GENES AND GENOMES

Bee Trypanosomatids: First Steps in the Analysis of the Genetic Variation and Population Structure of *Lotmaria passim*, *Crithidia bombi* and *Crithidia mellificae*

Carolina Bartolomé[1,2] · María Buendía-Abad[3] · Concepción Ornosa[4] · Pilar De la Rúa[5] · Raquel Martín-Hernández[3,6] · Mariano Higes[3] · Xulio Maside[1,2]

Abstract

Trypanosomatids are among the most prevalent parasites in bees but, despite the fact that their impact on the colonies can be quite important and that their infectivity may potentially depend on their genotypes, little is known about the population diversity of these pathogens. Here we cloned and sequenced three non-repetitive single copy *loci* (*DNA topoisomerase II*, *glyceraldehyde-3-phosphate dehydrogenase* and *RNA polymerase II large subunit*, *RPB1*) to produce new genetic data from *Crithidia bombi*, *C. mellificae* and *Lotmaria passim* isolated from honeybees and bumblebees. These were analysed by applying population genetic tools in order to quantify and compare their variability within and between species, and to obtain information on their demography and population structure. The general pattern for the three species was that (1) they were subject to the action of purifying selection on nonsynonymous variants, (2) the levels of within species diversity were similar irrespective of the host, (3) there was evidence of recombination among haplotypes and (4) they showed no haplotype structuring according to the host. *C. bombi* exhibited the lowest levels of synonymous variation ($\pi_S = 0.06 \pm 0.04$ %) — and a mutation frequency distribution compatible with a population expansion after a bottleneck — that contrasted with the extensive polymorphism displayed by *C. mellificae* ($\pi_S = 2.24 \pm 1.00$ %), which likely has a more ancient origin. *L. passim* showed intermediate values ($\pi_S = 0.40 \pm 0.28$ %) and an excess of variants a low frequencies probably linked to the spread of this species to new geographical areas.

Keywords Population genetics · Genetic diversity · Population structure · *Lotmaria passim* · *Crithidia mellificae* · *Crithidia bombi*

✉ Carolina Bartolomé
  carolina.bartolome@usc.es

1 Grupo de Medicina Xenómica, CIMUS, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Galicia, Spain

2 Instituto de Investigación Sanitaria de Santiago (IDIS), 15706 Santiago de Compostela, Galicia, Spain

3 Instituto Regional de Investigación y Desarrollo Agroalimentario y Forestal (IRIAF), Laboratorio de Patología Apícola, Centro de Investigación Apícola y Agroambiental (CIAPA), Consejería de Agricultura de la Junta de Comunidades de Castilla-La Mancha, 19180 Marchamalo, Spain

4 Departamento de Biodiversidad, Ecología y Evolución, Facultad de Ciencias Biológicas, Universidad Complutense de Madrid, 28040 Madrid, Spain

5 Departamento de Zoología y Antropología Física, Facultad de Veterinaria, Universidad de Murcia, 30100 Murcia, Spain

6 Instituto de Recursos Humanos para la Ciencia y la Tecnología, Fundación Parque Científico Tecnológico de Albacete, 02006 Albacete, Spain

Introduction

Trypanosomatids (Protozoa: Trypanosomatida: Trypanosomatidae) are unicellular flagellates that parasitise a wide variety of organisms [1, 2]. According to their life cycle, they can be divided into monoxenous (with a single host) and dixenous (with two hosts). The vast majority of monoxenous trypanosomatids, which are the predominant type, are restricted to insects, especially Diptera and Heteroptera but also Siphonaptera, Blattodea, Mecoptera, Lepidoptera and Hymenoptera [3, 4].

Despite the reduced number of trypanosomatid species reported in bees (all of them included within the subfamily Leishmaniinae) — *Crithidia mellificae* Langridge and McGhee, 1967, *Crithidia bombi* Lipa and Triggiani, 1988, *Crithidia expoeki* Schmid-Hempel and Tognazzo, 2010, *Lotmaria passim* Evans and Schwarz, 2014, and a few others [5, 6] — their prevalence [6–9] and impact on the host [10–13] can be quite important.

C. mellificae and *L. passim* infect predominantly honeybees, although both species have been found in other Hymenoptera [8, 14–17]; the same holds for *C. bombi* which, regardless of being a common parasite of bumblebees (as is *C. expoeki*), has been occasionally detected in *Apis mellifera* Linnaeus, 1758 [6, 14, 18]. The identification of these pathogens in species sharing the same environment highlights the importance of spillover events in their circulation among arthropod communities (reviewed by [19]). Considering that parasites may adapt to infect new hosts (with the concomitant risk that this represents for the ecosystems), and that the success of such infections has been associated with variation among genotypes [20], it is surprising how little is known about the diversity and the population genetics of bee trypanosomatids.

So far, most studies addressing these matters have been performed in *C. bombi* using microsatellite markers [21, 22]. These exhibit repetitive sequences that usually present a much higher mutation rate than the rest of the genome, which may interfere not only with the assessment of the levels of diversity of a species but also with that of other population genetics estimates [23, 24]. Another approach used to analyse the variability of bee trypanosomatids is the sequencing of other nuclear *loci*, either individually [8, 25, 26] or at the genomic level [27]. Although both approximations provided interesting insights into the amount of genetic variation and/or gene copy number, only a few of these studies used population genetic statistics to quantify and compare these parameters within and between species. The use of these tools is essential to interpret molecular data and to determine the genetic bases for the adaptation to the environment, such as revealing the influence of different evolutionary forces on the amount of variation detected or uncover the existence of genetic differentiation among parasite populations from different geographical regions or hosts. This in turn allows not only to evaluate potential changes in the structure and distribution of a species (that can help to develop better conservation strategies) but also to infer their demographic history, providing extremely valuable information for applied research in many different fields (ecology, environmental sciences, wildlife conservation or biogeography, among others) [28–30].

