New insights on the genetic diversity of the honeybee parasite *Nosema ceranae* based on *multilocus* sequence analysis

MATHIEU ROUDEL[1,2], JULIE AUFAUVRE[1,2], BRUNO CORBARA[1,2], FREDERIC DELBAC[1,2] *and* NICOLAS BLOT[1,2]*

[1] *Clermont Université, Université Blaise Pascal, Laboratoire 'Microorganismes : Génome et Environnement', BP 10448, 63000 Clermont-Ferrand, France*
[2] *CNRS, UMR 6023, LMGE, 63171 Aubière, France*

SUMMARY

The microsporidian parasite *Nosema ceranae* is a common pathogen of the Western honeybee (*Apis mellifera*) whose variable virulence could be related to its genetic polymorphism and/or its polyphenism responding to environmental cues. Since the genotyping of *N. ceranae* based on unique marker sequences had been unsuccessful, we tested whether a *multilocus* approach, assessing the diversity of ten genetic markers – encoding nine proteins and the small ribosomal RNA sub-unit – allowed the discrimination between *N. ceranae* variants isolated from single *A. mellifera* individuals in four distant locations. High nucleotide diversity and allele content were observed for all genes. Most importantly, the diversity was mainly present within parasite populations isolated from single honeybee individuals. In contrast the absence of isolate differentiation precluded any *taxa* discrimination, even through a *multilocus* approach, but suggested that similar populations of parasites seem to infect honeybees in distant locations. As statistical evolutionary analyses showed that the allele frequency is under selective pressure, we discuss the origin and consequences of *N. ceranae* heterozygosity in a single host and lack of population divergence in the context of the parasite natural and evolutionary history.

Key words: *Nosema ceranae*, parasite, honeybee, diversity.

INTRODUCTION

Nosema spp. are obligate intracellular parasites belonging to Microsporidia, a cosmopolitan phylum derived as a basic branch of the fungi (Keeling and Fast, 2002; Capella-Gutiérrez *et al*. 2012). Microsporidia are characterized by the production of resistant spores carrying a polar tube, used to transfer the infectious sporal content into target cells, and by their 70S ribosomes with bacterial-sized rRNAs. The genus *Nosema* comprises more than 150 species that mainly infect invertebrate hosts (Becnel and Andreadis, 1999). The western honeybee (*Apis mellifera*) is parasitized by the long-known *Nosema apis* (Zander, 1909) and by the emerging *Nosema ceranae*. Both are the etiologic agents of nosemosis, one of the most widespread diseases of the honeybee. *Nosema ceranae* is thought to have jumped host from the Asian (*Apis cerana*) to the Western honeybee some decades ago and has now become the predominant microsporidian species infecting *A. mellifera* (Klee *et al*. 2007; Chen *et al*. 2008; Martin-Hernandez *et al*. 2012). *Nosema ceranae* has also been reported in other oriental *Apis* and *Bombus* species (Plischuk *et al*. 2009; Chaimanee *et al*. 2010, 2011; Li *et al*. 2012).

* Corresponding author: Université Blaise Pascal, Laboratoire 'Microorganismes: Génome et Environnement' CNRS UMR 6023, 63177 Aubière, France. E-mail: nicolas.blot@univ-bpclermont.fr

The documented *A. mellifera* colony losses observed for the last 15 years has become an alarming issue for the beekeeping and agricultural fields as well as for conservation biology, especially due to the pollinating activity of the honeybees (Potts *et al*. 2010). Although the origin of the honeybee decline has not been clearly identified and is thought to be multicausal, some studies have suggested a role of *N. ceranae* (Martin-Hernandez *et al*. 2007; Higes *et al*. 2008, 2010). Following the ingestion of spores, the parasites invade and develop within the cytoplasm of the epithelial cells of the adult honeybee midgut. *N. ceranae* is considered as a major threat to the Western honeybee at both the individual and colony levels (Fries, 2010; Higes *et al*. 2013). It would increase the honeybee mortality, activate the degeneration of the infected cells, induce an energetic stress and inhibit transcripts involved in immunity (Antùnez *et al*. 2009; Martin-Hernandez *et al*. 2011; Dussaubat *et al*. 2012; Chaimanee *et al*. 2012). However the presence of *N. ceranae* is not systematically associated with honeybee weakening and mortality (Cox-Foster *et al*. 2007; Invernizzi *et al*. 2009; Gisder *et al*. 2010), suggesting modulations in the parasite virulence. Thus, not only the role of *N. ceranae* in the honeybee decline but also the origin of the variation in its virulence remains elusive. Some authors have suggested that such variation could be related to a polymorphism between *N. ceranae* isolates (Fries, 2010; Higes *et al*. 2013),

meaning that different parasite variants may infect the honeybee, some of them inducing a higher virulence.

Although the genome of the BRL01 strain of *N. ceranae* has been sequenced (Cornman *et al.* 2009), data are lacking regarding the comparison of the gene content between *N. ceranae* isolates and the assessment of *N. ceranae* genetic polymorphism *sensu stricto*, i.e. its genetically determined pathogenicity. Several studies have sought to evaluate the parasite genetic polymorphism *sensu lato*, i.e. the nucleotide variation in sequences that appeared relevant for the genotyping of strains in other microsporidia, but failed to discriminate between *N. ceranae* isolates. The nucleotide diversity of the ribosomal RNA coding region (rDNA) successfully discriminated between isolates of the mammals' microsporidian parasites *Encephalitozoon cuniculi*, *Encephalitozoon hellem* and *Enterocytozoon bieneusi* (Didier *et al.* 1995; Mathis *et al.* 1999; Haro *et al.* 2003; ten Hove *et al.* 2009) but not of *N. ceranae* sampled in *A. mellifera* (Huang *et al.* 2008; Chen *et al.* 2009a; Sagastume *et al.* 2011). The rDNA diversity observed in *N. ceranae* and the bumblebees' parasite *Nosema bombi* (Tay *et al.* 2005; O'Mahony *et al.* 2007; Sagastume *et al.* 2011) have suggested that rDNA may escape the concerted evolution maintaining sequence identity between the multiple gene copies present within *Nosema* genomes and would consequently be inappropriate to phylogenetic analyses (Ironside, 2007; Fries, 2010). The nucleotide diversity of virulence-related genes encoding Polar Tube Proteins (PTPs) and Spore Wall Proteins (SWPs) has allowed determining variants of *E. cuniculi* and *E. hellem* (Peuvel *et al.* 2000; Haro *et al.* 2003; Polonais *et al.* 2010). However, neither PTP1 nor PTP3 encoding genes permitted the distinction of *N. ceranae* isolates (Chaimanee *et al.* 2011; Hatjina *et al.* 2011). Since the polymorphism of rDNA and PTP3 genes seems to be a major obstacle for genotyping (Hatjina *et al.* 2011; Sagastume *et al.* 2011) one may wonder if the use of several independent gene markers may resolve parasite taxa.

In order to assess whether *N. ceranae* isolates can be discriminated through a *multilocus* approach, we compared the nucleotide and haplotype diversity of ten *N. ceranae* genes, including the small-subunit ribosomal RNA (SSU-rDNA) gene and nine protein encoding genes, that appeared both within and between parasite populations isolated from four geographically distant honeybees.

MATERIALS AND METHODS

Honeybee sampling

Apis mellifera individuals were collected in France, Morocco, Lebanon and Thailand in 2010 and 2011. The presence of microsporidian parasites has been verified by both the microscopic observation of spores and the amplification of the microsporidian SSU-rDNA using the Nc_D1 (CGACGATGTGA-TATGAAAATATTAA) and Nc_R1 (TCATTCT-CAAACAAAAAACCGTTC) primers. The absence of *N. apis* was checked using the Na_D1 (GCAT-GTCTTTGACGTACTATGTAC) and Na_R1 (CGTTTAAAATGTGAAACAACTATG) primers. All primers derived from Martin-Hernandez *et al.* 2007). The isolate concept was defined here as the parasite population infecting a single individual, thus only one bee from each location was used for subsequent cloning.

