemented in the software PAML v 4.8a [68].

Lower bounds of the recombination rate were estimated using two different statistics under the assumption of the infinite sites model (i.e. each segregating site has mutated only once): *Rm*, the minimum number of recombination events, is based on the four-gamete test [69], which infers a recombination event if all four possible two-locus haplotypes occur in the sample, and *Rh* [70], which bounds the number of recombination events by calculating the difference between the number of haplotypes in the sample and the number of segregating sites. Both statistics were calculated with RecMin [70] (http://www.stats.ox.ac.uk/~myers/RecMin.html). The population scaled recombination rate (ρ) at the three loci was estimated applying the composite-likelihood method of Hudson [71], adapted to finite-sites model (to account for sites that might have experienced more than one mutation), as implemented in LDhat v2.0 [72]. Since the likelihood of observing recombination is dependent on the order of sites, the statistical significance of a non-zero rate of recombination was evaluated with a permutation test, in which the maximum composite likelihood was calculated under random permutation of the physical position of the variants (1000 permutations) [72]. *Nosema* parasites are commonly found as single-cell diplokaryons, so that they harbor a minimum of two haploid genomes. Thus, the number of haploid genomes is assumed to be 2 x $2N_e$, and $\rho = 4N_e r$.

Haplotype diversity was estimated with DnaSP v5.10.02, the statistics K_{ST}^* [73] and S_{nn} [74], were used to investigate the population structure. Their significance was assessed using permutation tests (1000 replicates).

An analysis of molecular variance (AMOVA) was performed with Arlequin 3.5 [75] and the significance of covariance components was checked by applying non-parametric permutation procedures (3000 permutations).

Haplotype networks were generated with Network 4.6.1.0 (Fluxus Technology, http://www.fluxus-engineering.com/sharenet.htm) using the Median-Joining algorithm, which allows for more than one different nucleotide per site. The epsilon parameter was set to 0, 10 and 20 in successive runs in order for the resulting network to include all possible shortest trees in the resulting network. Since no significant differences were observed, only those networks generated with epsilon = 0 are presented. The Connection Host criterion was used as distance calculation method. The Reduced Median algorithm (with *r* = 2) was applied to obtain a simplified network containing all shortest trees. All networks were redrawn manually.

Identification of mating and meiotic genes

The identification in *N. apis* and *N. ceranae* of the components of a sex-related locus [76] and an inventory of genes involved in meiosis [77] was performed by means of Blastp and SQR Sequence Search (https://www.ncbi.nlm.nih.gov/Structure/seqr) using as queries protein sequences from other microsporidia such as *Encephalitozoon cuniculi*, *Enterocytozoon bieneusi*, *Antonospora locustae* and *Nosema bombycis*.

Results

Genetic diversity

The genetic variability of *N. apis* samples was initially assessed at *PTP2*, *PTP3* and *RPB1* in seven naturally infected *A. mellifera* colonies from Algeria, Argentina, Canada, Slovenia, Spain (*N* = 2) and Turkey (dataset[A] in Table 1 and S1 Table). To increase the resolution of the analysis, and given that the levels of diversity were similar for the three genes (see below), the *RPB1* locus was randomly selected to enlarge the former dataset with sequences of 15 additional samples from Canada (*N* = 2), Chile (*N* = 2), Germany (*N* = 3), Hungary, Netherlands (*N* = 2), Poland (*N* = 3), Slovenia (*N* = 2), (dataset[A + B] in Table 1 and S1 Table).

The three genes displayed similar levels of synonymous variation in *N. apis* (pooled π_S values of 1.17%, 1.27% and 1.50%, for *PTP2*, *PTP3* and *RPB1*[A], respectively; these values were estimated by pooling all the sequences of each locus). The enlarged *RPB1* dataset (*RPB1*[A+B]) produced comparable results (π_S = 1.68%) and was used hereafter. It is interesting to note that these pooled values are twice the observed average diversity across the seven samples (0.80%, 0.85% and 0.82% for *PTP2*, *PTP3* and *RPB1*[A], respectively). This discrepancy could reflect some level of differentiation among isolates and was further investigated in the "Population structure" section below.

The patterns of variation at these loci were also studied in *N. ceranae* (Table 2 and S2 Table). The pooled π_S values for *PTP2*, *PTP3* and *RPB1* were 1.00%, 0.85% and 1.58%, respectively and, in contrast to what was observed in *N. apis*, these estimates were very similar to the average diversities across samples (0.95%, 0.82% and 1.42%, for the same loci, respectively).

In both species the levels of polymorphism at nonsynonymous positions were much lower than those observed at synonymous sites (Tables 1 and 2) and θ_W estimates were usually higher than those of π (resulting in pooled negative Tajima's *D* values, especially in *N. ceranae*).

To verify if the observed patterns diversity and divergence at these three genes departed significantly from those of other genomic loci with available population data (*actin*, *Hsp70*, *HSWP4*, and *SWP30*) we performed a maximum likelihood HKA test [64]. *Actin* and *Hsp70* displayed lower diversity relative to their divergence levels (likelihood ratio test, *LRT* = 18.5, *P* < 0.001). No significant deviations were detected at any the other five loci. In addition, the ratio of nonsynonymous to synonymous divergence (d_N/d_S) was below unity for the seven loci, ranging from 0.02–0.04 for *actin*, *Hsp70* and *RPB1* to 0.21 for *SPW30* and *HSPW4* (S3 Table).

Considering the evidence for a recent population expansion in *N. ceranae* [32,38,39,42], we examined this possibility in *N. apis* by applying two alternative tests: Tajima's *D* [62] and Fu's F_S [63]. The former test can be used to compare the frequency spectrum of variants with neutral expectations, and revealed an excess of low frequency synonymous variants in *N. apis*. Although this deviation was not significant at individual loci (Table 1), it was significant when data from *PTP2*, *PTP3* and *RPB1* were combined (D_S = -1.82, *P* = 0.034). Similar results were obtained in *N. ceranae*, where the combined data revealed a significant excess of low frequency synonymous variants (D_S = -2.93, *P* < 0.002). The F_S test [63] provided additional evidence for a significant excess of haplotypes in *N. apis PTP2* and *RPB1* genes, as compared with neutral expectations (Table 3). *N. ceranae* sequences obtained from *A. mellifera* displayed a similar pattern (Table 3).

Given the possibility that the *N. apis* population was subdivided in two demes (see the "Population structure" section below), one in lineage A honey bees (from Africa and the Iberian Peninsula) and another one in the European honey bee lineages C and M, the *D* and F_S statistics were estimated separately in both groups. Tajima's *D* for lineage C isolates was significant both at synonymous and nonsynonymous sites (-2.18, *P* < 0.015 and -3.30, *P* < 0.0001, respectively), but no significant skew was observed at synonymous sites among isolates collected

Table 1. Nucleotide diversity at three loci from *N. apis*: *PTP2*, *PTP3* and *RPB1*.

	Isolate	Origin	PTP2							PTP3							RPB1						
			N	π_S	θ_{WS}	D_S	π_A	θ_{WA}	D_A	N	π_S	θ_{WS}	D_S	π_A	θ_{WA}	D_A	N	π_S	θ_{WS}	D_S	π_A	θ_{WA}	D_A
A	52	SLO	9	0.31	0.51	-1.36	0.00	0.00	NA	5	0.00	0.00	NA	0.00	0.00	NA							
	174	TUR	10	0.28	0.49	-1.40	0.15	0.26	-1.67	5	1.05	1.25	-1.05	0.34	0.41	-1.09							
	204	ARG	8	0.00	0.00	NA	0.09	0.14	-1.31	5	1.42	1.68	-1.09	0.13	0.10	-1.22							
	381	CAN	9	0.31	0.51	-1.36	0.12	0.20	-1.51	10	0.53	0.93	-1.56	0.24	0.30	-0.76							
	399	SPA	7	1.60	1.69	-0.34	0.16	0.23	-1.36	7	1.02	1.42	-1.43	0.12	0.09	1.34							
	410	SPA	10	1.79	1.46	0.85	0.15	0.26	-1.67	13	1.22	0.84	1.37	0.13	0.20	-1.23							
	854	ALG	10	1.30	0.98	1.23	0.04	0.07	-1.11	5	0.70	0.83	-0.97	0.17	0.20	-0.97							
	Pooled [A]		63	1.17	1.91	-1.53	0.10	0.58	-2.51***	50	1.27	2.14	-1.20	0.20	0.57	-1.91*							
B	59	SLO															8	1.97	1.30	2.29*	0.29	0.32	-0.34
	264	POL															10	1.94	1.36	1.78	0.18	0.14	0.85
	266	POL															10	1.96	1.70	0.58	0.21	0.29	-1.19
	268	POL															10	1.93	1.89	1.03	0.17	0.14	0.70
	363	CHI															8	2.09	1.66	1.10	0.18	0.16	0.58
	380	CAN															8	1.80	1.48	0.93	0.19	0.21	-0.52
	382	CAN															10	1.83	1.19	-2.16*	0.23	0.24	-0.23
	411	CHI															10	0.86	1.36	-1.64	0.16	0.29	-1.80*
	529	SLO															9	1.66	1.77	-0.37	0.09	0.15	-1.51
	569	HUN															10	1.93	1.36	1.69	0.20	0.19	0.14
	1074	GER															8	2.07	1.48	1.77	0.17	0.26	-1.60
	1098	GER															10	1.32	1.36	-0.21	0.18	0.24	-1.04
	1099	GER															10	0.79	0.85	-0.33	0.15	0.19	-0.94
	1511	NED															10	2.03	1.53	1.31	0.10	0.10	0.02
	1735	NED															9	0.11	0.18	-1.09	0.06	0.10	-1.36
	Pooled [A+B]																208	1.68	2.78	-1.16	0.22	1.37	-2.53***