Materials and Methods

Samples and DNA Extraction

Seven *A. mellifera* worker bees and five *Bombus terrestris* (Linnaeus, 1758) field samples from different areas of Spain (Supplemental Fig. 1), identified as trypanosomatid-positive by PCR [31], were selected for this study. Direct sequencing of their amplicons revealed the presence of *L. passim* in five of them (PA11-831, PA11-847, PA11-853, ITS2 PR13-21 and ITS3 PR13-21), *C. mellificae* in two (PA14-0015 and PA14-0044) and *C. bombi* in all *Bombus* specimens (B14.213, 14_349, 14_351, 14_373 and 14_395). To increase the size of the *C. mellificae* dataset, we also included the reference strain ATCC 30254. This was first grown in ATCC medium 355 and later sub-cultivated in Brain Heart Infusion (BHI) medium. Individual colonies were re-suspended in milliQ water (PCR-quality) and processed for DNA extraction.

DNA extractions from *A. mellifera* and the ATCC 30254 strain were carried out as described in Cepero et al. [25], using the BS96 DNA Tissue extraction protocol on a Bio-Sprint workstation (Qiagen), whereas those from *B. terrestris* were performed using a modification of the Chelex protocol [32].

Primer Design

The amino acid sequences of three *loci* (*DNA topoisomerase II*, *TOPII*; *glyceraldehyde-3-phosphate dehydrogenase*, *GAPDH*; and *RNA polymerase II large subunit*, *RPB1*) were obtained from the draft genome of *L. passim* SF. These were used to conduct TBLASTN searches for orthologs in other genomes of the Trypanosomatidae family. Their outputs were aligned using Bioedit [33] and used to design broad-range primers with primer BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The selected oligonucleotides were located within the most conserved regions of the markers and their sequences were as follows: Tryp-1F (CCGAGT ACTTCKCSTACCAG) and Tryp-1R (AGCCGAGGATGC CCTTCAT) [25] for *GAPDH*, TrypRPB1-F (AGGCGGAGC

TGATYGAGATG) and TrypRPB1-R (ACCGAGAAGGCR AAGCARTAG) for *RPB1* and TrypTOP-F (CACAAGCGC ATYATGGACCT) and TrypTOP-R (TTRCTCTGCGAG TCGAACTT) for *TOPII*, respectively.

In several samples, these primers amplified simultaneously more than one trypanosomatid species (see 'Results' section), so, in order to amplify them separately, new sets of species-specific primers were designed (Supplemental Table 1).

PCR, Cloning and Sequencing

PCR reactions were performed using the Phusion High-Fidelity DNA Polymerase (Thermo Scientific) in 20-µl volumes containing 9.4 µl of H_2O, 4 µl of 5X Phusion HF Buffer, 0.4 µl of dNTP mix 10 mM, 2 µl of each primer 5 µM, 0.2 µl of Phusion DNA Polymerase and 2 µl of DNA. Cycling conditions consisted of an initial denaturalization at 98 °C for 30 s, followed by 35 cycles of denaturalization at 98 °C for 10 s, primer annealing at 62°C (for *TOPII* and *GAPDH*) or 65 °C (for *RPB1*) for 30 s (the Phusion DNA polymerase requires higher annealing temperatures than Taq-based polymerases) and an extension of 72 °C for 10 s, and a final extension of 8 min at 72 °C.

When the annealing temperature was ≥ 69 °C (as in the case of some of the species-specific primers shown in Supplemental Table 1), a two-step PCR protocol — with no annealing step — was used.

The resulting amplicons were gel-purified (NZYGelpure, NZYTech, Portugal) and cloned using CloneJET PCR Cloning Kit, Thermo Scientific). Plasmid DNA of ten clones from each positive sample was purified (NZYMiniprep, NZYTech, Portugal) and sequenced on an ABI 3730XL sequencing machine (GATC, Eurofins Genomics, Germany).

Population Genetics Analyses

Based on the expression $P = 1 - (1 - p)^\wedge n$, where P is the probability of finding a variant at frequency p (in this case, $p = 0.25$) in a sample of size n, it was estimated that the sequencing of ten amplicons provided about a 95% likelihood of identifying any variant segregating at a frequency ≥ 25% in each isolate, which ensured enough power for the analysis of genetic variation.

It should be noted that although the ATCC 30254 strain was originally isolated from *A. mellifera*, for subsequent analyses, it was considered apart on the grounds that it came from an axenic culture. In the case of *GADPH,* two additional GenBank datasets encompassing the same region were included for comparison with ours (PopSets 1169070972 and 663527929).

The nucleotide diversity in each species was estimated at synonymous and nonsynonymous sites by means of the π [average number of nucleotide differences per site between two sequences; 34] and θ_W [number of segregating sites; 35] statistics, which were calculated applying the Jukes and Cantor correction [36]. Both measures are complementary, since π is sensitive to the frequency of polymorphisms while θ_W is not. These parameters as well as the Tajima's D statistic [37] — a measure of the mutation frequency spectrum — were calculated with DnaSP v6 [38]. D values are negative when there is an excess of rare variants and positive when there is an excess of high-frequency mutations. The pooled value of D across *loci* was estimated manually using Tajima's formulae [37]. The divergence (or genetic differentiation) between species was quantified as the mean number of synonymous (K_s) and nonsynonymous (K_a) substitutions per site, using the Jukes and Cantor correction, as implemented in DnaSP v6. Alignment gaps were excluded from all the analyses.

DnaSP v6 was also used to carry out the McDonald-Kreitman test [39], which was applied to each *locus* to assess if the levels of variability within and between species fitted the predictions of the neutral theory. Under this model, both should be correlated over time [40], and therefore the ratio of nonsynonymous to synonymous variation between species (*Dn/Ds*) should be equal to the ratio of nonsynonymous to synonymous variation within a species (*Pn/Ps*).