Genetic markers amplification and cloning

Ten genetic markers have been amplified by PCR (Platinum *Taq* Polymerase High Fidelity, Invitrogen) in a mixture containing $2\,mM\,MgSO_4$ and $0.4\,\mu M$ of each primer (Table 1) in a total volume of $25\,\mu L$, and performing annealing steps at $52\,°C$. PCR products were inserted into the pCRII-TOPO cloning vector (TOPO TA Cloning Kit Dual Promoter, Invitrogen). After transformation (NEB 5-alpha cells, New England Biolabs), several clones were cultured and their plasmids content extracted. Inserted fragments were sequenced (GATC Biotech, Germany) using both the T7 and M13 Reverse flanking primers. To check the absence of experimentally induced sequence variation, a control has been performed using a clonal plasmid preparation carrying the SWP30 marker as an initial PCR template.

Statistical evolutionary analyses

To determine variable sites, nucleotide sequences were aligned using ClustalW under the BioEdit editor (Hall, 1999), selecting the amplified fragments only, i.e. removing primer sequences. All polymorphic sites were visually checked for quality on electropherograms. Alignments with insertions-deletions (*indels*) of tandem repeats were manually improved for better parsimony. Diversity, transition to transversion and $d_N:d_S$ analyses were conducted in MEGA5, removing ambiguous positions in pairwise comparisons (pairwise deletion) allowing to consider *indels* as polymorphic characters (Nei and Kumar, 2000; Tamura *et al.* 2011). The nucleotide diversity has been given by averaging the pairwise *p*-distances between sequences, i.e. the mean number of base differences per site, for the whole allelic population (π) and for populations within (π_w) or between (π_b) honeybee isolates. To measure how isolates subpopulations differed, the nucleotide diversity explained by population structure and the magnitude of gene differentiation among subpopulations were given by the fixation index (F_{ST}) and the coefficient of evolutionary differentiation (G_{ST}), respectively.

Table 1. Marker genes and primers used for genetic variability (^{a}) and RT-PCR (^{b}) analyses

ORF name in *N. ceranae* BRL01 genome	Forward and reverse primers (5′ to 3′)	References
SSU-rDNA (DQ486027)	MICRO-F2 CACCAGGTTGATTCTGCCTG[a,b]	From Visvesvara *et al*. (1995)
	Univ-1390 GACGGGCGGTGTGTACAA[a,b]	Zheng *et al*. (1996)
EnP1B (NCER_100768)	EnP1B-F GCATTTGTATTTGGGTGG[a,b]	This study
	EnP1B-R ACCAGCAGGTCTGCAATC[a,b]	
PTP2 (NCER_101590)	PTP2-F GTAGCCAAGTTGCCACCT[a,b]	This study
	PTP2-R GGGTTCTGCATCCTTGTC[a,b]	
HSWP4 (NCER_100828)	SWP4-F CATGAACTTAAGGGATTAGCAAG[a,b]	This study
	SWP4-R CTCTTCGTGTGCATCATCCG[a,b]	
SWP25 (NCER_100064)	SWP25-F GGTTTATGTAGCATTGAGGATGC[a,b]	This study
	SWP25-R TGCTGGCATACAAGTTCCATC[a,b]	
SWP30 (NCER_100566)	SWP30-F ATGCGGGAATATGAACATCC[a,b]	This study
	SWP30-R CTTTAACAACTGTGTTCTTTCGTAC[a,b]	
NCER_100070	NCER_100070-F GCCGGCAACAAAACCTATC[a]	This study
	NCER_100070-R CGTTACTGGCATCACTGTTGG[a]	
	NCER_100070_midF AGGCAGGACTCTTCGATCAC[b]	This study
	NCER_100070_midR TCATCCTGCGGTTTTATTGG[b]	
NCER_100533	NCER_100533-F GAATTGGGATTTGGAGGAGG[a]	This study
	NCER_100533-R TCTGAAGGGTCGGGTGAAG[a]	
	NCER_100533_endF CCGAGGGTCAACAAAATGAC[b]	This study
	NCER_100533_endR ACATTCCCTGTGGCAAACTC[b]	
NCER_101600	NCER_101600-F GCCTATGACTTCATTACAAGAAGAATC[a,b]	This study
	NCER_101600-R TCACTCGTAAGTTTCGCATTTG[a,b]	
NCER_101165	NCER_101165-F AATCCGCTAACTCGACCTTC[a]	This study
	NCER_101165-R GAGATTTCGGGTCGACAATG[a]	
	NCER_101165_endF TCGACCCGAAATCTCAACTC[b]	This study
	NCER_101165_endR GTGTGCCTTCATCATCATCG[b]	

Standard error estimates were obtained by a bootstrap procedure with 10 000 replicates. The overall transition/transversion bias ($R_{s/v}$) was computed from the Maximum Composite Likelihood estimate of the pattern of nucleotide substitution. The codon-based one-tailed test of selection has been performed by calculating the probability of rejecting the null hypothesis of strict-neutrality ($d_N = d_S$) in favour of the alternative hypotheses, $d_N < d_S$ or $d_N > d_S$ for purifying or positive selection respectively, using the Li–Wu–Luo method. Statistical significance (*P*-values <0·05 were considered significant) was computed by a 10 000-replicates bootstrap resampling procedure. All tests were in favour of $d_N < d_S$ rather than $d_N > d_S$.

Other selection neutrality tests, linkage disequilibrium (LD) and recombination analyses were conducted in DnaSP (Librado and Rozas, 2009; Rozas, 2009) removing all positions containing gaps and missing data (complete deletion). Divergence from the null hypothesis, i.e. from the hypothesis that all mutation are either neutral or strongly deleterious, was estimated computing the Tajima's *D* (Tajima, 1989) and Ramos-Onsins & Rozas's R_2 (Ramos-Onsins and Rozas, 2002), based on the nucleotide polymorphism, and the Fu's F_s (Fu, 1997), based on the haplotype distribution. The LD, i.e. the non-random association of alleles, was estimated by calculating *ZZ*, which equals the average LD between adjacent polymorphic sites minus the standardized average intragenic LD (Rozas *et al*. 2001). Statistical significance (*P*-values <0·05 were considered significant) was given by the probability of obtaining under the null hypothesis absolute values equal or higher than the observed ones. It was computed by performing coalescent simulation with 10 000 replicates using fixed theta per sequence values and no recombination, no strong difference being observed using a recombination parameter. All significant values were out of the confidence interval comprising 95% of the expected values under the null hypothesis (data not shown), i.e. they were very unlikely under the standard neutral model. The number of pairs of polymorphic sites that exhibit statistically significant LD (*P* < 0·05) performing a two-tailed Fisher's exact test, with and without Bonferroni correction against multiple comparisons biases, is given together with the total amount of considered pairwise comparisons.

Phylogenetic trees were computed under MEGA 5, using the Maximum Likelihood algorithm, with a 10 000-replicates bootstrap procedure, a nucleotidic substitution model using the Tamura 3-parameters method and a Nearest-Neighbour-Interchange heuristic method.

Sequence prediction and modelling

Prediction of potentially secreted proteins was based on the screening for potential N-terminal secretory signal sequences using SignalP v3.0 (Bendtsen *et al*.

2004). Peptide repeated motifs in EnP1B and NCER_101600 were manually extracted and schematized using the WebLogo tool (Crooks *et al.* 2004).

RNA extraction and RT-PCR

Total RNA was extracted from the midgut of one naturally infected honeybee in France (SV total RNA Isolation System Kit, Promega). cDNA was synthesized (ImProm-II™ Reverse Transcription System, Promega) using an oligo-d(T) primer (Invitrogen) and 100 ng of RNA. Subsequent PCR amplifications were performed (GoTaq® Flexi DNA Polymerase, Promega) using the RT-PCR primers listed in Table 1. Negative and positive controls were performed using reactions without the reverse transcriptase and *N. ceranae* genomic DNA as a template respectively. All controls validated the experiment.