[A], Seven isolates used to estimate *N. apis* diversity;

[B], 15 additional isolates used for further analysis of *RPB1*;

ALG: Algeria, ARG: Argentina, CAN: Canada, CHI: Chile, GER: Germany, HUN: Hungary, NED: Netherlands, POL.: Poland, SLO: Slovenia, SPA: Spain, TUR: Turkey; *N*: number of cloned sequences; π_S and π_A, pairwise nucleotide diversity at synonymous and nonsynonymous sites expressed as percentage [59]; θ_{WS} and θ_{WA}, the average number of synonymous and nonsynonymous segregating sites expressed as percentage, respectively [60]; the average number of synonymous and nonsynonymous sites analyzed across loci were 155.6 and 582.4, respectively; D_S and D_A, Tajima´s *D* [62] at synonymous and nonsynonymous sites, respectively; NA: not available; statistical significance of Tajima´s *D*:

*, *P* <0.05;

***, *P* <0.001.

doi:10.1371/journal.pone.0145609.t001

Table 2. Nucleotide diversity at three loci from *N. ceranae*: *PTP2*, *PTP3* and *RPB1*.

Isolate	Origin	PTP2 [a]							PTP3 [a]							RPB1 [b]						
		N	π_S	θ_{WS}	D_S	π_A	θ_{WA}	D_A	N	π_S	θ_{WS}	D_S	π_A	θ_{WA}	D_A	N	π_S	θ_{WS}	D_S	π_A	θ_{WA}	D_A
3	AUS															8	1.68	1.66	0.24	0.00	0.00	NA
4	AUS															8	0.81	1.23	-1.67*	0.13	0.19	-1.53
57	SPA	9	1.16	1.02	0.54	0.12	0.14	-0.58	8	1.10	1.08	-0.92	0.07	0.06	0.33	10	1.08	1.13	-0.24	0.05	0.09	-1.40
169	BRA	7	0.66	0.56	0.69	0.15	0.16	-0.27	9	1.14	1.07	0.24	0.22	0.28	-0.91	9	1.46	1.00	1.90	0.00	0.00	NA
253	SPA	10	1.53	1.47	0.11	0.12	0.21	-1.56	12	0.76	0.77	-0.06	0.29	0.46	-1.52	10	1.90	1.61	0.69	0.05	0.09	-1.40
376	CAN	8	0.60	0.53	0.41	0.08	0.07	0.33	6	0.70	0.76	-0.45	0.08	0.07	0.85	7	1.59	1.48	0.26	0.00	0.00	NA
377	CAN	6	1.11	1.21	-0.47	0.13	0.17	-1.13	8	0.88	1.34	-1.64	0.16	0.18	-0.30	10	1.87	1.61	0.41	0.05	0.09	-1.40
440	HUN	8	0.70	0.53	1.10	0.08	0.07	0.33	9	0.39	0.64	-1.51	0.13	0.17	-0.94	9	1.26	1.17	0.25	0.00	0.00	NA
526	NED	8	0.97	0.80	0.84	0.16	0.15	0.24	10	0.99	0.82	0.75	0.24	0.27	-0.38	7	1.85	1.48	1.18	0.07	0.10	-1.24
531	SLO	13	0.93	0.67	1.20	0.13	0.12	0.10	5	0.94	1.11	-1.09	0.21	0.22	-0.17	7	1.72	1.30	1.55	0.04	0.05	-1.01
839	ALG	7	0.40	0.56	-1.24	0.11	0.16	-1.24	10	0.35	0.62	-1.56	0.09	0.16	-1.56	7	0.00	0.00	NA	0.04	0.05	-1.01
911	TWN	6	0.74	0.60	1.03	0.30	0.34	-0.68	10	0.53	0.62	-0.51	0.14	0.16	-0.76	10	1.23	1.13	0.33	0.00	0.00	NA
912	SPA	8	0.87	1.33	-1.60	0.14	0.22	-1.45	12	0.70	0.96	-1.02	0.21	0.30	-1.17	6	0.61	0.80	-1.30	0.13	0.17	-1.23
1175	CRO	9	0.66	0.51	0.98	0.12	0.14	-0.58	11	0.57	0.79	-1.03	0.13	0.16	-0.51	7	1.99	1.86	0.27	0.04	0.05	-1.01
1244	ARG	7	1.06	0.85	1.11	0.09	0.08	0.56	11	0.40	0.59	-1.11	0.12	0.16	-0.75	10	1.59	1.45	0.36	0.05	0.09	-1.40
1251	HI	10	0.76	0.49	1.74	0.15	0.14	0.22	10	1.08	1.07	-0.54	0.30	0.43	-1.36	8	1.50	1.05	1.89	0.10	0.15	-1.45
1299	GRE	10	1.18	0.98	0.75	0.33	0.41	-0.84	12	1.06	0.96	0.33	0.11	0.10	0.22	8	0.97	1.23	-1.04	0.10	0.15	-1.43
1319	HI	7	1.06	0.85	1.11	0.09	0.08	0.56	5	1.17	1.39	-1.12	0.18	0.22	-1.05	11	1.55	1.09	1.62	0.05	0.09	-1.43
1324	HI	8	1.45	1.07	1.50	0.10	0.07	1.17	12	0.91	0.77	0.59	0.13	0.15	-0.38	6	1.51	1.20	1.39	0.13	0.17	-1.23
1610	USA	5	0.84	0.66	1.46	0.12	0.09	1.22	11	0.81	0.79	0.04	0.11	0.10	0.20	10	1.76	1.77	-0.10	0.05	0.09	-1.40
1994	CHI	11	1.09	1.41	-0.94	0.25	0.46	-1.90*	12	0.86	0.77	0.39	0.10	0.15	-1.38	10	1.69	1.29	1.28	0.05	0.09	-1.40
2032	SOL	12	1.20	0.92	1.03	0.14	0.13	0.15	11	1.11	1.78	-1.62	0.22	0.31	-1.22	10	1.39	1.12	0.90	0.10	0.18	-1.67
KI	JAP															8	1.60	1.23	1.35	0.00	0.00	NA
Pooled		169	1.00	2.19	-1.46	0.15	0.88	-2.34**	194	0.82	3.50	-2.21**	0.17	1.35	-2.49***	196	1.58	3.19	-1.54	0.05	0.78	-2.71***

[a] , Sequence data from [39];

[b] , sequence data for isolates 440, 1251 and 1324 from [32];

ALG: Algeria, ARG: Argentina, AUS: Australia, BRA: Brazil, CAN: Canada, CHI: Chile, CRO: Croatia, GRE: Greece, HI: Hawaii (USA), HUN: Hungary, JAP: Japan, NED: Netherlands, SLO: Slovenia, SOL: Solomon Islands, SPA: Spain, TWN: Taiwan, USA: United States of America; N: number of cloned sequences; π_S and π_A, pairwise nucleotide diversity at synonymous and nonsynonymous sites expressed as percentage, respectively [59]; θ_{WS} and θ_{WA}, nucleotide site variability based on the number of synonymous and nonsynonymous segregating sites expressed as percentage, respectively [60]; the average number of synonymous and nonsynonymous sites analyzed across loci were 178.3 and 650.7, respectively; D_S and D_A, Tajima´s D [62] at synonymous and nonsynonymous sites, respectively; NA: not available; statistical significance of Tajima´s D:

*, $P < 0.05$;

**, $P < 0.01$;

***, $P < 0.001$.

doi:10.1371/journal.pone.0145609.t002

Table 3. *F*S test for detecting population expansion.