Median-joining haplotype networks [41] were constructed using PopART [42] (http://popart.otago.ac.nz/index.shtml) departing from nexus haplotype files generated with DnaSP v6. This type of networks includes nodes to represent inferred sequences by iteratively adding 'median' sequence vectors.

The analysis of molecular variance (AMOVA) among haplotypes was performed with Arlequin 3.5 [43]. The significance of the covariance components associated with different levels of genetic structure (within isolates, among isolates within host species and among host species) was tested using non-parametric permutation procedures [44].

The existence of recombination was initially tested with RecMin (http://www.stats.ox.ac.uk/~myers/RecMin.html), which provides two minimum numbers of recombination events: the statistic *Rm* of Hudson and Kaplan [45] based on the four-gamete test, and *Rh* [46], which reflects the number of recombination events required to explain the history of a sample under the assumption that each segregating site has mutated only once. However, recombination was only confirmed after performing a permutation-based analysis with LDhat [47], which uses the composite-likelihood method of Hudson [48], adapted to finite-sites models [47]. The maximum composite likelihood was calculated under random permutation of the physical position of the variants (1000 permutations) to test the hypothesis of no-recombination.

Results

The protocol of PCR cloning and sequencing the amplicons obtained with broad-range primers revealed the presence of multiple trypanosomatid species in samples where *a priori* only one had been detected by direct sequencing (Table 1).

L. passim was the most prevalent trypanosomatid of the dataset, as it was identified in the seven field isolates of *A. mell*ifera and in three out of the five *B. terrestris*. In half of these samples, this pathogen co-occurred with, at least, a second species (Table 1).

In bumblebees, the most prevalent trypanosomatid species was *C. bombi*, which was detected in the five isolates studied.

Population Genetics Analyses

Diversity Within Species

The first observation derived from these analyses was that the *loci* used in this study exhibited some heterogeneity in the levels of synonymous nucleotide variation (π_S), *RPB1* being considerably more polymorphic than *GAPDH* or *TOPII* in *L. passim* and *C. mellificae* (Table 2 and Supplemental Tables 2, 3 and 4). This *locus* also showed slightly higher diversity in *C. bombi*, although it was not statistically different from that of *GAPDH*.

Overall, *C. mellificae* was far more diverse at synonymous positions (2.24 ± 1.00%) than *L. passim* (0.40 ± 0.28 %) or *C. bombi*, which showed extremely low levels of diversity (0.06 ± 0.04 %) (mean π_S ± SE estimated using data from Supplemental Tables 2, 3 and 4). These values remained nearly unchanged after weighting each *locus* by its number of synonymous sites, to avoid any bias caused

Table 2 Mean synonymous (π_S) and nonsynonymous (π_A) pairwise diversity per *locus* and species (± standard error, SE)

	Species	Mean π_S ± SE (%)	Mean π_A ± SE (%)
GAPDH-A	*C. bombi*	0.07 ± 0.03	0.03 ± 0.02
	C. mellificae	0.93 ± 0.17	0.07 ± 0.05
	L. passim	0.06 ± 0.03	0.06 ± 0.02
GAPDH-B	*C. bombi*	0.00 ± 0.00	0.00 ± 0.00
	C. mellificae	1.40 ± 0.29	0.30 ± 0.04
	L. passim	0.38 ± 0.06	0.76 ± 0.42
RPB1	*C. bombi*	0.13 ± 0.06	0.12 ± 0.03
	C. mellificae	4.21 ± 0.39	0.03 ± 0.02
	L. passim	0.96 ± 0.12	0.17 ± 0.02
TOPII	*C. bombi*	0.00 ± 0.00	0.05 ± 0.02
	C. mellificae	1.58 ± 0.41	0.06 ± 0.03
	L. passim	0.17 ± 0.07	0.16 ± 0.03

GAPDH-A: data from this work (ATCC 30254 excluded)

GAPDH-B: *C. bombi* (Popset 1169070972), *C. mellificae* (Popset 663527929 + ATCC 30254 from this work) and *L. passim* (Popset 663527929)

by the different length of the genes: weighted mean π_S ± SE = 2.02 ± 0.91% for *C. mellificae*, 0.35 ± 0.30% for *L. passim* and 0.07 ± 0.04% for *C. bombi*.

The ATCC 30254 strain displayed a little more diversity across loci than the *C. mellificae* field samples, both using data from this study (weighted mean π_S ± SE = 3.01 ± 1.15%, estimated using data from Supplemental Tables 2, 3 and 4) and when considering sequences from other *C. mellificae* type isolates available in GenBank (*GAPDH*-B in Table 2). In the case of *L. passim*, the difference between the two groups was highly significant (0.38 ± 0.06 in cultured cell lines vs. 0.06 ± 0.03 in wild populations), whereas the field samples from the *C. bombi* Popset 1169070972 (*GAPDH*-B in Table 2) exhibited no variability.