RESULTS

Selection of N. ceranae *marker genes*

Ten marker genes have been selected within the *N. ceranae* BRL01 genome (Cornman *et al.* 2009), including the nearly complete rRNA small subunit encoding gene. Four spore wall and polar tube protein encoding genes have been identified and named according to their closest known homologues in *Nosema bombycis* (Table 1). They encode the polar tube protein PTP2, the spore wall proteins SWP25 and SWP30, and the hypothetical spore wall protein HSWP4 (Wu *et al.* 2008, 2009). Except for its first 32 codons, the NCER_100768 gene encodes a protein homologous to the spore wall protein EnP1 of *Encephalitozoon cuniculi* (Peuvel-Fanget *et al.* 2006). We propose to shift the start codon from the suggested GTG to the first upstream ATG codon and to name the resulting open reading frame EnP1B, as another gene (NCER_101082) with higher homology with *E. cuniculi* EnP1 is already present in *N. ceranae* genome. Lastly, four genes encoding hypothetical proteins carrying tandem peptide repeats and potential N-terminus secretion signal were selected as potential surface and/or structural proteins that could show high variability. RT–PCR experiments (data not shown) proved that all the chosen open reading frames are transcribed in a naturally *N. ceranae*-infected honeybee, demonstrating that they are true protein-encoding genes that ought to be subjected to evolutionary processes.

Within- vs between-isolate distance of N. ceranae *marker genes*

Following amplification and cloning, 8–15 clone sequences were obtained for each marker and each isolate and have been deposited in GenBank (Accession Numbers KC680230 to KC680656). The experimental control performed using a clonal matrix did not show any sequence variation in the 15 obtained clones, showing that no error-prone bias was brought throughout the procedure, i.e. any sequence diversity observed ought to depict true allelic variability. Whatever the tested marker, a surprisingly high allele content – mostly due to singletons – has been observed, even for populations isolated from single host individuals (Table 2). In order to compare the genetic variability of the markers, pairwise *p*-distances between sequences have been averaged to the nucleotide diversities present in the whole allelic population (π), within a single honeybee (π_w) and between geographically distant honeybees (π_b, Table 2 and Fig. 1). For each marker, the π_w and π_b were very similar, showing that there is as much variability within a single bee as between geographically distant bees. Even *p*-distance maxima were very similar within and between isolates (Fig. 1), showing that the most divergent sequences can be found within a single parasitized individual. Moreover, whatever the marker, the minimum *p*-distances were null both within and between isolates, showing that there are always identical alleles within an isolate but also between distant isolates. Altogether, these data suggest that the parasite populations infecting honeybees were similar in the four locations studied. Indeed, the F_{ST} and G_{ST} indices for population differentiation between isolates were low (Table 2), indicating that the parasite populations from distinct isolates are poorly divergent.

The nucleotide diversity appeared variable between markers, with π ranging from 0·001 to 0·018 for HSWP4 and EnP1B respectively (Fig. 1). However, these values are still too low, and especially much lower than their corresponding π_w, to differentiate the isolate populations. Indeed phylogenetic analyses failed to separate isolate-specific taxa: whatever the chosen marker and the evolutionary model, trees had star-like topologies with poorly supported nodes by bootstrap procedures (Fig. 2). Altogether, these data suggested that the markers high nucleotide diversity but low divergence between isolates preclude any *multilocus* approach for the genotyping of *N. ceranae* strains.

Nature of polymorphic characters

The high transition to transversion bias observed for all markers suggests that they are submitted to evolutionary processes including selection. Indeed all markers exhibited rejection of at least one neutrality test (shown in bold in Table 2). For most protein encoding genes, the codon-based one-tailed test of selection significantly rejected the null hypothesis in favour of the purifying hypothesis ($d_N < d_S$). However, four markers also exhibited nonsense mutations, leading to the encoding of non-functional proteins, suggesting that the markers may be submitted to complex evolutionary processes. EnP1B

Table 2. Statistics of nucleotide polymorphism, divergence, linkage disequilibrium and recombination in ten gene regions of *N. ceranae*

| Marker	SSU-rDNA	EnP1B	PTP2	HSWP4	SWP25	SWP30	NCER_100070	NCER_100533	NCER_101600	NCER_101165
Length of analysed sequence	1116 bp	732 bp	474 bp	650 bp	655 bp	537 bp	698 bp	671 bp	902 bp	660 bp
Nb of sequences	39	41	44	42	39	42	40	37	59	44
Nb of alleles	37	33	22	13	20	25	22	14	27	25
Variable sites	66	56	26	14	23	35	23	18	48	31
Unique variable sites	55	24	22	14	22	28	16	17	31	24
Size (and number) of *indels*	1(x6)1, 2, 4	21(x2),24, 45, 87	–	–	–	–	1(×4),3	1,36,144	1,60,85,108,114, 150,216,228,444	–
Entire population nucleotide diversity, π	0·00446±0·00081	0·01782±0·00283	0·00503±0·00154	0·00102±0·00026	0·00231±0·00067	0·00766±0·00199	0·00635±0·00194	0·00155±0·00036	0·00716±0·00161	0·00695±0·00193
Within isolate mean p-distance, π_w	0·00454±0·00083	0·01468±0·00245	0·00470±0·00140	0·00094±0·00024	0·00209±0·00051	0·00779±0·00206	0·00513±0·00149	0·00157±0·00037	0·00737±0·00162	0·00515±0·00133
Between isolate mean p-distance, π_b	0·00453±0·00092	0·01723±0·00306	0·00495±0·00162	0·00095±0·00033	0·00233±0·00083	0·00765±0·00206	0·00664±0·00207	0·00157±0·00052	0·00725±0·00161	0·00744±0·00217
Mean interpopulation diversity, FST	−0·00008±0·00008	0·00204±0·00044	0·00031±0·00022	0·00008±0·00004	0·00023±0·00022	−0·00010±0·00015	0·00118±0·00004	−0·00001±0·00006	−0·00021±0·00010	0·00180±0·00067
Isolate differentiation coefficient, G_{ST}	−0·02±0·02	0·12±0·02	0·06±0·03	0·08±0·03	0·10±0·07	−0·01±0·02	0·19±0·03	−0·01±0·04	−0·03±0·01	0·26±0·04
Mean pairwise transition to transversion ratio, s/v	2·86±0·69	6·01±2·29	2·14±0·92	0·31±0·15	1·53±0·79	3·92±1·47	5·17±2·21	ND (only transitions)	3·86±2·06	2·77±2·16
Overall transition to transversion bias, $R_{s/v}$	4·10	5·46	3·63	13·02	8·93	6·86	4·81	345·8	4·70	3·56
Nb of synonymous, missense, nonsense SNPs	–	29, 26, 1	7, 19, 0	6, 8, 0	10, 12, 1	13, 21, 0	6, 17, 0	13, 4, 1	23, 25, 0	13, 17, 1
Rejection of $d_N = d_S$ in favour of $d_N < d_S$ or $d_N > d_S$ (p)	–	$d_N < d_S$ (0·000)	$d_N < d_S$ (0·101)	$d_N < d_S$ (0·049)	$d_N < d_S$ (0·055)	$d_N < d_S$ (0·025)	$d_N < d_S$ (0·340)	$d_N < d_S$ (0·001)	$d_N < d_S$ (0·001)	$d_N < d_S$ (0·071)
Tajima's D (p)	**−2·45** (0·000)	−0·36 (0·400)	**−2·07** (0·002)	**−2·49** (0·000)	**−2·43** (0·000)	**−1·70** (0·022)	−0·55 (0·339)	**−2·43** (0·000)	−0·94 (0·186)	−1·28 (0·094)
Fu's F_S (p)	**−36·36** (0·000)	**−17·18** (0·000)	**−17·54** (0·000)	**−12·40** (0·000)	**−21·22** (0·000)	**−15·77** (0·000)	**−6·73** (0·009)	−9·36 (0·000)	−2·41 (0·182)	−13·58 (0·000)
Ramos-Onsins and Rozas's R_2 (p)	**0·030** (0·000)	0·093 (0·314)	**0·038** (0·000)	**0·043** (0·000)	**0·030** (0·000)	**0·047** (0·004)	0·085 (0·225)	0·051 (0·000)	0·079 (0·251)	0·061 (0·042)
Haplotype diversity (expected heterozygosity), Hd (p)	0·985 (0·999)	0·974 (0·999)	0·918 (0·999)	0·460 (0·531)	0·849 (0·982)	0·928 (0·990)	0·845 (0·567)	0·511 (0·522)	0·617 (0·105)	0·920 (0·822)
Intragenic recombination, ZZ (p)	0·052 (0·822)	0·062 (0·926)	−0·023 (0·395)	−0·033 (0·208)	−0·017 (0·382)	0·002 (0·559)	0·019 (0·657)	0·022 (0·738)	0·201 (0·996)	0·096 (0·946)
Polymorphic pairs with significant LD ^{w}/, ^{w}/$_o$ correction (total pairs)	69·0 (2016)	104·29 (903)	9·1 (300)	3·0 (91)	4·0 (253)	25·5 (595)	23·21 (231)	5·0 (91)	45·27 (171)	22·10 (435)
Minimum number of recombination events (R_m)	2	6	0	0	0	2	2	0	4	1
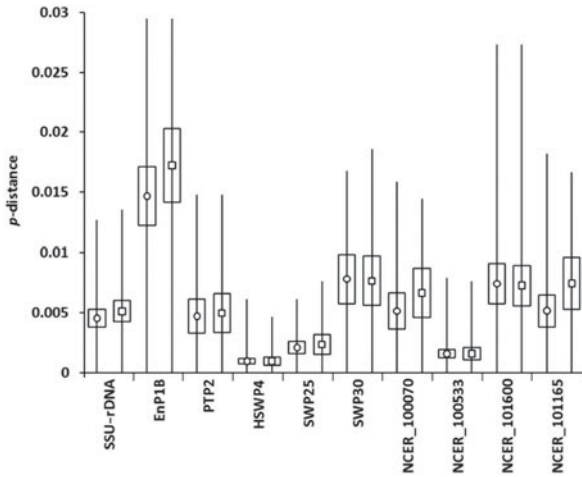
Fig. 1. Nucleotide diversity of ten *N. ceranae* genes. The nucleotide diversity is represented as the mean *p*-distance (i.e. the number of base differences per site) between sequences within an isolate (π_w, circles) and between isolates (π_b, squares). Standard error estimates obtained by a 10 000-replicates bootstrap procedure are shown by boxes. Lines represent the complete range of pairwise *p*-distances, showing the minimum and maximum obtained values.