Dataset	Fs					
	PTP2		**PTP3**		**RPB1**	
N. apis	-18.66	***	-8.17	ns	-87.22	***
Lineage A	-1.52	ns	0.46	ns	-3.54	*
Lineage C	-13.24	***	-5.70	**	-34.15	***
Lineage M					-0.77	ns
N. ceranae [a, b]	-34.35	***	-120.89	***	-77.55	***

Fs: Fu's *Fs* [63]; The significance was evaluated by comparing the values of the statistic, and the observed levels of recombination per gene, with a null distribution generated by 1,000 coalescent simulations (ns, non-significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$);

[a], sequence data for *PTP2* and *PTP3* from [39];

[b], sequence data for isolates 440, 1251 and 1324 (*RPB1*) from [32];

the remaining ones are from this work. Lineages A, C and M indicate the evolutionary lineage of the honey bee colonies where the isolates come from (isolate 410 was excluded from this analysis).

doi:10.1371/journal.pone.0145609.t003

from honey bees of lineages A or M (D_S = -0.19 and 1.83, ns; respectively.). The F_S test produced a similar scenario (Table 3).

Recombination

The estimate of the population scaled recombination rate (ρ) was significantly greater that zero at *PTP2* and *RPB1* in *N. apis* (Table 4) and several recombination events were detected at these two loci (*Rm* = 2 and 6, *Rh* = 2 and 14, respectively).

The outcome of these tests did not change after removing singleton mutations, which usually contain little information and could cause interferences in the assessment of recombination. In addition, the four-allele combinations expected after a recombination event were found both in individual samples (e.g. AT, TT, AC and TC at positions 276 and 453 of *PTP2* in

Table 4. Statistics used to detect recombination in *N. apis*.

Dataset	Locus	N	Rm	Rh	ρ	P
N. apis	*PTP2*	63	2	2	36	*
	PTP3	50	0	0	2	ns
	RPB1	208	6	14	11	***
N. ceranae	*PTP2* [a]	169	5	10	68	ns
	PTP3 [a]	194	3	6	61	*
	RPB1 [b]	196	6	12	14	*

N, number of cloned sequences; *Rm*, minimum number of recombination events [69]; *Rh*: lower bound on the number of recombination events [70] (http://www.stats.ox.ac.uk/~myers/RecMin.html); ρ, estimate of the population scaled recombination rate; *P*, probability of Lkmax $\leq$ estimated in a likelihood permutation-based test as implemented in LDhat [72]; ns, non-significant,

*, $P < 0.05$ and ***, $P < 0.001$;

[a], sequence data from [39];

[b], sequence data for isolates 440, 1251 and 1324 from [32];

the remaining ones are from this work.

doi:10.1371/journal.pone.0145609.t004

sample 854 –haplotypes 1, 15, 23 and 24, respectively–; S4 Table) and in samples from different populations (e.g. CC, CT, TC and TT at positions 771 and 963 of *RPB1*[A+B]; S7 Table), further supporting the existence of low, although statistically significant, levels of recombination in *N. apis*. Likewise, evidence for recombination was found in *N. ceranae*, although in this case it only reached statistical significance for *PTP3* and *RPB1* (Table 4).

Sex and meiotic loci

A Blastp analysis of the genomes of *N. apis* and *N. ceranae* revealed the presence of several components of a sex-related locus [78] that encode a *triose phosphate transporter* (*TPT*), a *high-mobility group* (*HMG*), and an *RNA helicase* (with accession numbers EQB61312, EQB61310, EQB60627 and KKO76186, KKO76188, KKO75161, respectively). In both species *TPT* and *HMG* were syntenic and harbored an additional predicted ORF between them, whereas the *RNA helicase* was unlinked to the former.

These genomes also contained meiotic genes, although not all of them presented orthologs in both species (S8 Table). However, it must be noted that these results should be taken with caution due to the difficulty of distinguishing orthologs encoding different members of gene families (i.e. *Smc1* and *Smc4* in *N. ceranae*; S8 Table).

Population structure

N. apis exhibited high levels of haplotype diversity at the three loci under study (*Hd* = 0.79–0.91). To explore the distribution of the haplotypes among samples these were plotted into networks (Fig 1), revealing the following patterns: (i) all loci presented a reduced number of common haplotypes (e.g. h1 in *PTP2* and in *PTP3*, and h2 and h5 in *RPB1*); (ii) most other haplotypes differed from these by a reduced number of substitutions (usually one or two); and (iii), there was a hinted association among haplotypes obtained from Spanish and Algerian samples. For example, *PTP2* haplotypes h15, h16, h17 were shared by the three samples from these two countries (Fig 1A), and most other closely linked haplotypes (e.g. h14, h18, h19, h21, h22 and h24) were found in either of them. A similar effect was observed for *PTP3* haplotypes h3 and h13 (Fig 1B), and h2 of *RPB1* (Fig 1C). This raised the possibility of a structuring of the parasite and host populations, which was not apparent in *N. ceranae* networks (S1 Fig), and it was further explored by determining the evolutionary lineage of all sampled honey bee colonies and by quantifying the relative contribution of three covariance components to the observed parasite haplotype diversity: (i) covariance within isolates (i.e., within honey bee colonies), (ii) among isolates obtained from honey bees of the same evolutionary lineage, and (iii) among host evolutionary lineages. Most *N. apis* isolates were collected from *A. mellifera* colonies of lineage C, except isolates 399 and 854 (from Spain and Algeria) which were of lineage A, and isolates 382 (Canada) and 411 (Chile) which belonged to lineage M (S1 Table). A third isolate from Spain, 410, displayed mixed results, with evidence for the presence of honey bees of both lineages A and M, probably due to drifting workers and the existence of colonies of both evolutionary lineages in the same apiary (S1 Table).