Along with its higher rate of synonymous variation, *C. mellificae* displayed a greater tendency to present

Table 1 Trypanosomatid species detected in *A. mellifera* and *B. terrestris* samples

Host	Isolate	Direct sequencing	Cloning sequencing
A. mellifera	ITS2 PR13-21	*L. passim*	*L. passim*
	ITS3 PR13-21	*L. passim*	*L. passim*
	PA11-831	*L. passim*	*L. passim*
	PA11-847	*L. passim*	*L. passim*
	PA11-853	*L. passim*	*L. passim*
	PA14-0015	*C. mellificae*	*L. passim*; *C. bombi*; *C. mellificae*
	PA14-0044	*C. mellificae*	*L. passim*; *C. bombi*; *C. mellificae*
B. terrestris	B14.213	*C. bombi*	*L. passim*; *C. bombi*
	14_349	*C. bombi*	*C. bombi*
	14_351	*C. bombi*	*L. passim*; *C. bombi*; *C. mellificae*
	14_373	*C. bombi*	*C. bombi*
	14_395	*C. bombi*	*L. passim*; *C. bombi*; *C. mellificae*

intermediate frequency mutations at the three *loci* (D_S pooled across *loci*= 1.67; data from Supplemental Tables 2, 3 and 4) than the other two species, which exhibited slight excesses of rare variants (D_S pooled across *loci*= −1.59 and −1.52 for *C. bombi* and *L. passim*, respectively; data from Supplemental Tables 2, 3 and 4). At any rate, D_S deviations only reached significant values at *GAPDH* in *C. mellificae* and *L. passim* (Supplemental Table 2).

At nonsynonymous sites, *C. mellificae* showed an important reduction in variability with respect to that at synonymous positions, both when the three *loci* were considered individually (Table 2) and when they were pooled together (weighted mean $\pi_A \pm SE = 0.05 \pm 0.08\%$ vs. weighted mean $\pi_S \pm SE = 2.02 \pm 0.91\%$), in line with the values observed in the ATCC 30254 strain (weighted mean $\pi_A \pm SE = 0.11 \pm 0.40\%$ vs. weighted mean $\pi_S \pm SE = 3.01 \pm 1.15\%$).

In contrast, *L. passim* and *C. bombi* displayed similar rates of polymorphism at both sites (weighted mean $\pi_A \pm SE = 0.12 \pm 0.08\%$ vs. weighted mean $\pi_{S=} 0.35 \pm 0.30\%$ for *L. passim*, and weighted mean $\pi_A \pm SE = 0.06 \pm 0.09\%$ vs. weighted mean $\pi_S \pm SE = 0.07 \pm 0.04\%$ for *C. bombi*, respectively).

The *GAPDH*-B datasets of *C. mellificae* and *L. passim* followed the same trends as those described above for each of these species, but with much higher rates of nonsynonymous diversity than those obtained in the natural populations analysed here (Table 2). Yet again, the *C. bombi* Popset 1169070972 showed no variability (*GAPDH*-B, Table 2).

In most cases, the frequency spectrum of replacement substitutions at individual genes did not deviate from neutral expectations (D_A values from Supplemental Tables 2, 3 and 4). However, when the mean D_A was calculated across *loci*, it displayed a significant excess of rare variants for the three species (D_A pooled across *loci* = −3.21, $P < 0.001$, for *C. bombi*; −2.62, $P < 0.01$, for *L. passim*; and −1.97, $P < 0.05$, for *C. mellificae*).

In terms of diversity (both at synonymous and nonsynonymous positions), there were no significant differences among trypanosomatid species isolated from different hosts, although this point will be addressed in detail in the 'Population Structure' section.

Divergence Between Species

The analysis of divergence between species revealed important differences among *loci*. *GAPDH* was the *locus* with least synonymous differentiation among species (K_s), whereas *TOPII* exhibited saturation of synonymous mutations ($K_s > 1$) each time that *C. bombi* was involved in the comparisons (Table 3), which derived in larger mean $K_s \pm SE$ values across *loci* (0.85 ± 0.39 between *C. bombi* and *C. mellificae* or 0.79 ± 0.35 between *C. bombi* and *L. passim*) than those

Table 3 Pairwise nucleotide divergence, at synonymous (K_s) and nonsynonymous sites (K_a)

Locus	C. bombi vs. C. mellificae		
	Ks	Ka	Ka/Ks
GAPDH	0.24	0.01	0.04
RPB1	0.72	0.01	0.01
TOPO	1.59	0.06	0.04
Locus	C. bombi vs. L. passim		
	Ks	Ka	Ka/Ks
GAPDH	0.23	0.02	0.09
RPB1	0.71	0.01	0.01
TOPO	1.44	0.07	0.05
Locus	L. passim vs. C. mellificae		
	Ks	Ka	Ka/Ks
GAPDH	0.28	0.01	0.04
RPB1	0.54	0.00	0.01
TOPO	0.67	0.10	0.15

obtained from the comparison between *C. mellificae* and *L. passim* (0.50 ± 0.12).

Again, *TOPII* was the gene accumulating most replacement variants between species (K_a) although the average $K_a \pm SE$ values across *loci* were much more similar among them than those of K_s ($K_a \pm SE = 0.04 \pm 0.03$ between *C. mellificae* and *L. passim*, 0.03 ± 0.02 between *C. mellificae* and *C. bombi* or 0.03 ± 0.02 between *C. mellificae* and *L. passim*).

In all cases, the K_a/K_s ratios, which allow inferring the existence and direction of selective forces acting on a sequence, were lower than one, which is consistent with a predominant purifying selection (Table 3).

To further investigate the heterogeneity observed among genes, both in terms of polymorphism and divergence, we performed McDonald-Kreitman tests [39] to assess if the ratios of nonsynonymous and synonymous mutations within and between species fitted the predictions of the neutral model [40].

However, contrary to the expectations under neutrality (in which *Pn/Ps =Dn/Ds*; see 'Materials and Methods' section), the vast majority of the comparisons resulted in significantly higher *Pn/Ps* than *Dn/Ds* ratios (0.88 > 0.14 for pooled data between *C. bombi* and *C. mellificae*, 2.38 > 0.19 between *C. bombi* and. *L. passim* and 0.81 > 0.22 between *C. mellificae* and *L. passim*, respectively), indicating an excess of amino acid polymorphisms (*Pn*) as compared with fixations (*Dn*; Table 4).