and NCER_101600 showed large *indel* events corresponding to discrete numbers of base triplets thus not inducing frameshift. For EnP1B, *indels* were linked to a difference in the repetition of a 7-amino acid motif (Figs 3 and 4). A more complex pattern could be highlighted for the *indels* observed in NCER_101600, which contained two close but distinct 14- or 16-amino acid repeated motifs (Fig. 4). These *indels*, up to 444 base pairs (bp), generated a large variation in the length of NCER_101600 amplified products. Although they were not present in all isolates in our dataset, the dominant amplification products at 573, 801 and 951 bp were visible by agarose gel electrophoresis for all isolates (data not shown).

Among the segregating sites observed in the SSU-rDNA in the present and other deposited data, most were single nucleotide polymorphisms (SNPs), 70·8 and 16·7% being transitions and transversions respectively, but 12·5% of the sites showed *indel* events. The variable sites have been located on a proposed secondary structure model of the corresponding RNA subunit (Supplementary Fig. S1 – in Online version only). Helices were less submitted to variability than non-helix domains and most of the mutations in helices did not strongly modify the base pairing of the structure, i.e. changing a canonical (or non-canonical) bond to another one, suggesting non-random mutagenic events.

DISCUSSION

The variability of N. ceranae *gene sequences*

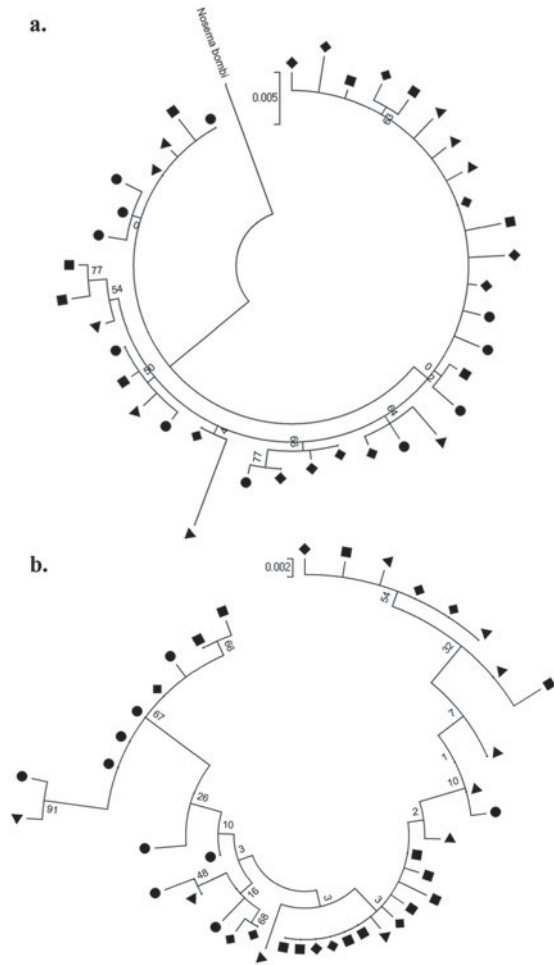Since one gene itself does not allow assessing the diversity of *N. ceranae* parasites in *A. mellifera*,

Fig. 2. Phylogenetic analysis of the *N. ceranae* small ribosomal RNA subunit (a) and hypothetical spore wall protein EnP1B (b) encoding genes from four geographically distant locations. The evolutionary history analysis was inferred by the maximum likelihood method using the multiple sequences obtained from four individuals sampled in France, Lebanon, Morocco and Thailand (indicated by triangle, star, square and circle respectively). The first tree was rooted using the close *Nosema bombi* SSU sequence as an outgroup while the second tree was unrooted since no close homologue of EnP1B was found in databases. The percentage of trees in which the associated taxa clustered together, inferred by a 10 000-replicates bootstrap analysis, is shown next to the branches.