The analysis of molecular variance revealed a structured *N. apis* population, where between 20 and 34% of the total variance at the three loci corresponded to differences among host-lineages (*P* < 0.05, in permutation tests; Table 5). These results held irrespective of whether sample 410 was considered as belonging to lineage A or M. Differences among haplotypes of the same isolate accounted for the best part of the variance (between 62 and 70%), and differences among isolates of the same lineage represented ≤11% of the variance. In *N. ceranae* the differentiation of haplotypes within isolates was even greater (between 91 and
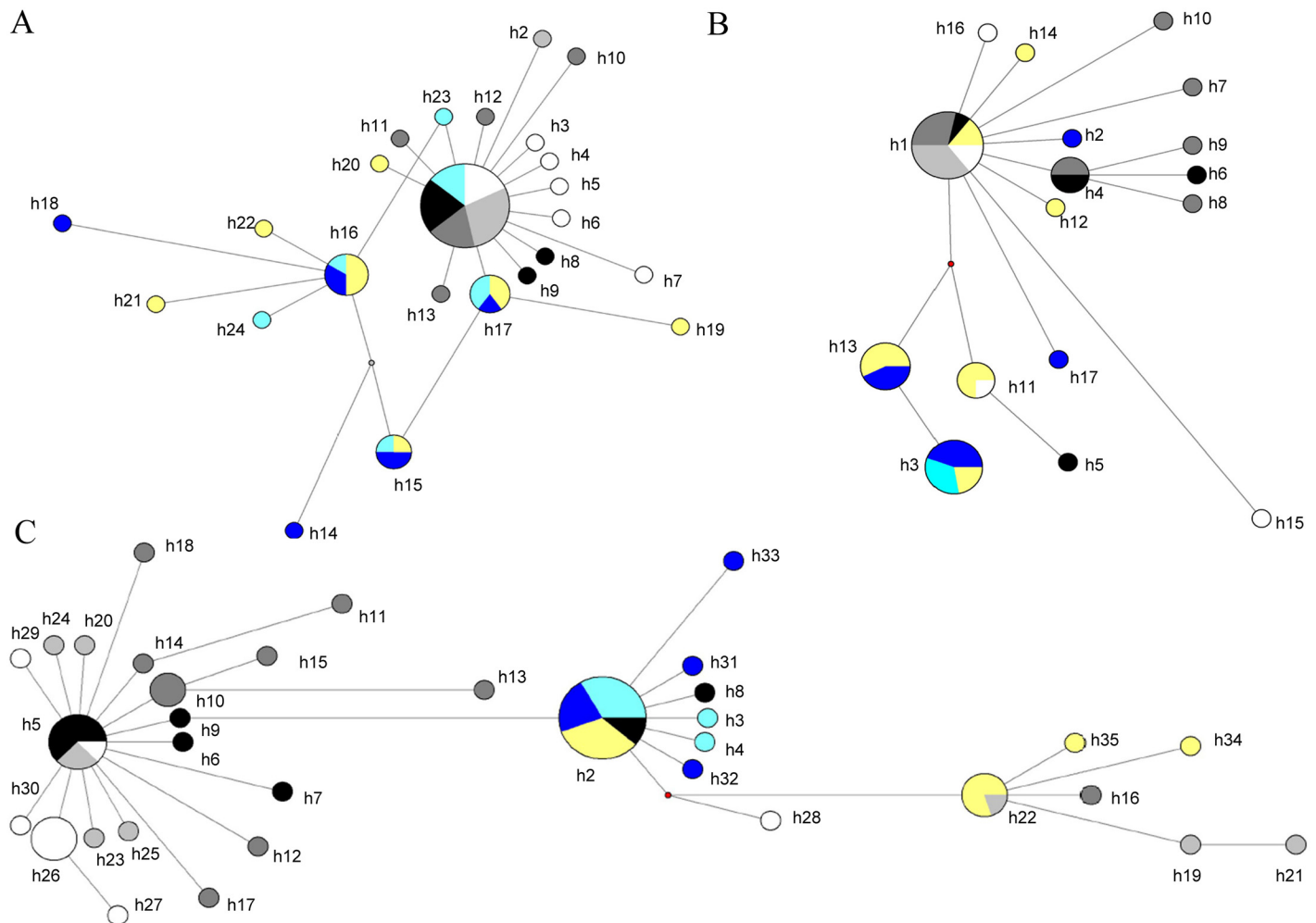
Fig 1. Median-joining haplotype network for three *N. apis* loci according to their geographical origin and *A. mellifera* lineage: *PTP2* (A), *PTP3* (B) and *RPB1* (C). Haplotypes are depicted by circles, the width being proportional to their frequencies. Color codes are as follows; blue: lineage A (light blue: isolate 854 (Algeria); dark blue: isolate 399 (Spain)); yellow: lineages A/M (isolate 410 (Spain)); greyscale: lineage C (black: isolate 204 (Argentina); dark grey: isolate 381 (Canada); light grey: isolate 52 (Slovenia); white: isolate 174 (Turkey)); red dots represent median vectors (hypothesized haplotypes required to connect existing sequences within the network with maximum parsimony).

doi:10.1371/journal.pone.0145609.g001

99%) but, contrastingly, the differences among host-lineages were not significant as they represented just a tiny fraction of the observed variance ($< 4\%$, Table 5).

A pairwise analysis of *N. apis* differentiation between host lineages uncovered that most variation occurred between lineage A and the other two lineages, which were otherwise indistinguishable (Table 6). K_{ST}^*, which measures the average pairwise differences within populations with respect to the total, revealed significant differentiation between *N. apis* sequences obtained from *A. mellifera* colonies of lineages A and C (K_{ST}^* between 0.09 and 0.25, $P < 0.001$, Table 6) or between A and M ($K_{ST}^* = 0.30$, $P < 0.001$). Similarly, S_{nn}, which estimates how often related sequences are found in the same population, reached significant values between group A and C ($S_{nn} = 0.82–0.93$, $P < 0.001$) and between A and M ($S_{nn} = 1.00$, $P < 0.001$). Both tests failed to detect significant differences between sequences obtained from lineages C and M.

The divergence between two populations is a direct function of the mutation rate times twice the number of generations since their split. Considering that: (i) the net divergence (*D*a)

Table 5. Analysis of molecular variance (AMOVA) in *N. apis* haplotypes according to *A. mellifera* lineages.

Dataset	Locus	Source of variation	d.f.	SS	VC	% var	P
N. apis	*PTP2*	Among lineages	1	17.2	0.51	32.6	*
		Among isolates within lineages	5	6.1	0.02	1.3	ns
		Within isolates	56	58.7	1.05	66.1	***
	PTP3	Among lineages	1	13.9	0.50	33.8	*
		Among isolates within lineages	5	6.8	0.07	4.6	*
		Within isolates	43	38.9	0.90	61.6	***
	RPB1	Among lineages	2	54.3	0.56	19.6	**
		Among isolates within lineages	19	94.1	0.32	10.8	***
		Within isolates	186	374.3	2.01	69.6	***
N. ceranae	*PTP2* [a]	Among lineages	2	6.0	0.04	3.5	*
		Among isolates within lineages	17	20.5	0.01	0.9	ns
		Within isolates	149	166.3	1.12	95.6	ns
	PTP3 [a]	Among lineages	2	242.1	-13.96	-2.2	ns
		Among isolates within lineages	17	14014.6	20.29	3.2	ns
		Within isolates	174	109558.1	629.64	99.0	ns
	RPB1 [b]	Among lineages	2	10.0	0.04	1.9	ns
		Among isolates within lineages	20	63.9	0.14	6.7	*
		Within isolates	173	339.7	1.96	91.4	**

[a], Sequence data from [39];

[b], sequence data for isolates 440, 1251 and 1324 from [32];

the remaining ones are from this work; isolate 410 was considered as if sampled from a lineage A honey bee colony (see text); d.f., degrees of freedom; SS, sum of squares; VC, variance components; % var, percentage of variation; p, probability of a random variance component value $\leq$ observed value, in 3024 permutations;

*, $P < 0.05$; **, $P < 0.01$ and ***, $P < 0.001$; ns, non-significant.

doi:10.1371/journal.pone.0145609.t005

between *N. apis* from lineage A and C/M colonies is 0.002. (ii) The number of spores can double as fast as every 24 hours (e.g. [79,80]). (iii) The substitution rate in these pathogens is about two times faster than that observed in other fungi (as estimated for *Encephalitozoon cuniculi* [81]), and that (iv) the per site mutation rates in fungi is of the order of 7.2×10^{-11} for *Neurospora crassa* or 2.2×10^{-10} for *Saccharomyces cerevisiae* [82], the split between the two parasite populations can be dated between 6,200 and 19,000 years ago.

Discussion

Here we report a population genetic analysis of *N. apis* based on the study of the patterns of diversity of three unlinked single copy genes: *PTP2* and *PTP3*, which encode polar tube proteins (reviewed in [83]), and the largest subunit of the RNA polymerase II (*RPB1*), a housekeeping gene that has been frequently used as phylogenetic marker in microsporidian species [84,85]. The levels of synonymous variation at these unlinked coding genes should be a good proxy for the extant neutral variation of the species [86]. The fact that they are single copy markers makes them a preferred choice than the commonly used ribosomal loci, as substantial levels of genetic variation and recombination between paralogous rDNA copies have been previously reported in microsporidia [36]. So far only Ironside [44] has published diversity data for a single copy locus (*RPB1*) in *N. apis*.

Table 6. Analysis of population differentiation in *N. apis* according to *A. mellifera* lineages.

Dataset	Locus	Host Lineages	K_{ST}*	P	S_{nn}	P
N. apis	PTP2	A & C	0.16	***	0.82	***
	PTP3	A & C	0.25	***	0.95	***
	RPB1	A & C	0.09	***	0.93	***
		A & M	0.30	***	1.00	***
		C & M	0.00	ns	0.81	ns
N. ceranae	PTP2 [a]	A & C	0,02	**	0,69	*
		A & M	0,02	*	0,55	ns
		C & M	0,00	ns	0,69	ns
	PTP3 [a]	A & C	0,00	ns	0,67	*
		A & M	0,01	*	0,61	**
		C & M	0,00	ns	0,73	**
	RPB1 [b]	A & C	0.02	*	0.68	ns
		A & M	0.00	ns	0.54	ns
		C & M	0.00	ns	0.76	ns

[a], sequence data from [39];
[b], sequence data for isolates 440, 1251 and 1324 from [32];
the remaining ones are from this work; isolate 410 was excluded; K_{ST}* [73], estimates the amount of within-deme nucleotide diversity relative to the overall diversity; S_{nn} [74], measures how often related sequences are found in the same deme; P, significance in permutation tests: ns, non-significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

doi:10.1371/journal.pone.0145609.t006

The three loci analyzed displayed similar average levels of synonymous diversity (π_S), about 1% (Table 1), analogous to what was found for these and other loci in *N. ceranae* [32,38,39,41], and somewhat higher than those estimated for *RPB1* (0.41%) in cloned sequences from a single isolate of *N. apis* [44]. In terms of diversity at all sites *N. ceranae* and *N. apis* displayed values of the order of 0.40%, which are lower than those described in *N. bombycis* (1.83%, [44]), but higher than those of other microsporidia of the genus *Hamiltosporidium* (between 0.06% and 0.28%, [87]).