When variants with frequencies below 5% were removed from the analyses, in order to increase the power to detect adaptive evolution [49], the differences between these ratios turned non-significant.

Table 4 McDonald-Kreitman tests

Cb vs. *Cm*	*Locus*	Fixed		Polymorphic		*P*
		Syn	Non-syn	Syn	Non-syn	
	GAPDH	33	5	6	7	**
	RPB1	43	2	15	11	***
	TOPII	63	12	5	5	*
	Pooled	139	19	26	23	***
Cb vs. *Lp*	*Locus*	Fixed		Polymorphic		*P*
		Syn	Non-syn	Syn	Non-syn	
	GAPDH	33	9	8	14	**
	RPB1	47	2	4	15	***
	TOPII	60	15	4	9	***
	Pooled	140	26	16	38	***
Cm vs. *Lp*	*Locus*	Fixed		Polymorphic		*P*
		Syn	Non-syn	Syn	Non-syn	
	GAPDH	37	6	8	11	**
	RPB1	33	0	15	7	***
	TOPII	43	19	9	8	ns
	Pooled	113	25	32	26	***

Cb, *C. bombi*; *Cm*, *C. mellificae*; *Lp*, *L. passim*; *Fixed*, fixed mutations between species; *Ds and Dn*, number of synonymous and nonsynonymous fixed variants, respectively; *Polymorphic*, polymorphic changes within species; *Ps and Pn*, number of synonymous and nonsynonymous polymorphic variants, respectively; *P*, significance of the deviation from the neutral model calculated using the two-tailed Fisher's exact test; *ns*, non-significant; *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$

Population Structure

As previously mentioned, isolates obtained from honey-bees and bumblebees showed non-significant differences in nucleotide diversity measured as π [34].

An alternative way to investigate if there was any kind of genetic structure among the haplotypes detected in these hosts was to estimate the significance and relative contribution of each of the covariance components to the total haplotypic variance (see AMOVA analysis in the 'Materials and Methods' section). This analysis showed that most of the variation resided within isolates (range 82.2–100%, Table 5), and this was valid for all species and markers except for *GAPDH* in *C. mellificae*, where the major contributor to the total variance was the variation among isolates of the same host ($P < 0.001$ in permutation tests). This component was the second contributor to the variance in *L. passim*, accounting for 14.1% of the total variance (average across the three *loci*). None of the three parasite species exhibited significant haplotype differentiation among hosts.

The graphical representation of the haplotypes revealed that they were organised in one or more (usually two) core haplotypes, present in both hosts at intermediate frequencies, from which derived low-frequency haplotypes that, in most cases, differed from the former by a single mutation (Fig. 1 and Supplemental Figs. 2 and 3). This became even more obvious when the sequences retrieved from GenBank were included in the networks, since their haplotypes were mostly represented by single sequences (data not shown).

Recombination

L. passim and *C. mellificae* displayed reticulated haplotype networks (Fig. 1 and Supplemental Figs. 2 and 3), which suggest the existence of recombination events in these species. Consistent with these patterns, *L. passim* showed evidences of potential recombination at *RPB1* (*Rm*= 2; *Rh*= 3), whereas *C. mellificae* exhibited them at all *loci* (*Rm*= 1 and *Rh*= 2 at *GAPDH*, *Rm*= 2 and *Rh*= 5 at *RPB1* and *Rm*= 1and *Rh*= 3 at *TOPII*). The statistical significance of these potential recombination events was evaluated with permutation tests (LDhat; [47]) and it proved significant ($P < 0.01$) for *RPB1* and *TOPII* in *C. mellificae*.

Discussion

There is increasing evidence of the simultaneous presence of multiple trypanosomatids in bees [6, 14–16]. One of the reasons why these co-occurrences have not been reported until recently might be due to the methodology employed for their detection, which usually relies on the PCR amplification of DNA isolates followed by the direct sequencing of the PCR products [5, 9, 10, 50, 51]. However, as seen in this study, direct sequencing might overlook

Table 5 Analysis of molecular variance (AMOVA)

Species	Locus	Source of variation	d.f.	SS	VC	% var	P
C. bombi	GAPDH	Among host species	1	0.1	0.0	−1.4	
		Among isolates within host species	5	0.7	0.0	0.3	
		Within isolates	63	8.1	0.1	101.1	
		Total	69	8.8	0.1		
	RPB1	Among host species	1	0.7	0.0	4.5	
		Among isolates within host species	5	1.5	0.0	1.9	
		Within isolates	63	16.1	0.3	93.6	*
		Total	69	18.3	0.3		
	TOPII	Among host species	1	0.2	0.0	3.8	
		Among isolates within host species	5	0.4	0.0	1.9	
		Within isolates	63	4.3	0.1	94.3	
		Total	69	4.9	0.1		
C. mellificae†	GAPDH	Among host species	1	2.5	−0.4	−45.1	
		Among isolates within host species	2	22.1	1.1	113.2	***
		Within isolates	36	10.9	0.3	32.0	***
		Total	39	35.4	0.9		
	RPB1	Among host species	1	4.2	0.1	5.5	
		Among isolates within host species	2	3.7	0.0	−0.8	
		Within isolates	36	72.4	2.0	95.3	
		Total	39	80.3	2.1		
	TOPII	Among host species	1	0.9	0.0	0.0	
		Among isolates within host species	2	1.7	0.0	1.2	
		Within isolates	36	27.4	0.8	98.9	
		Total	39	30.0	0.8		
L. passim	GAPDH	Among host species	1	0.3	0.0	−1.6	
		Among isolates within host species	8	3.2	0.0	11.3	***
		Within isolates	90	16.0	0.2	90.2	***
		Total	99	19.5	0.2		
	RPB1	Among host species	1	0.9	0.0	−4.4	
		Among isolates within host species	8	19.3	0.2	22.2	***
		Within isolates	90	58.7	0.7	82.2	***
		Total	99	78.9	0.8		
	TOPII	Among host species	1	0.6	0.0	0.1	
		Among isolates within host species	8	4.4	0.0	8.9	**
		Within isolates	90	24.9	0.3	91.0	**
		Total	99	29.8	0.3		