we tested whether a *multilocus* approach could discriminate between populations of the parasite. This work showed that every gene presents unexpected high nucleotide diversity within a single host individual (π_w) but a poor genetic variation between parasite populations isolated from geographically distant honeybees (F_{ST} and G_{ST}, Table 2 and Fig. 1). Thus, while *N. ceranae* showed high sequence variability, the later did not allow any discrimination between parasite variants, even using concomitantly several genetic markers, implying that any phylogenetic assertions on *N. ceranae* must be considered with care.
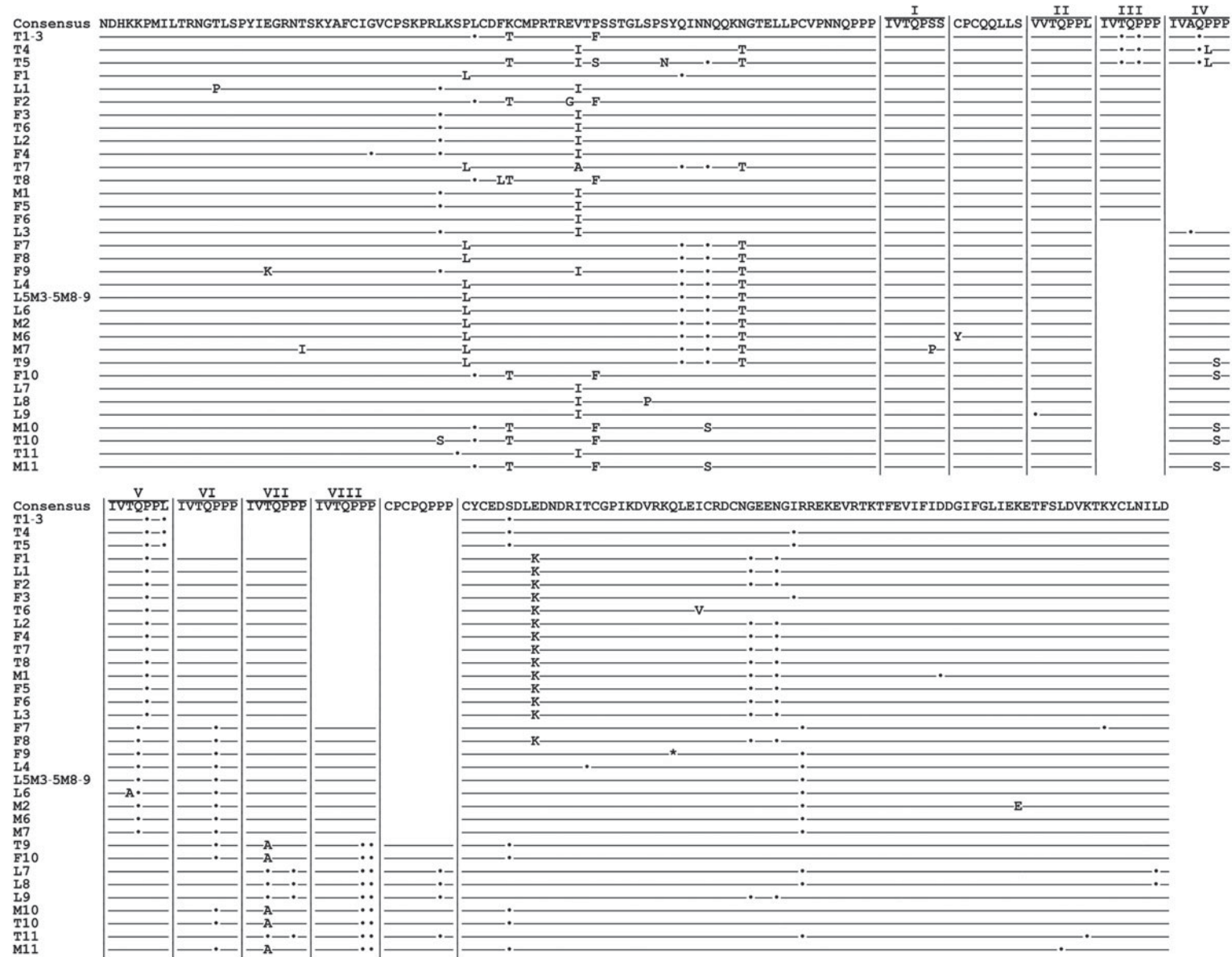
Fig. 3. Alignment of the partial EnP1B protein sequences inferred by the sequencing of the corresponding gene in single individuals sampled in France, Lebanon, Morocco and Thailand (as indicated by their initials on the left). The consensus sequence is indicated in the first row. Below, lines stand for identical sequences, letters for amino acid substitutions, dots for synonymous nucleotide substitutions (i.e. without change of the corresponding amino acid), blank areas for deletions. The star indicates a nonsense substitution (i.e. inducing a stop codon). Repeated motifs are separated by columns and have been numbered with Latin letters on their top.
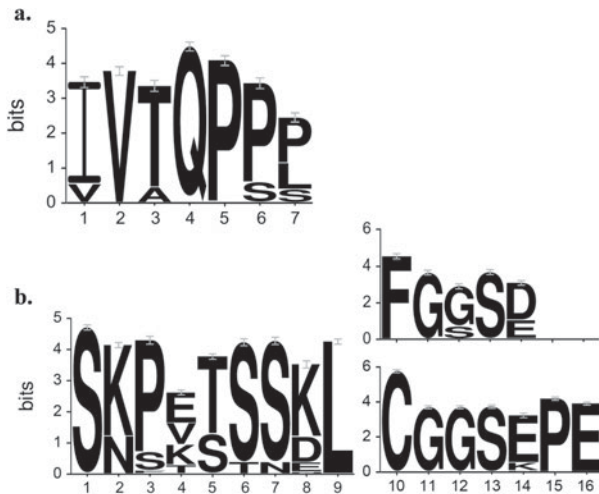
Fig. 4. Logo of the repeated motifs of EnP1B (a) and NCER_101600 (b) inferred by aligning the 264 and 678 respective peptide sequences extracted from our data. The height of each letter is proportional to the observed frequency of the corresponding amino acid and the overall height of each stack of letters is proportional to the sequence conservation at that position (with 4·32 bits as a maximum sequence conservation).

For all the tested markers, most of the alleles observed were singletons (Table 2). This low redundancy suggests that the true allelic richness was certainly underestimated. For instance not all *indels* of the NCER_101600 marker have been cloned and sequenced in every isolate whereas corresponding bands had been observed in agarose gel electrophoresis. Thus variations in nucleotide diversity between genes (Fig. 1) may be due to their intrinsic variability and allele selection as well as to a lack of data. Concerning the nucleotide diversity of the SSU-rDNA, although it seems variable in microsporidia, similar levels have been observed in the rDNA regions of *N. bombycis, Nosema granulosis* and *Vairimorpha cheracis* (π ranging from 0·004 to 0·009; Ironside, 2013) and a heterogeneity in rDNA sequences had also been observed in *N. apis* and *N. bombi* (Gatehouse and Malone, 1998; Tay *et al.* 2005; O'Mahony *et al.* 2007), suggesting that high nucleotide diversity might be common in the *Nosema/Vairimorpha* group.

A lack of N. ceranae *population divergence between geographically distinct* A. mellifera *hosts*

The low F_{ST} and G_{ST} indices (Table 2) indicated that no differentiation between *N. ceranae* populations could be assessed upon the geographic origin of their *A. mellifera* host. Such lack of divergence had already been observed for the rDNA marker and the genes encoding the polar tube proteins PTP1 and PTP3 (Huang *et al.* 2008; Chen *et al.* 2009*a*; Chaimanee *et al.* 2011; Hatjina *et al.* 2011; Sagastume *et al.* 2011). Our data inferred that this can certainly be

extrapolated to any coding genes of *N. ceranae* although there are no data based on selection-free silent sites and microsatellites to enlarge this to the whole genome sequence. Moreover, our results clearly confirmed the previously suggested hypothesis that geographically distant honeybees may be infected by similar parasite populations (Huang *et al.* 2008, Chen *et al.* 2009*a*, Sagastume *et al.* 2011).

The significant R_2 and the significantly negative Tajima's D and Fu's F_s statistics (Table 2) are characteristic of a population expansion (Ramirez-Soriano *et al.* 2008). This is particularly supported by the almost always significantly negative F_s (for all markers but NCER_101600) that is thought to be a more sensitive indicator of population expansion than D. The very weak population divergence observed through F_{ST} and G_{ST} (Table 2) and the star-like phylogenetic dendrograms (Fig. 2) suggest that such population expansion would have emerged from a poorly diverse initial population. It has been shown that the nucleotide diversity of the PTP1 gene allowed separating specific taxa of *N. ceranae* in *Apis florae* and in *Apis dorsata*, but no distinction could be found amongst parasites isolated from *Apis cerana* and *A. mellifera* (Chaimanee *et al.* 2011). Thus parasite populations may yet be discriminated relative to their host niche but a very similar population infects the Asian and the European honeybee species. Altogether, these findings support the hypothesis that *N. ceranae* experienced a recent host-jump from the Asian honeybee *A. ceranae* to the Western honeybee *A. mellifera* followed by a rapid spreading in its new host niche (Paxton *et al.* 2007; Klee *et al.* 2007; Chen *et al.* 2008; Botias *et al.* 2012) that may have been favoured by sustained intercontinental honeybee exchanges (Klee *et al.* 2007; Mutinelli, 2011). Such a host-jump may have constituted a bottleneck-like event, with a strong genetic drift followed by a rapid increase in population size. Such evolutionary history would also be reflected by negative D and F_s values and significant R_2 (Fay and Wu, 1999; Ramirez-Soriano *et al.* 2008) and by the short and weakly supported branching of the markers' evolutionary models (Fig. 2). However if most lineages did not survive the bottleneck event, a long branch would be expected between the taxonomic units related to the *A. cerana* and the *A. mellifera* hosts. This was not observed by Chaimanee *et al.* (2011), suggesting that the parasite populations infecting the former and new hosts may not have been completely isolated yet.

Origin and maintenance of the N. ceranae *genetic variation*

Mixed infection of one host by several parasite genotypes may affect the host–parasite interactions with benefits to the parasite infectivity and development,

as shown for the microsporidian *Octosporea bayeri* in *Daphnia magna* (Vizoso and Ebert, 2005). However, in agreement with Sagastume *et al.* (2011), it is hardly conceivable that geographically distant honeybees are systematically infected by a similar non-clonal population of parasites, and that this genotype diversity is maintained throughout the parasite development. Thus the high *N. ceranae* haplotype diversity or expected heterozygosity *Hd* (Table 2), with many haplotypes per isolates, would rather be linked to a combination of non-haploid stages, recombination events and/or gene exchange.