Although it has been postulated that the polar tube proteins could be factors of virulence and thus subject to adaptive selection [88], the MLHKA test revealed that the relative levels of diversity and divergence of the *PTP* loci do not differ from those observed at three other unlinked loci (including a housekeeping gene, *RPB1*), which suggests that they evolve under the effect of similar evolutionary forces. Consistently low d_N/d_S values can be reconciled with a predominant effect of purifying selection over the seven loci. Although the large synonymous divergence between these species means that these results should be taken with caution, the fact that it applies to all loci supports that the genes used in the current study are a good proxy of the patterns of variation across the parasites' genome.

The detection of substantially lower variation (π_A) coupled with significantly negative Tajima's D_A values at nonsynonymous sites at the three loci in both species indicate that amino acid replacement variants are readily removed from the populations, which reflects that these loci are likely to be functional and subject to purifying selection, as previously suggested in *N. apis* [44] and *N. ceranae* [32,36,38,39,44] for these and other genes. This fits well with the finding of just 29 putative pseudogenes in the genome of *N. ceranae* [42], which indicates that the majority of coding sequences retained in these reduced genomes [33,48] are essential for the survival of these parasites. The relatively lower variability at a*ctin* and *Hsp70* can probably be

attributed to the effect of negative selection at linked deleterious sites (background selection) at these loci [89], which is consistent with the strong purifying selection—low d_N/d_S values—observed in these highly conserved genes [90,91] and the low recombination rates reported for these parasites [40] (see below).

The results of the Tajima´s D test at silent sites revealed an overall excess of low frequency variants in the two parasite species (Tables 1 and 2). Although some of these could correspond to nucleotide misincorporations introduced during the PCR process (despite the use of a high-fidelity enzyme blend), previous assays using either invariant DNA templates [38,39] or single DNA molecules [40], confirmed that the vast majority of the mutations detected in *N. ceranae* were actually present in the sample mixture, and that no error-prone bias was brought throughout the procedure. The same pattern was also observed at the genomic level [42], so there are no reasons to think that this would be different in *N. apis*.

All isolates presented substantial levels of nucleotide diversity (Tables 1 and 2). In fact, many of them harbored various distinct haplotypes, sometimes as many as nine (e.g. S7 Table). Given that the three genes are present as a single copy in the genome of both parasites, there are two possible and non-mutually exclusive explanations for the observed within-isolate variation: one is to assume that the two nuclei present in each cell are diploid, as it has been recently proposed for *N. ceranae* [42], so that they can harbor up to four different haplotypes for each loci, and the other is the existence of genetic heterogeneity among parasites in each host colony (mixed infections) [32,39,41].

At any rate, the accumulation of alleles at low frequencies observed in these two species is compatible with a demographic growth, in which most mutations present in the expanding populations have a recent origin and, therefore, are rare [92]. The greater D_S value obtained for *N. ceranae* in the combined sample (D_S = -2.93, $P < 0.002$), suggests that the expansion of this population might have taken place more recently or it has been more accentuated than that experienced by *N. apis* (D_S = -1.82, $P < 0.034$). This would agree with its recent jump to *A. mellifera* and spread throughout the worldwide distribution range of its new host [32,38,39,41,42].

A. mellifera originated between 6 and 8 million years ago somewhere in Asia, where all other species of the genus are confined, and from where it expanded to its historical geographic distribution range across sub-Saharan Africa, Europe and Western Asia [93,94]. The species now comprises several locally adapted and anatomically distinct subspecies, which split between 0.3 and one million years ago and can be clustered into four major groups: lineage A, includes subspecies that can be found in Africa and the Iberian Peninsula; lineage M is distributed along Western and Northern Europe; lineage C, in South Eastern Europe, and lineage O, in the Middle East and Western Asia [93,94]. Our results suggest that between 20% and 34% of the genetic variance of the *N. apis* population corresponds to differences between samples collected from honey bee colonies of different lineages (Table 5). It should be noted that the sampling scheme might influence the observed frequency spectra of variants, as the retrieval of alleles from distant locations of a structured population is likely to cause departures from neutral expectations assuming panmixia. The reduced between-sample variation in *N. ceranae* means that this effect is unlikely in this species, but the evidence for genetic differentiation between *N. apis* collected from different host lineages (lineage A *vs.* C or M) suggests that the assumed panmixia might not be met. The separate analysis of the parasite subpopulations revealed that only those isolated from European honey bees of lineage C depart from mutation-drift equilibrium.

Remarkably, the observed structure of the parasite population does not match that of the host: lineages C and M display the highest differentiation amongst the four *A. mellifera* lineages [93,94]. In contrast, *N. apis* isolates from these lineages are genetically indistinguishable. This lack of a full correspondence between the structures of parasite and host populations suggests

that they only share a fraction of their demographic history. Indeed, the split of the *N. apis* populations retrieved from honey bee lineages A and C (or M) was dated between 6,200 and 19,000 years ago, that is just after the Last Glacial Maximum, about 20,000 years ago [93]. Thus, a reduction of the parasite's geographic distribution range during the last glacial period might have prompted the isolation of the two populations. Contrastingly, the absence of genetic differentiation between the parasites from lineages C and M suggests that this population has spread across Europe recently (which is also consistent with the results of the Tajima´s *D* and *Fs* tests), and much later than the honey bees did. This expansion might have been associated with the practice of beekeeping by humans, whose origin has been traced to the Middle East and Egypt about five thousand years ago [95], and also with the human-driven colonization by *A. mellifera* of the New World, East Asia and Oceania in the last few centuries [94,96].

Evidence for low levels of recombination from nucleotide variation data had been previously reported [32,36,38,39,41,42], and further supported by Single Genome Amplification (SGA) analysis, in *N. ceranae* [40]. This new evidence for a second *Nosema* species, suggests that recombination might be a common feature of the genus. Genetic exchange between chromosomes is crucial in the evolution of organisms, and its detection has important consequences since it can generate new genetic combinations that result in individuals better adapted to confront novel environments or hosts.

If exclusively clonal reproduction is assumed, high levels of genetic differentiation would be expected between homologous sequences in the two nuclei of each individual (an adaptation of the Meselson effect for a diplokaryon), as observed in the asexual microsporidian *Hamiltosporidium tvarminnensis* [87]. But the absence of genetic structuration of the haplotypes retrieved in each colony, along with the observed neutral diversity within samples and the evidence for genetic recombination, suggest that there might exist mechanisms for occasional genetic flow between the nuclei. Whether this exchange takes place between nuclei of the same cell or between those of different cells during the multinucleated stages of the merogonia [97] remains to be determined.

Genetic exchange can occur during meiosis in sexual stages of the life cycle but also during parasexual processes such as mitotic crossover, non-homologous recombination and gene conversion. Although both mechanisms have been proposed in microsporidians (e.g.*Encephalitozoon cuniculi* [98], *Hamiltosporidium magnivora* [87], *Nematocida* [99] or *Nosema* spp. [36,44,100]), their occurrence and frequency are still a matter of debate as, in absence of cytological observation, the outcomes of these processes are difficult to distinguish [101,102]. The finding of a putative sex-related locus in *N. ceranae* and in *N. apis* goes in line with similar evidence in other microsporidians [76,87,99,103], but it should not be taken as a definite proof of sexual reproduction, which would require the presence of idiomorphs of this locus in different isolates and their expression during the mating phase [76]. The same could be said regarding the existence of core meiotic genes in these genomes [76,99,103,104] whose detection, although suggestive that meiosis may occur, does not ensure their functionality during this process [103].

The genetic diversity patterns at the three loci analyzed in this study suggest that *N. apis* and *N. ceranae* have experienced different recent evolutionary histories and provide new data on the relationship between these parasites and the honey bees that host them. In addition, they extend the evidence for genetic recombination to a second species of the genus, further supporting the idea that mechanisms of genetic exchange between chromosomes play an important role in modeling the genetic configuration of these organisms. However, further studies are needed in order to determine the extent to which the observed patterns extend to other parts of these species' genomes, to elucidate the molecular mechanisms responsible for the observed recombination and whether or not it implies sexual reproduction.