†ATCC 30254 excluded from the dataset

Negative values should be considered as zero, and values ≥100 as 100

d.f., degrees of freedom; *SS*, sum of squares; *VC*, variance components; *% var*, percentage of variation; *p*, probability of a random variance component value ≤ observed value in 3024 permutations; *$p < 0.05$, **$p < 0.01$ and ***$p < 0.001$

the presence of species with lower parasitic loads. This drawback can be solved by adding a cloning step prior to sequencing that although more expensive and time consuming is one of the best options for carrying out reliable population genetics analyses, not only because it allows the detection of additional species but also because it enables to uncover the presence of low-frequency variants within a species [8, 15, 25]. As an example, we can mention the Popset 1169070972 that we intended to use for comparison with our *C. bombi* sequences and that, despite of being a fairly large dataset, showed no variation likely due to direct sequencing (V. Vavilova personal communication).

Thus, in the current study, we cloned and sequenced 10 amplicons of *GAPDH*, one of the genes most frequently used for the molecular identification of trypanosomatids [8, 15, 25, 52], and of two other single-locus nuclear markers (*RPB1* and *TOPII*) for which there were no previous population data.

Fig. 1 Median-joining haplotype network for *GAPDH*. Haplotypes are depicted by circles, the width being proportional to their frequencies. Black, grey and white circles/sections represent haplotypes obtained from *B. terrestris*, *A. mellifera* and the ATCC 30254 strain (this work), respectively. **a** *Crithidia bombi* haplotypes; **b** *Crithidia mellificae* haplotypes; **c** *Lotmaria passim* haplotypes. Mutations are shown as hatch marks along edges and inferred haplotypes ('median' sequence vectors) are represented by stars
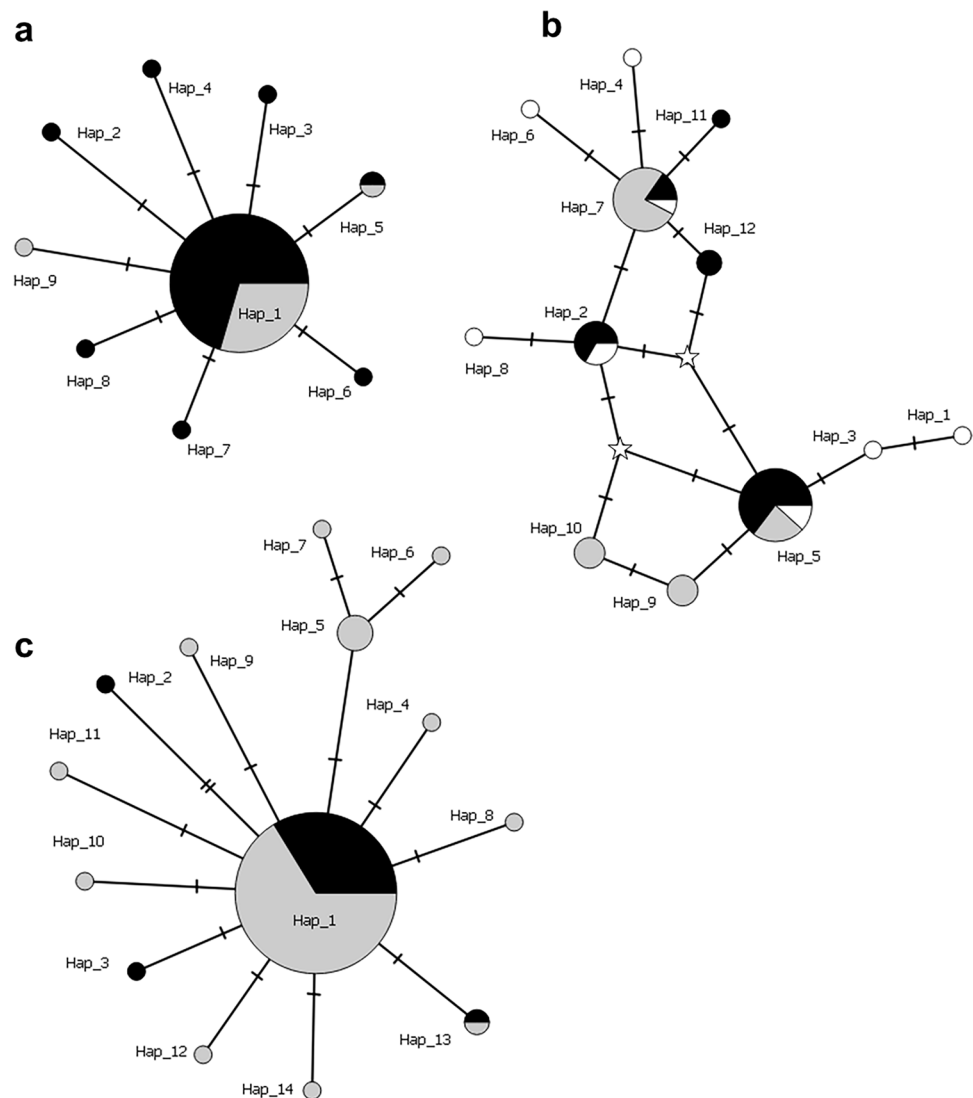
One of the findings resulting from their analysis, and that contrasts with previous reports based on microsatellite data [21, 53], was the low variability detected in *C. bombi*, both at synonymous and non-synonymous sites. The discrepancy between both types of variation data probably relies on the higher mutation rate shown by microsatellites with respect to single-copy coding regions of the genome [54]. Besides, microsatellites are often subject to homoplasy (identical sequences and/or amplicon sizes resulting from independent evolutionary paths) that, together with other factors [reviewed by 24], may confound the population genetic inference. Although the analysis of a few genes does not necessarily reflect the general patterns of variation of a species, the consistency of the trends detected across the three *loci* seems to suggest that *C. bombi* is little polymorphic, at least in these isolates. The reduced nucleotide variation together with a negative Tajima's D_S shown by *C. bombi* is consistent with a recent population expansion [55] that could be related with the radiation of bumblebees from Central