The rDNA is present in multiple *loci* within *N. ceranae* genome (Cornman *et al.* 2009) and its nucleotide diversity could be related to an intragenomic polymorphism of these gene copies, as observed within a single spore of *N. bombi* (O'Mahony *et al.* 2007). As the SSU-rDNA marker did not show stronger diversity than protein encoding markers (Fig. 1, Table 2), one may wonder whether all genes would be present in multiple polymorphic copies, questioning the polyploidy of the parasite. The observation of frameshift mutations, due to nonsense SNPs or 1-bp deletions, supports this hypothesis, as it would allow the accumulation of mutations in inactivated genes, leading to pseudogenization while maintaining functional allele copies under selective pressure. The observed *Hd* is not significantly different to what would be expected for a diploid organism (Table 2). *Nosema ceranae* is a dikaryotic single cell organism (Chen *et al.* 2009*b*) and, although the ploidy of the two nuclei is not known, *N. ceranae* could be considered to be at least diploid.

High heterozygosity with low F_{ST} could be interpreted as a clonal and asexual life cycle, with transient stages of di- or polyploidy, resulting in the same allelic population in isolates and leading to incorrect phylogenetic topology assertions (Birky, 1996; Haag *et al.* 2013). However a clonal mode of reproduction should lead to strong LD whereas the intragenic recombination given as *ZZ* values was never significant (Table 2). Yet the R_m values and the detection of polymorphic site pairs with significant LD suggest that recombinations did occur between alleles. Through meiosis, sex is a source of recombination and gene exchange, but it would have been lost independently in distinct microsporidian clades (Ironside, 2007). Despite the lack of evidence for sexual reproduction in *N. ceranae*, its genome contains sex-related *loci* and genes related to the meiotic recombination machinery (Lee *et al.* 2010*a*, *b*) and a potential sexual haplo-diploid cycle of the parasite has been hypothesized (Sagastume *et al.* 2011). Only deep molecular studies of clonal cultured populations of *N. ceranae* could resolve its degree of ploidy, recombination and sexual/asexual cycle.

EnP1B and NCER_101600, which are the two markers showing numerous *indels* of repeated motifs (Fig. 4), were the most polymorphic markers (π),

with the highest isolate differentiation (F_{ST}), and showed more polymorphic site pairs with significant LD and higher R_m, i.e. more recombination events are required to explain the observed haplotype diversity (Table 2). These two genes seem to have experienced slightly independent evolutionary history that could be related to their potential involvement in pathogenesis. Indeed the *E. cuniculi* EnP1 contains repeated peptides related to heparin-binding motifs and is involved in the parasite adhesion to the host cell surface *in vitro* (Southern *et al.* 2007). Polymorphism is a common feature of genes involved in the gene-for-gene relationships between a parasite and its host. Some models have suggested that such polymorphism could be stabilized in parasite populations associated with polycyclic diseases (Tellier and Brown, 2007), i.e. when several parasite generations develop in a host through autoinfection, which is thought to be the case for microsporidia (Cali and Takvorian, 1999). NCER_101600 may well be involved in a gene-for-gene host–parasite interaction.

CONCLUSION

Taken as a phenotypic variation, differences in a parasite virulence would reflect a non-exclusive combination of its (1) genetic polymorphism *sensu stricto*, i.e. its genetically determined virulence strategies adapted to certain environmental conditions – including the host niche – and (2) polyphenism, i.e. the expression of alternative virulence strategies depending on environmental cues (Pfennig, 2001; Schwander and Leimar, 2011). The present work showed that despite a relatively high nucleotide diversity, a similar parasite population is found within geographically distant honeybee isolates. Thus the virulence of *N. ceranae* should rather be related to an alteration of the host–parasite interactions in response to environmental cues, including environmental stressors (Bromenshenk *et al.* 2010; Gisder *et al.* 2010) and the honeybee genetic background (Fontbonne *et al.* 2013), than to its polymorphism.

Lastly, if the present work showed that *N. ceranae* populations are poorly divergent in distant isolates, it questions the origin and maintenance of a strong haplotype diversity within a single honeybee host. Our data support the hypothesis of a recent host jump of *N. ceranae* from the Asian to the European honeybee that would be still followed by a population expansion – and possibly further radiation – in its new host niche, without clear isolation from the *A. cerana*-infecting parasites. Altogether, these findings suggest that *N. ceranae* may constitute a valuable model to monitor the evolution of an emerging parasite.

REFERENCES

Antùnez, K., Martin-Hernandez, R., Prieto, L., Meana, A., Zunino, P. and Higes, M. (2009). Immune suppression in the honey bee (*Apis mellifera*) following infection by *Nosema ceranae* (Microsporidia). *Environmental Microbiology* **11**, 2284–2290. doi: 10.1111/j.1462-2920.2009.01953.x.

Bendtsen, J. D., Nielsen, H., von Heijne, G. and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**, 783–795.

Becnel, J. J. and Andreadis, T. G. (1999). Microsporidia in insects. In *The Microsporidia and Microsporidiosis* (ed. Wittner, M. and Weiss, L. M.), pp. 447–501. ASM Press, Washington, DC, USA.

Birky, C. W., Jr. (1996). Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* **144**, 427–437.

Botias, C., Anderson, D. L., Meana, A., Garrido-Bailon, E., Martin-Hernandez, R. and Higes, M. (2012). Further evidence of an oriental origin for *Nosema ceranae* (Microsporidia: Nosematidae). *Journal of Invertebrate Pathology* **110**, 108–113. doi: 10.1016/j.jip.2012.02.014.

Bromenshenk, J. J., Henderson, C. B., Wick, C. H., Stanford, M. F., Zulich, A. W., Jabbour, R. E., Deshpande, S. V., McCubbin, P. E., Seccomb, R. A., Welch, P. M., Williams, T., Firth, D. R., Skowronski, E., Lehmann, M. M., Bilimoria, S. L., Gress, J., Wanner, K. W. and Cramer, R. A., Jr. (2010). Iridovirus and microsporidian linked to honey bee colony decline. *PLoS ONE* **5**, e13181. doi: 10.1371/journal.pone.0013181.

Cali, A. and Takvorian, P. M. (1999). Developmental morphology and life cycles of the microsporidia. In *The Microsporidia and Microsporidiosis* (ed. Wittner, M. and Weiss, L. M.), pp. 85–128. ASM Press, Washington, DC, USA.

Capella-Gutiérrez, S., Marcet-Houben, M. and Gabaldón, T. (2012). Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. *BMC Biology* **10**, 47. doi: 10.1186/1741-7007-10-47.

Chaimanee, V., Warrit, N. and Chantawannakul, P. (2010). Infections of *Nosema ceranae* in four different honeybee species. *Journal of Invertebrate Pathology* **105**, 207–210. doi: 10.1016/j.jip.2010.06.005.

Chaimanee, V., Chen, Y., Pettis, J. S., Scott Cornman, R. and Chantawannakul, P. (2011). Phylogenetic analysis of *Nosema ceranae* isolated from European and Asian honeybees in Northern Thailand. *Journal of Invertebrate Pathology* **107**, 229–233. doi: 10.1016/j.jip.2011.05.012.

Chaimanee, V., Chantawannakul, P., Chen, Y., Evans, J. D. and Pettis, J. S. (2012). Differential expression of immune genes of adult honey bee (*Apis mellifera*) after inoculated by *Nosema ceranae*. *Journal of Insect Physiology* **58**, 1090–1095. doi: 10.1016/j.jinsphys.2012.04.016.

Chen, Y., Evans, J. D., Smith, I. B. and Pettis, J. S. (2008). *Nosema ceranae* is a long-present and wide-spread microsporidian infection of the European honey bee (*Apis mellifera*) in the United States. *Journal of Invertebrate Pathology* **97**, 186–188. doi: 10.1016/j.jip.2007.07.010.

Chen, Y., Evans, J. D., Zhou, L., Boncristiani, H., Kimura, K., Xiao, T., Litkowski, A. M. and Pettis, J. S. (2009a). Asymmetrical coexistence of *Nosema ceranae* and *Nosema apis* in honey bees. *Journal of Invertebrate Pathology* **101**, 204–209. doi: 10.1016/j.jip.2009.05.012.

Chen, Y. P., Evans, J. D., Murphy, C., Gutell, R., Zuker, M., Gundensen-Rindal, D. and Pettis, J. S. (2009b). Morphological, molecular, and phylogenetic characterization of *Nosema ceranae*, a microsporidian parasite isolated from the European honey bee, *Apis mellifera*. *Journal of Eukaryotic Microbiology* **56**, 142–147. doi: 10.1111/j.1550-7408.2008.00374.x.