Supporting Information

S1 Fig. Median-joining haplotype network for three *N. ceranae* loci according to their *A. mellifera* lineage: *PTP2* (A), *PTP3* (B) and *RPB1* (C). Haplotypes are depicted by circles, the width being proportional to their frequencies (only shared haplotypes are named). Color codes are as follows; blue: lineage A (isolates 839 (Algeria), 57 and 253 (Spain), 169 (Brazil)); yellow: lineage M (isolates 912 (Spain), 526 (Netherlands), 1251 (Hawaii)); grey: lineage C (isolates 1244 (Argentina), 3 and 4 (Australia), 376 and 377 (Canada), 440 (Hungary), 531 (Slovenia), 911 (Taiwan), 1175 (Croatia), 1299 (Greece), 1319 and 1324 (Hawaii), 1610 (USA), 2032 (Solomon Islands), 1994 (Chile), KI (Japan)); red dots represent median vectors (hypothesized haplotypes required to connect existing sequences within the network with maximum parsimony).
(TIF)

S1 Table. Origin and accession numbers of *N. apis* sequences obtained from *A. mellifera* honey bees.
(XLS)

S2 Table. Origin and accession numbers of *N. ceranae* sequences obtained from *A. mellifera* honey bees.
(XLS)

S3 Table. Ratio of nonsynonymous to synonymous divergence (d_N/d_S) between *N. apis* and *N. ceranae* (Yang and Nielsen method).
(XLS)

S4 Table. Number of occurrences and nucleotide variants of *PTP2*[A] haplotypes from *N. apis*.
(XLS)

S5 Table. Number of occurrences and nucleotide variants of *PTP3*[A] haplotypes from *N. apis*.
(XLS)

S6 Table. Number of occurrences and nucleotide variants of *RPB1*[A] haplotypes from *N. apis*.
(XLS)

S7 Table. Number of occurrences and nucleotide variants of *RPB1*[A+B] haplotypes from *N. apis*.
(XLS)

S8 Table. Meiotic genes in different microsporidian species.
(XLS)

Acknowledgments

Author Contributions

Conceived and designed the experiments: CB XM MH RM-H. Performed the experiments: TG-M LJ. Analyzed the data: CB XM TG-M LJ PDR. Contributed reagents/materials/analysis tools: MH RM-H CB XM PDR. Wrote the paper: CB XM MH RM-H PDR.

References

1. Keeling P. Five questions about microsporidia. PLoS Pathog. 2009; 5: e1000489. doi: 10.1371/journal.ppat.1000489 PMID: 19779558

2. James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N, et al. Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. Curr Biol. 2013; 23: 1548–1553. doi: 10.1016/j.cub.2013.06.057 PMID: 23932404

3. Kudo R. Notes on Nosema apis Zander. J Parasitol. 1920; 7: 85–90.

4. Fries I, Feng F, da Silva A, Slemenda SB, Pieniazek NJ. Nosema ceranae sp. (Microspora, Nosematidae), morphological and molecular characterization of a microsporidian parasite of the Asian honey bee Apis cerana (Hymenoptera, Apidae). Eur J Protistol. 1996; 32: 356–365.

5. Higes M, Martín-Hernández R, Meana A. *Nosema ceranae*, a new microsporidian parasite in honeybees in Europe. J Invertebr Pathol. 2006; 92: 93–95. PMID: 16574143

6. Huang WF, Jiang JH, Chen YW, Wang CH. A *Nosema ceranae* isolate from the honeybee *Apis mellifera*. Apidologie. 2007; 38: 30–37.

7. Klee J, Besana AM, Genersch E, Gisder S, Nanetti A, Tam DQ, et al. Widespread dispersal of the microsporidian *Nosema ceranae*, an emergent pathogen of the western honey bee, *Apis mellifera*. J Invertebr Pathol. 2007; 96: 1–10. PMID: 17428493

8. Chen Y, Evans JD, Smith IB, Pettis JS. Nosema ceranae is a long-present and wide-spread microsporidian infection of the European honey bee (Apis mellifera) in the United States. J Invertebr Pathol. 2008; 97: 186–188. PMID: 17880997

9. Medici SK, Sarlo EG, Porrini MP, Braunstein M, Eguaras MJ. Genetic variation and widespread dispersal of *Nosema ceranae* in *Apis mellifera* apiaries from Argentina. Parasitol Res. 2012; 110: 859–864. doi: 10.1007/s00436-011-2566-2 PMID: 21808980

10. Giersch T, Berg T, Galea F, Hornitzky M. Nosema ceranae infects honey bees (Apis mellifera) and contaminates honey in Australia. Apidologie. 2009; 40: 117–123.

11. Higes M, Martín-Hernández R, Garrido-Bailón E, Botías C, Meana A. The presence of Nosema ceranae (Microsporidia) in North African honey bees (Apis mellifera intermissa). J Apicult Res. 2009; 48: 217–219.

12. Botías C, Anderson DL, Meana A, Garrido-Bailón E, Martín-Hernández R, Higes M. Further evidence of an oriental origin for Nosema ceranae (Microsporidia: Nosematidae). J Invertebr Pathol. 2012; 110: 108–113. doi: 10.1016/j.jip.2012.02.014 PMID: 22425522

13. Chaimanee V, Chen Y, Pettis JS, Scott Cornman R, Chantawannakul P. Phylogenetic analysis of *Nosema ceranae* isolated from European and Asian honeybees in Northern Thailand. J Invertebr Pathol. 2011; 107: 229–233. doi: 10.1016/j.jip.2011.05.012 PMID: 21600213

14. Graystock P, Yates K, Darvill B, Goulson D, Hughes WO. Emerging dangers: Deadly effects of an emergent parasite in a new pollinator host. J Invertebr Pathol. 2013; 114: 114–119. doi: 10.1016/j.jip.2013.06.005 PMID: 23816821

15. Plischuk S, Martín-Hernández R, Prieto L, Lucía M, Botías C, Meana A, et al. South American native bumblebees (Hymenoptera: Apidae) infected by Nosema ceranae (Microsporidia), an emerging pathogen of honeybees (Apis mellifera). Environ Microbiol Rep. 2009; 1: 131–135. doi: 10.1111/j.1758-2229.2009.00018.x PMID: 23765744

16. Webster TC. *Nosema apis* spore transmission among honey bees. Am Bee J. 1993; 133: 869–870.

17. Smith ML. The honey bee parasite Nosema ceranae: transmissible via food exchange? PLoS One. 2012; 7: e43319. doi: 10.1371/journal.pone.0043319 PMID: 22916241

18. OIE. Nosemosis of honey bees (NB: Version adopted in May 2013). In: Manual of Diagnostic Tests and Vaccines for Terrestrial Animals 2014. Available: http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.02.04_NOSEMOSIS_FINAL.pdf.

19. Higes M, Martín-Hernández R, Meana A. *Nosema ceranae* in Europe: an emergent type C nosemosis. Apidologie. 2010; 41: 375–392.