Asia [56]. In agreement with this, Central European samples of *C. bombi* were found to be less diverse than those from Alaska, which are closer to their basal populations [27]. Besides, Gerasimov et al. [27] found lower variability in the mitochondrial DNA of *C. bombi* than in its nuclear counterpart, which is largely compatible with a bottleneck scenario: the population undergoes a drastic decrease in size that results in a reduction in the levels of variation that, at first, is more evident in the mitochondrial DNA because its smaller effective population size makes it more responsive to demographic changes than the nuclear one [57].

To the best of our knowledge, the only other population genetic study of genetic variation in *L. passim* is that of Cepero et al. [25], who reported the first genetic description of this trypanosomatid prior to its formal naming and characterization by Schwarz et al. [8]. The re-analysis of the section of Cepero's sequences that overlapped with ours (KJ704252.1-KJ704272.1; positions 50 to 499) revealed significantly higher diversity than that obtained in the current

study both at synonymous ($\pi_S \pm$ SE= 0.32 $\pm$ 0.12 vs. 0.06 $\pm$ 0.03) and nonsynonymous sites ($\pi_A \pm$ SE= 0.50 $\pm$ 0.30 vs. 0.06 $\pm$ 0.02), in line with the results obtained for the *L. passim* sequences obtained from axenic cultures (Popset 663527929; $\pi_S \pm$ SE= 0.38 $\pm$ 0.06 and $\pi_A \pm$ SE= 0.76 $\pm$ 0.42), which suggest large differences in the levels of polymorphism across samples (especially at nonsynonymous positions, as proved by the large SE values). Both groups of field isolates (Cepero's and those obtained in this work) exhibited significantly negative Tajima's *D* at synonymous sites ($-$ 1.87 and $-$ 1.91, respectively; $P < 0.05$). The significance of this statistic at *GAPDH*, together with the excess of rare neutral variants observed in our dataset of *L. passim* (D_S pooled across *loci*= -1.52), seems to indicate that the natural populations of this species have been subject to a recent demographic expansion which could hypothetically be associated with its spread to new geographical areas.

On the other hand, *C. mellificae* proved to be considerably more diverse at synonymous positions than *C. bombi* or *L. passim*, which suggests that it may have a larger population size than any of these two species. It also exhibited a greater tendency to present intermediate frequency mutations than the former ones (represented by positive Tajima's *D* values). This can indicate either the existence of balancing selection, population structuring or a sudden population contraction; the first possibility was discarded on the basis that balancing selection usually affects very specific targets, whereas in this case, the three *loci* showed a similar pattern; the results of the analysis of molecular variance also permitted to disregard the existence of population subdivision (most of the diversity resided within isolates), so the most probable explanation for the presence of this slight excess of intermediate frequency alleles is the existence of a reduction in population size (following an incomplete bottleneck, low-frequency mutations are lost more rapidly than common ones, deriving in transient positive Tajima's *D* values) [55, 57].

Independently of the demographic trends shown by each of these trypanosomatids, the most relevant pattern for the three species is that they are subject to the action of purifying selection on nonsynonymous variants, which is a process that removes deleterious mutations (usually causing amino acid changes) from the populations. Its effect was detected even in *C. bombi*, where the efficiency of purifying selection was likely to be lower due to its recent bottleneck and concomitant reduction in effective population size [58].

The existence of such selective constraints on the three species comes from different evidences: (i) the frequency spectrum of replacement changes, represented by significantly negative D_A values, indicated that a large part of these mutations were segregating at low frequencies [37, 54] and (ii) *Ka/Ks* ratios lower than one suggested that mutations modifying a protein were less likely to get fixed between two species than synonymous substitutions [59].

Both findings were clearly reflected in the results of the McDonald-Kreitman test, in which the removal of polymorphisms with a frequency below 5% confirmed that the excess of replacement polymorphisms (*Pn*) observed prior to their exclusion was mostly due to the presence of slightly deleterious mutations, which usually segregate at low frequencies and rarely become fixed in the populations [49, 60]. This surplus of rare mutations could also be inferred from the haplotype networks, which in most cases exhibited one or two intermediate-frequency haplotypes connected to other closely related haplotypes that often represented singleton mutations, whose abundance is also dependent on the recombination rate [61]. However, it should be taken into account that pooling potentially different allopatric samples into single specific groups may interfere with the analysis of this parameter through the risk of a Wahlund effect [62].

The apparent absence of recombination events observed in this study for *C. bombi* would lead to a reduction in the effectiveness of selection [63] that would contradict the results of the MacDonald-Kreitman test; this suggests that, although not detected here, there must be a certain genetic exchange that helps to maintain the integrity of the genome. Consistent with this, *C. bombi*, as well as other trypanosomatids, showed evidence of Mendelian segregation in hybridization experiments [64, 65], a finding that was further supported by the presence of meiosis-specific genes in their genomes [66–68]. In this study, we also found signs of recombination in *L. passim* and *C. mellificae*, especially in the latter, where the evidence for this process was highly significant and in good agreement with the high levels of synonymous diversity and the efficiency of purifying selection ($\pi_A << \pi_S$) observed in this species.