Cornman, R. S., Chen, Y. P., Schatz, M. C., Street, C., Zhao, Y., Desany, B., Egholm, M., Hutchison, S., Pettis, J. S., Lipkin, W. I. and Evans, J. D. (2009). Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. *PLoS Pathogens* **5**, e1000466. doi: 10.1371/journal.ppat.1000466.

Cox-Foster, D. L., Conlan, S., Holmes, E. C., Palacios, G., Evans, J. D., Moran, N. A., Quan, P. L., Briese, T., Hornig, M., Geiser, D. M.,

Martinson, V., vanEngelsdorp, D., Kalkstein, A. L., Drysdale, A., Hui, J., Zhai, J., Cui, L., Hutchison, S. K., Simons, J. F., Egholm, M., Pettis, J. S. and Lipkin, W. I. (2007). A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* **318**, 283–287. doi: 10.1126/science.1146498.

Crooks, G. E., Hon, G., Chandonia, J. M. and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Research* **14**, 1188–1190. doi: 10.1101/gr.849004.

Didier, E. S., Vossbrinck, C. R., Baker, M. D., Rogers, L. B., Bertucci, D. C. and Shadduck, J. A. (1995). Identification and characterization of three *Encephalitozoon cuniculi* strains. *Parasitology* **111**, 411–421. doi: 10.1017/S0031182000065914.

Dussaubat, C., Brunet, J.-L., Higes, M., Colbourne, J. K., Lopez, J., Choi, J. H., Martín-Hernández, R., Botías, C., Cousin, M., McDonnell, C., Bonnet, M., Belzunces, L. P., Moritz, R. F., Le Conte, Y. and Alaux, C. (2012). Gut pathology and responses to the microsporidium *Nosema ceranae* in the honey bee *Apis mellifera*. *PLoS ONE* **7**, e37017. doi: 10.1371/journal.pone.0037017.

Fay, J. C. and Wu, C. I. (1999). A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Molecular Biology and Evolution* **16**, 1003–1005.

Fontbonne, R., Garnery, L., Vidau, C., Aufauvre, J., Texier, C., Tchamitchian, S., El Alaoui, H., Brunet, J.L., Delbac, F. and Biron, D. G. (2013). Comparative susceptibility of three Western honeybee taxa to the microsporidian parasite *Nosema ceranae*. *Infection Genetics and Evolution* **17**, 188–194.

Fries, I. (2010). *Nosema ceranae* in European honey bees (*Apis mellifera*). *Journal of Invertebrate Pathology* **103**, S73–S79. doi: 10.1016/j.jip.2009.06.017.

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

Gatehouse, H. S. and Malone, L. A. (1998). 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. *Journal of Invertebrate Pathology* **71**, 97–105. doi: 10.1006/jipa.1997.4737.

Gisder, S., Hedtke, K., Mockel, N., Frielitz, M. C., Linde, A. and Genersch, E. (2010). Five-year cohort study of *Nosema* spp. in Germany: does climate shape virulence and assertiveness of *Nosema ceranae*? *Applied and Environmental Microbiology* **76**, 3032–3038. doi: 10.1128/AEM.03097-09.

Haag, K. L., Traunecker, E. and Ebert, D. (2013). Single-nucleotide polymorphisms of two closely related microsporidian parasites suggest a clonal population expansion after the last glaciation. *Molecular Ecology* **22**, 314–326. doi: 10.1111/mec.12126.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95–98.

Haro, M., Del Aguila, C., Fenoy, S. and Henriques-Gil, N. (2003). Intraspecies genotype variability of the microsporidian parasite *Encephalitozoon hellem*. *Journal of Clinical Microbiology* **41**, 4166–4171. doi: 10.1128/JCM.41.9.4166-4171.2003.

Hatjina, F., Tsoktouridis, G., Bouga, M., Charistos, L., Evangelou, V., Avtzis, D., Meeus, I., Brunain, M., Smagghe, G. and de Graaf, D. C. (2011). Polar tube protein gene diversity among *Nosema ceranae* strains derived from a Greek honey bee health study. *Journal of Invertebrate Pathology* **108**, 131–134. doi: 10.1016/j.jip.2011.07.003.

Higes, M., Martín-Hernández, R. and Meana, A. (2010). *Nosema ceranae* in Europe: an emergent type C nosemosis. *Apidologie* **41**, 375–392. doi: 10.1051/apido/2010019.

Higes, M., Martin-Hernandez, R., Botias, C., Bailon, E. G., Gonzalez-Porto, A. V., Barrios, L., Del Nozal, M. J., Bernal, J. L., Jimenez, J. J., Palencia, P. G. and Meana, A. (2008). How natural infection by *Nosema ceranae* causes honeybee colony collapse. *Environmental Microbiology* **10**, 2659–2669. doi: 10.1111/j.1462-2920.2008.01687.x.

Higes, M., Meana, A., Bartolomé, C., Botias, C. and Martin-Hernandez, R. (2013). *Nosema ceranae* (Microsporidia), a controversial 21st century honey bee pathogen. *Environmental Microbiology Reports* **5**, 17–29. doi: 10.1111/1758-2229.12024.

Huang, W. F., Bocquet, M., Lee, K. C., Sung, I. H., Jiang, J. H., Chen, Y. W. and Wang, C. H. (2008). The comparison of rDNA spacer regions of *Nosema ceranae* isolates from different hosts and locations. *Journal of Invertebrate Pathology* **97**, 9–13. doi: 10.1016/j.jip.2007.07.001.

Invernizzi, C., Abud, C., Tomasco, I. H., Harriet, J., Ramallo, G., Campa, J., Katz, H., Gardiol, G. and Mendoza, Y. (2009). Presence of *Nosema ceranae* in honeybees (*Apis mellifera*) in Uruguay. *Journal of Invertebrate Pathology* **101**, 150–153. doi: 10.1016/j.jip.2009.03.006.

Ironside, J. E. (2007). Multiple losses of sex within a single genus of Microsporidia. *BMC Evolutionary Biology* **7**, 48. doi: 10.1186/1471-2148-7-48.

Ironside, J. E. (2013) Diversity and recombination of dispersed ribosomal DNA and protein coding genes in Microsporidia. *PLoS ONE* **8**, e55878. doi: 10.1371/journal.pone.0055878.

Keeling, P. J. and Fast, N. M. (2002). Microsporidia: biology and evolution of highly reduced intracellular parasites. *Annual Review of Microbiology* **56**, 93–116. doi: 10.1146/annurev.micro.56.012302.160854.

Klee, J., Besana, A. M., Genersch, E., Gisder, S., Nanetti, A., Tam, D. Q., Chinh, T. X., Puerta, F., Ruz, J. M., Kryger, P., Message, D., Hatjina, F., Korpela, S., Fries, I. and Paxton, R. J. (2007). Widespread dispersal of the microsporidian *Nosema ceranae*, an emergent pathogen of the western honey bee, *Apis mellifera*. *Journal of Invertebrate Pathology* **96**, 1–10. doi: 10.1016/j.jip.2007.02.014.

Lee, S. C., Corradi, N., Doan, S., Dietrich, F. S., Keeling, P. J. and Heitman, J. (2010*a*). Evolution of the sex-related locus and genomic features shared in microsporidia and fungi. *PLoS ONE* **5**, e10539. doi: 10.1371/journal.pone.0010539.

Lee, S. C., Ni, M., Li, W., Shertz, C. and Heitman, J. (2010*b*). The evolution of sex: a perspective from the fungal kingdom. *Microbiology and Molecular Biology Reviews* **74**, 298–340. doi: 10.1128/MMBR.00005-10.

Li, J., Chen, W., Wu, J., Peng, W., An, J., Schmid-Hempel, P. and Schmid-Hempel, R. (2012). Diversity of *Nosema* associated with bumblebees (*Bombus* spp.) from China. *International Journal for Parasitology* **42**, 49–61. doi: 10.1016/j.ijpara.2011.10.005.

Librado, P. and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452. doi: 10.1093/bioinformatics/btp187.