20. Goblirsch M, Huang ZY, Spivak M. Physiological and behavioral changes in honey bees (Apis mellifera) induced by Nosema ceranae Infection. PLoS One. 2013; 8: e58165. doi: 10.1371/journal.pone.0058165 PMID: 23483987

21. Dussaubat C, Maisonnasse A, Crauser D, Beslay D, Costagliola G, Soubeyrand S, et al. Flight behavior and pheromone changes associated to Nosema ceranae infection of honey bee workers (Apis mellifera) in field conditions. J Invertebr Pathol. 2013; 113: 42–51. doi: 10.1016/j.jip.2013.01.002 PMID: 23352958

22. Alaux C, Crauser D, Pioz M, Saulnier C, Le Conte Y. Parasitic and immune modulation of flight activity in honey bees tracked with optical counters. J Exp Biol. 2014; 217: 3416–3424. doi: 10.1242/jeb.105783 PMID: 25063861

23. Schmid-Hempel R. Parasites in social insects. Chichester, West Sussex, UK: Princeton University Press; 1998.

24. Higes M, Martín-Hernández R, Botías C, Bailón EG, González-Porto AV, Barrios L, et al. How natural infection by *Nosema ceranae* causes honeybee colony collapse. Environ Microbiol. 2008; 10: 2659–2669. doi: 10.1111/j.1462-2920.2008.01687.x PMID: 18647336

25. Runckel C, Flenniken ML, Engel JC, Ruby JG, Ganem D, Andino R, et al. Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, Nosema, and Crithidia. PLoS One. 2011; 6: e20656. doi: 10.1371/journal.pone.0020656 PMID: 21687739

26. Bailey L. The epidemiology and control of *Nosema* disease of the honey-bee. Ann Appl Biol. 1955; 43: 379–389.

27. Perry CJ, Sovik E, Myerscough MR, Barron AB. Rapid behavioral maturation accelerates failure of stressed honey bee colonies. Proc Natl Acad Sci USA. 2015; 112: 3427–3432. doi: 10.1073/pnas.1422089112 PMID: 25675508

28. Botías C, Martín-Hernández R, Barrios L, Meana A, Higes M. Nosema spp. infection and its negative effects on honey bees (Apis mellifera iberiensis) at the colony level. Vet Res. 2013; 44: 25. doi: 10.1186/1297-9716-44-25 PMID: 23574888

29. Gisder S, Hedtke K, Mockel N, Frielitz MC, Linde A, Genersch E. Five-year cohort study of Nosema spp. in Germany: does climate shape virulence and assertiveness of Nosema ceranae? Appl Environ Microbiol. 2010; 76: 3032–3038. doi: 10.1128/AEM.03097-09 PMID: 20228103

30. Dainat B, Evans JD, Chen YP, Gauthier L, Neumann P. Predictive markers of honey bee colony collapse. PLoS One. 2012; 7: e32151. doi: 10.1371/journal.pone.0032151 PMID: 22384162

31. Higes M, Meana A, Bartolomé C, Botías C, Martín-Hernández R. Nosema ceranae (Microsporidia), a controversial 21st century honey bee pathogen. Environ Microbiol Rep. 2013; 5: 17–29. doi: 10.1111/1758-2229.12024 PMID: 23757127

32. Gómez-Moracho T, Maside X, Martín-Hernández R, Higes M, Bartolomé C. High levels of genetic diversity in *Nosema ceranae* within *Apis mellifera* colonies. Parasitology. 2014; 141: 475–481. doi: 10.1017/S0031182013001790 PMID: 24238365

33. Chen Y, Pettis JS, Zhao Y, Liu X, Tallon LJ, Sadzewicz LD, et al. Genome sequencing and comparative genomics of honey bee microsporidia, *Nosema apis* reveal novel insights into host-parasite interactions. BMC Genomics. 2013; 14: 451. doi: 10.1186/1471-2164-14-451 PMID: 23829473

34. Fries I. Nosema ceranae in European honey bees (Apis mellifera). J Invertebr Pathol. 2010; 103 Suppl 1: S73–79. doi: 10.1016/j.jip.2009.06.017 PMID: 19909977

35. Chen YP, Evans JD, Murphy C, Gutell R, Zuker M, Gundensen-Rindal D, et al. Morphological, molecular, and phylogenetic characterization of *Nosema ceranae*, a microsporidian parasite isolated from the European honey bee, *Apis mellifera*. J Eukaryot Microbiol. 2009; 56: 142–147. doi: 10.1111/j.1550-7408.2008.00374.x PMID: 19457054

36. Sagastume S, del Aguila C, Martín-Hernández R, Higes M, Henriques-Gil N. Polymorphism and recombination for rDNA in the putatively asexual microsporidian *Nosema ceranae*, a pathogen of honeybees. Environ Microbiol. 2011; 13: 84–95. doi: 10.1111/j.1462-2920.2010.02311.x PMID: 21199250

37. Sagastume S, Martín-Hernández R, Higes M, Henriques-Gil N. Ribosomal gene polymorphism in small genomes: analysis of different 16S rRNA sequences expressed in the honeybee parasite Nosema ceranae (Microsporidia). J Eukariot Microbiol. 2014; 61: 42–50.

38. Roudel M, Aufauvre J, Corbara B, Delbac F, Blot N. New insights on the genetic diversity of the honeybee parasite *Nosema ceranae* based on multilocus sequence analysis. Parasitology. 2013; 140: 1346–1356. doi: 10.1017/S0031182013001133 PMID: 23880415

39. Gómez-Moracho T, Bartolomé C, Bello X, Martín-Hernández R, Higes M, Maside X. Recent worldwide expansion of Nosema ceranae (Microsporidia) in Apis mellifera populations inferred from

multilocus patterns of genetic variation. Infect Genet Evol. 2015; 31: 87–94. doi: 10.1016/j.meegid.2015.01.002 PMID: 25583446

40. Gómez-Moracho T, Bartolomé C, Martín-Hernández R, Higes M, Maside X. Evidence for weak genetic recombination at the *PTP2* locus of *Nosema ceranae*. Environ Microbiol. 2014.

41. Van der Zee R, Gómez-Moracho T, Pisa L, Sagastume S, García-Palencia P, Maside X, et al. Virulence and polar tube protein genetic diversity of Nosema ceranae (Microsporidia) field isolates from Northern and Southern Europe in honeybees (Apis mellifera iberiensis). Environ Microbiol Rep. 2014; 6: 401–413. doi: 10.1111/1758-2229.12133 PMID: 24992540

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

43. Burgher-MacLellan KL, Williams GR, Shutler D, MacKenzie K, Rogers REL. Optimization of duplex real-time PCR with melting-curve analysis for detecting the microsporidian parasites *Nosema apis* and *Nosema ceranae* in *Apis mellifera*. Can Entomol. 2010; 142: 271–283.

44. Ironside JE. Diversity and recombination of dispersed ribosomal DNA and protein coding genes in microsporidia. PLoS One. 2013; 8: e55878. doi: 10.1371/journal.pone.0055878 PMID: 23405227

45. Gatehouse HS, Malone LA. Genetic variability among Nosema apis isolates. J Apicult Res. 1999; 38: 79–85.

46. Gatehouse HS, Malone LA. The ribosomal RNA gene region of Nosema apis (Microspora): DNA sequence for small and large subunit rRNA genes and evidence of a large tandem repeat unit size. J Invertebr Pathol. 1998; 71: 97–105. PMID: 9547137

47. Eickbush TH, Eickbush DG. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics. 2007; 175: 477–485. PMID: 17322354

48. Cornman RS, Chen YP, Schatz MC, Street C, Zhao Y, Desany B, et al. Genomic analyses of the microsporidian Nosema ceranae, an emergent pathogen of honey bees. PLoS Pathog. 2009; 5: e1000466. doi: 10.1371/journal.ppat.1000466 PMID: 19503607

49. Gong J, Dong J, Liu X, Massana R. Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. Protist. 2013; 164: 369–379. doi: 10.1016/j.protis.2012.11.006 PMID: 23352655

50. Botías C, Martín-Hernández R, Dias J, García-Palencia P, Matabuena M, Juarranz A, et al. The effect of induced queen replacement on Nosema spp. infection in honey bee (Apis mellifera iberiensis) colonies. Environ Microbiol. 2012; 14: 845–859. doi: 10.1111/j.1462-2920.2011.02647.x PMID: 22118366

51. Martín-Hernandez R, Botías C, Bailón EG, Martínez-Salvador A, Prieto L, Meana A, et al. Microsporidia infecting Apis mellifera: coexistence or competition. Is Nosema ceranae replacing Nosema apis? Environ Microbiol. 2012; 14: 2127–2138. doi: 10.1111/j.1462-2920.2011.02645.x PMID: 22176602

52. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32: 1792–1797. PMID: 15034147

53. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser. 1999; 41: 95–98.

54. Evans JD, Schwarz RS, Chen YP, Budge G, Cornman RS, De la Rúa P, et al. Standard methodologies for molecular research in Apis mellifera. In: Dietemann V, Ellis JD, Neumann P, editors. The COLOSS BEEBOOK, Volume I: standard methods for Apis mellifera research: J Apic Res 52(4). 2013.

55. Walsh PS, Metzger DA, Higushi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10(4): 506–13 (April 1991). Biotechniques. 2013; 54: 134–139. PMID: 23599926

56. Garnery L, Solignac M, Celebrano G, Cornuet JM. A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. Experientia. 1993; 49: 1016–1021.

57. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013; 30: 2725–2729. doi: 10.1093/molbev/mst197 PMID: 24132122

58. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25: 1451–1452. doi: 10.1093/bioinformatics/btp187 PMID: 19346325

59. Nei M. Molecular evolutionary genetics. New York: Columbia University Press; 1987.

60. Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 1975; 7: 256–276. PMID: 1145509

61. Jukes TH, Cantor CR. Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism. New York: Academic Press. pp. 21–132. 1969.

62. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123: 585–595. PMID: 2513255

63. Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 1997; 147: 915–925. PMID: 9335623

64. Wright SI, Charlesworth B. The HKA test revisited: A maximum-likelihood-ratio test of the standard neutral model. Genetics. 2004; 168: 1071–1076. PMID: 15514076

65. Hudson RR, Kreitman M, Aguade M. A test of neutral molecular evolution based on nucleotide data. Genetics. 1987; 116: 153–159. PMID: 3110004

66. Polonais V, Prensier G, Metenier G, Vivares CP, Delbac F. Microsporidian polar tube proteins: highly divergent but closely linked genes encode *PTP1* and *PTP2* in members of the evolutionarily distant *Antonospora* and *Encephalitozoon* groups. Fungal Genet Biol. 2005; 42: 791–803. PMID: 16051504

67. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molecular Biology and Evolution. 2000; 17: 32–43. PMID: 10666704

68. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution. 2007; 24: 1586–1591. PMID: 17483113

69. Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics. 1985; 111: 147–164. PMID: 4029609

70. Myers SR, Griffiths RC. Bounds on the minimum number of recombination events in a sample history. Genetics. 2003; 163: 375–394. PMID: 12586723

71. Hudson RR. Two-locus sampling distributions and their application. Genetics. 2001; 159: 1805–1817. PMID: 11779816

72. McVean G, Awadalla P, Fearnhead P. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics. 2002; 160: 1231–1241. PMID: 11901136

73. Hudson RR, Boos DD, Kaplan NL. A statistical test for detecting geographic subdivision. Mol Biol Evol. 1992; 9: 138–151. PMID: 1552836

74. Hudson RR. A new statistic for detecting genetic differentiation. Genetics. 2000; 155: 2011–2014. PMID: 10924493

75. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010; 10: 564–567. doi: 10.1111/j.1755-0998.2010.02847.x PMID: 21565059

76. Lee SC, Corradi N, Doan S, Dietrich FS, Keeling PJ, Heitman J. Evolution of the sex-related locus and genomic features shared in microsporidia and fungi. PLoS One. 2010; 5: e10539. doi: 10.1371/journal.pone.0010539 PMID: 20479876

77. Malik SB, Pightling AW, Stefaniak LM, Schurko AM, Logsdon JM Jr. An expanded inventory of conserved meiotic genes provides evidence for sex in Trichomonas vaginalis. PLoS One. 2008; 3: e2879.

78. Lee SC, Corradi N, Byrnes EJ 3rd, Torres-Martinez S, Dietrich FS, Keeling PJ, et al. Microsporidia evolved from ancestral sexual fungi. Curr Biol. 2008; 18: 1675–1679. doi: 10.1016/j.cub.2008.09.030 PMID: 18976912

79. Fries I, Granados RR, Morse RA. Intracellular germination of spores of *Nosema apis* Z. Apidologie. 1992; 23: 61–70.

80. Singh RN, Sasidharan TO, Santha PC, Daniel AGK, Kamble CK. Role of temperature on multiplication and sporulation of *Nosema bombycis* in silk moth, *Bombyx mori* L. J Agric Sci. 2005; 18: 398–400.

81. Thomarat F, Vivares CP, Gouy M. Phylogenetic analysis of the complete genome sequence of *Encephalitozoon cuniculi* supports the fungal origin of microsporidia and reveals a high frequency of fast-evolving genes. J Mol Evol. 2004; 59: 780–791. PMID: 15599510

82. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. Genetics. 1998; 148: 1667–1686. PMID: 9560386

83. Xu Y, Weiss LM. The microsporidian polar tube: a highly specialised invasion organelle. Int J Parasitol. 2005; 35: 941–953. PMID: 16005007

84. Hirt RP, Logsdon JM Jr., Healy B, Dorey MW, Doolittle WF, Embley TM. Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. Proc Natl Acad Sci USA. 1999; 96: 580–585. PMID: 9892676

85. Ironside JE. Multiple losses of sex within a single genus of Microsporidia. BMC Evol Biol. 2007; 7: 48. PMID: 17394631

86. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Segurel L, Venkat A, et al. Revisiting an old riddle: what determines genetic diversity levels within species? PLoS Biol. 2012; 10: e1001388. doi: 10.1371/journal.pbio.1001388 PMID: 22984349

87. Haag KL, Traunecker E, Ebert D. Single-nucleotide polymorphisms of two closely related microsporidian parasites suggest a clonal population expansion after the last glaciation. Mol Ecol. 2013; 22: 314–326. doi: 10.1111/mec.12126 PMID: 23163569

88. Pombert JF, Xu J, Smith DR, Heiman D, Young S, Cuomo CA, et al. Complete genome sequences from three genetically distinct strains reveal high intraspecies genetic diversity in the microsporidian *Encephalitozoon cuniculi*. Eukaryot Cell. 2013; 12: 503–511. doi: 10.1128/EC.00312-12 PMID: 23291622

89. Charlesworth D, Morgan MT, Charlesworth B. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993; 134: 1289–1303. PMID: 8375663

90. Gupta RS, Golding GB. Evolution of *Hsp70* gene and its implications regarding relationships between archaebacteria, eubacteria, and eukaryotes. J Mol Evol. 1993; 37: 573–582. PMID: 8114110

91. Galkin VE, VanLoock MS, Orlova A, Egelman EH. A new internal mode in *F-actin* helps explain the remarkable evolutionary conservation of actin's sequence and structure. Curr Biol. 2002; 12: 570–575. PMID: 11937026

92. Tajima F. The effect of change in population size on DNA polymorphism. Genetics. 1989; 123: 597–601. PMID: 2599369

93. Wallberg A, Han F, Wellhagen G, Dahle B, Kawata M, Haddad N, et al. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. Nat Gen. 2014; 46: 1081–1088.

94. Han F, Wallberg A, Webster MT. From where did the Western honeybee (Apis mellifera) originate? Ecol Evol. 2012; 2: 1949–1957. doi: 10.1002/ece3.312 PMID: 22957195

95. Crane E. The World History of Beekeeping and Honey Hunting. London, UK: Gerald Duckworth and Company Ltd; 1999.

96. Whitfield CW, Behura SK, Berlocher SH, Clark AG, Johnston JS, Sheppard WS, et al. Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera*. Science. 2006; 314: 642–645. PMID: 17068261

97. Gray FH, Cali A, Briggs JD. Intracellular stages in the life cycle of the microsporidian *Nosema apis*. J Invertebr Pathol. 1969; 14: 391–394. PMID: 4983939

98. Selman M, Sak B, Kvac M, Farinelli L, Weiss LM, Corradi N. Extremely reduced levels of heterozygosity in the vertebrate pathogen *Encephalitozoon cuniculi*. Eukaryot Cell. 2013; 12: 496–502. doi: 10.1128/EC.00307-12 PMID: 23376943

99. Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, Becnel JJ, et al. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. Genome Res. 2012; 22: 2478–2488. doi: 10.1101/gr.142802.112 PMID: 22813931

100. Krebes L, Zeidler L, Frankowski J, Bastrop R. (Cryptic) sex in the microsporidian Nosema granulosis —evidence from parasite rDNA and host mitochondrial DNA. Infect Genet Evol. 2014; 21: 259–268. doi: 10.1016/j.meegid.2013.11.007 PMID: 24269340

101. Weedall GD, Hall N. Sexual reproduction and genetic exchange in parasitic protists. Parasitology. 2015; 142 Suppl 1: S120–127. doi: 10.1017/S0031182014001693 PMID: 25529755

102. Birky CW Jr. Giardia sex? Yes, but how and how much? Trends Parasitol. 2010; 26: 70–74. doi: 10.1016/j.pt.2009.11.007 PMID: 20022561

103. Desjardins CA, Sanscrainte ND, Goldberg JM, Heiman D, Young S, Zeng Q, et al. Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. Nat Commun. 2015; 6: 7121. doi: 10.1038/ncomms8121 PMID: 25968466

104. Lee SC, Ni M, Li W, Shertz C, Heitman J. The evolution of sex: a perspective from the fungal kingdom. Microbiol Mol Biol Rev. 2010; 74: 298–340. doi: 10.1128/MMBR.00005-10 PMID: 20508251