Martin-Hernandez, R., Meana, A., Prieto, L., Salvador, A. M., Garrido-Bailon, E. and Higes, M. (2007). Outcome of colonization of *Apis mellifera* by *Nosema ceranae*. *Applied and Environmental Microbiology* **73**, 6331–6338. doi: 10.1128/AEM.00270-07.

Martin-Hernandez, R., Botias, C., Barrios, L., Martinez-Salvador, A., Meana, A., Mayack, C. and Higes, M. (2011). Comparison of the energetic stress associated with experimental *Nosema ceranae* and *Nosema apis* infection of honeybees (*Apis mellifera*). *Parasitology Research* **109**, 605–612. doi: 10.1007/s00436-011-2292-9.

Martin-Hernandez, R., Botias, C., Bailon, E. G., Martinez-Salvador, A., Prieto, L., Meana, A. and Higes, M. (2012). Microsporidia infecting *Apis mellifera*: coexistence or competition. Is *Nosema ceranae* replacing *Nosema apis*? *Environmental Microbiology* **14**, 2127–2138. doi: 10.1111/j.1462-2920.2011.02645.x.

Mathis, A., Tanner, I., Weber, R. and Deplazes, P. (1999). Genetic and phenotypic intraspecific variation in the microsporidian *Encephalitozoon hellem*. *International Journal for Parasitology* **29**, 767–770. doi: 10.1016/S0020-7519(99)00025-9.

Mutinelli, F. (2011). The spread of pathogens through trade in honey bees and their products (including queen bees and semen): overview and recent developments. *Revue Scientifique et Technique/Office international des épizooties* **30**, 257–271.

Nei, M. and Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York, NY, USA.

O'Mahony, E. M., Tay, W. T. and Paxton, R. J. (2007). Multiple rRNA variants in a single spore of the microsporidian *Nosema bombi*. *Journal of Eukaryotic Microbiology* **54**, 103–109. doi: 10.1111/j.1550-7408.2006.00232.x.

Paxton, R. J., Klee, J., Korpela, S. and Fries, I. (2007). *Nosema ceranae* has infected *Apis mellifera* in Europe since at least 1998 and may be more virulent than *Nosema apis*. *Apidologie* **38**, 558–565. doi: 10.1051/apido:2007037.

Peuvel, I., Delbac, F., Metenier, G., Peyret, P. and Vivares, C. P. (2000). Polymorphism of the gene encoding a major polar tube protein PTP1 in two microsporidia of the genus *Encephalitozoon*. *Parasitology* **121**, 581–587. doi: 10.1017/S0031182000006910.

Peuvel-Fanget, I., Polonais, V., Brosson, D., Texier, C., Kuhn, L., Peyret, P., Vivarès, C. and Delbac, F. (2006). EnP1 and EnP2, two proteins associated with the *Encephalitozoon cuniculi* endospore, the chitin-rich inner layer of the microsporidian spore wall. *International Journal for Parasitology* **36**, 309–318. doi: 10.1016/j.ijpara.2005.10.005.

Pfennig, K. S. (2001). Evolution of pathogen virulence: the role of variation in host phenotype. *Proceedings of the Royal Society of London. Series B* **268**, 755–760. doi: 10.1098/rspb.2000.1582.

Plischuk, S., Martín-Hernández, R., Prieto, L., Lucía, M., Botías, C., Meana, A., Abrahamovich, A. H., Lange, C. and Higes, M. (2009). South American native bumblebees (Hymenoptera: Apidae) infected by *Nosema ceranae* (Microsporidia), an emerging pathogen of honeybees

(*Apis mellifera*). *Environmental Microbiology Reports* **1**, 131–135. doi: 10.1111/j.1758-2229.2009.00018.x.

Polonais, V., Mazet, M., Wawrzyniak, I., Texier, C., Blot, N., El Alaoui, H. and Delbac, F. (2010). The human microsporidian *Encephalitozoon hellem* synthesizes two spore wall polymorphic proteins useful for epidemiological studies. *Infection and Immunity* **78**, 2221–2230. doi: 10.1128/IAI.01225-09.

Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O. and Kunin, W. E. (2010). Global pollinator declines: trends, impacts and drivers. *Trends in Ecology and Evolution* **25**, 345–353. doi: 10.1016/j.tree.2010.01.007.

Ramirez-Soriano, A., Ramos-Onsins, S. E., Rozas, J., Calafell, F. and Navarro, A. (2008). Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* **179**, 555–567. doi: 10.1534/genetics.107.083006.

Ramos-Onsins, S. E. and Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**, 2092–2100.

Rozas, J. (2009). DNA sequence polymorphism analysis using DnaSP. *Methods in Molecular Biology* **537**, 337–350. doi: 10.1007/978-1-59745-251-9_17.

Rozas, J., Gullaud, M., Blandin, G. and Aguade, M. (2001). DNA variation at the *rp49* gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. *Genetics* **158**, 1147–1155.

Sagastume, S., Del Aguila, C., Martin-Hernandez, R., Higes, M. and Henriques-Gil, N. (2011). Polymorphism and recombination for rDNA in the putatively asexual microsporidian *Nosema ceranae*, a pathogen of honeybees. *Environmental Microbiology* **13**, 84–95. doi: 10.1111/j.1462-2920.2010.02311.x.

Schwander, T. and Leimar, O. (2011). Genes as leaders and followers in evolution. *Trends in Ecology and Evolution* **26**, 143–151. doi: 10.1016/j.tree.2010.12.010.

Southern, T. R., Jolly, C. E., Lester, M. E. and Hayman, J. R. (2007). EnP1, a microsporidian spore wall protein that enables spores to adhere to and infect host cells *in vitro*. *Eukaryotic Cell* **6**, 1354–1362. doi: 10.1128/EC.00113-07.

Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics* **123**, 597–601.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731–2739. doi: 10.1093/molbev/msr121.

Tay, W. T., O'Mahony, E. M. and Paxton, R. J. (2005). Complete rRNA gene sequences reveal that the microsporidium *Nosema bombi* infects diverse bumblebee (*Bombus* spp.) hosts and contains multiple polymorphic sites. *Journal of Eukaryotic Microbiology* **52**, 505–513. doi: 10.1111/j.1550-7408.2005.00057.x.

Tellier, A. and Brown, J. K. (2007). Polymorphism in multilocus host parasite coevolutionary interactions. *Genetics* **177**, 1777–1790. doi: 10.1534/genetics.107.074393.

ten Hove, R. J., Van Lieshout, L., Beadsworth, M. B., Perez, M. A., Spee, K., Claas, E. C. and Verweij, J. J. (2009). Characterization of genotypes of *Enterocytozoon bieneusi* in immunosuppressed and immunocompetent patient groups. *Journal of Eukaryotic Microbiology* **56**, 388–393. doi: 10.1111/j.1550-7408.2009.00393.x.

Visvesvara, G. S., da Silva, A. J., Croppo, G. P., Pieniazek, N. J., Leitch, G. J., Ferguson, D., de Moura, H., Wallace, S., Slemenda, S. B. and Tyrrell, I. (1995). *In vitro* culture and serologic and molecular identification of *Septata intestinalis* isolated from urine of a patient with AIDS. *Journal of Clinical Microbiology* **33**, 930–936.

Vizoso, D. B. and Ebert, D. (2005). Mixed inoculations of a microsporidian parasite with horizontal and vertical infections. *Oecologia* **143**, 157–166. doi: 10.1007/s00442-004-1771-4.

Wu, Z., Li, Y., Pan, G., Tan, X., Hu, J., Zhou, Z. and Xiang, Z. (2008). Proteomic analysis of spore wall proteins and identification of two spore wall proteins from *Nosema bombycis* (Microsporidia). *Proteomics* **8**, 2447–2461. doi: 10.1002/pmic.200700584.

Wu, Z., Li, Y., Pan, G., Zhou, Z. and Xiang, Z. (2009). SWP25, a novel protein associated with the *Nosema bombycis* endospore. *Journal of Eukaryotic Microbiology* **56**, 113–118. doi: 10.1111/j.1550-7408.2008.00375.x.

Zander, E. (1909). Tierische Parasiten als Krankenheitserreger bei der Biene. *Leipziger Bienenzeitung* **24**, 147–150.

Zheng, D., Alm, E. W., Stahl, D. A. and Raskin, L. (1996). Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. *Applied and Environmental Microbiology* **62**, 4504–4513.