Another characteristic common to the three species is that most of the variation resided within isolates, which means that the haplotypes detected for the different trypanosomatids showed no structuring according to the host, suggesting that direct and indirect interactions between pollinators enable the dispersal of the most common haplotypes of each species, which are present both in honeybees and bumblebees (as depicted in the haplotype networks from the previous section). Also, the fact that parasites show similar levels of diversity in both hosts contradicts the assumption that those present in their main host should have a larger population size than those carried by incidental vectors, which, again, seems to indicate that most of the parasite variability is effectively spread and maintained into the environment by the former. Finally, it should be stressed that although some population structuring was detected for *C. mellificae* and *L. passim* within hosts, the limited sample size prevents to draw any firm conclusion about the possible existence of genetic subdivisions within these species.

Altogether, this study represents a first approximation to the population genetics of these highly prevalent pathogens,

which could serve as a basis to carry out further research involving more samples and genes.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. Maslov DA, Opperdoes FR, Kostygov AY, Hashimi H, Lukes J, Yurchenko V (2019) Recent advances in trypanosomatid research: genome organization, expression, metabolism, taxonomy and evolution. Parasitology 146:1–27. https://doi.org/10.1017/S0031182018000951

2. Kaufer A, Ellis J, Stark D, Barratt J (2017) The evolution of trypanosomatid taxonomy. Parasit Vectors 10:287. https://doi.org/10.1186/s13071-017-2204-7

3. Maslov DA, Votýpka J, Yurchenko V, Lukeš J (2013) Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. Trends Parasitol 29:43–52. https://doi.org/10.1016/j.pt.2012.11.001

4. Lukes J, Butenko A, Hashimi H, Maslov DA, Votypka J, Yurchenko V (2018) Trypanosomatids are much more than just trypanosomes: clues from the expanded family tree. Trends Parasitol 34:466–480. https://doi.org/10.1016/j.pt.2018.03.002

5. Gallot-Lavallée M, Schmid-Hempel R, Vandame R, Vergara CH, Schmid-Hempel P (2016) Large scale patterns of abundance and distribution of parasites in Mexican bumblebees. J Invertebr Pathol 133:73–82. https://doi.org/10.1016/j.jip.2015.12.004

6. Bartolomé C, Buendía-Abad M, Benito M, Sobrino B, Amigo J, Carracedo A, Martín-Hernández R, Higes M, Maside X (2020) Longitudinal analysis on parasite diversity in honeybee colonies: new taxa, high frequency of mixed infections and seasonal patterns of variation. Sci Rep 10:10454. https://doi.org/10.1038/s41598-020-67183-3

7. Tritschler M, Retschnig G, Yanez O, Williams GR, Neumann P (2017) Host sharing by the honey bee parasites *Lotmaria passim* and *Nosema ceranae*. Ecol Evol 7:1850–1857. https://doi.org/10.1002/ece3.2796

8. Schwarz RS, Bauchan GR, Murphy CA, Ravoet J, de Graaf DC, Evans JD (2015) Characterization of two species of Trypanosomatidae from the honey bee *Apis mellifera*: *Crithidia mellificae* Langridge and McGhee, and *Lotmaria passim* n. gen., n. sp. J Eukaryot Microbiol 62:567–583. https://doi.org/10.1111/jeu.12209

9. Plischuk S, Antúnez K, Haramboure M, Minardi GM, Lange CE (2017) Long-term prevalence of the protists *Crithidia bombi* and *Apicystis bombi* and detection of the microsporidium *Nosema bombi* in invasive bumble bees. Environ Microbiol Rep 9:169–173. https://doi.org/10.1111/1758-2229.12520

10. Ravoet J, Maharramov J, Meeus I, De Smet L, Wenseleers T, Smagghe G, de Graaf DC (2013) Comprehensive bee pathogen screening in Belgium reveals *Crithidia mellificae* as a new contributory factor to winter mortality. PLoS One 8:e72443. https://doi.org/10.1371/journal.pone.0072443

11. Brown MJF, Schmid-Hempel R, Schmid-Hempel P (2003) Strong context-dependent virulence in a host–parasite system: reconciling genetic evidence with theory. J Anim Ecol 72:994–1002. https://doi.org/10.1046/j.1365-2656.2003.00770.x

12. Gegear RJ, Otterstatter MC, Thomson JD (2006) Bumble-bee foragers infected by a gut parasite have an impaired ability to utilize floral information. Proc Biol Sci 273:1073–1078. https://doi.org/10.1098/rspb.2005.3423

13. Gómez-Moracho T, Buendía-Abad M, Benito M, García-Palencia P, Barrios L, Bartolomé C, Maside X, Meana A, Dolores Jiménez-Antón M, Isabel Olías-Molero A, María Alunda J, Martín-Hernández R, Higes M (2020) Experimental evidence of harmful effects of Crithidia mellificae and Lotmaria passim on honey bees. Int J Parasitol. https://doi.org/10.1016/j.ijpara.2020.06.009

14. Bartolomé C, Buendía M, Benito M, De la Rúa P, Ornosa C, Martín-Hernández R, Higes M, Maside X (2018) A new multiplex PCR protocol to detect mixed trypanosomatid infections in species of *Apis* and *Bombus*. J Invertebr Pathol 154:37–41. https://doi.org/10.1016/j.jip.2018.03.015

15. Ravoet J, Schwarz RS, Descamps T, Yañez O, Tozkar CO, Martín-Hernández R, Bartolomé C, De Smet L, Higes M, Wenseleers T, Schmid-Hempel R, Neumann P, Kadowaki T, Evans JD, de Graaf DC (2015) Differential diagnosis of the honey bee trypanosomatids *Crithidia mellificae* and *Lotmaria passim*. J Invertebr Pathol 130:21–27. https://doi.org/10.1016/j.jip.2015.06.007

16. Tripodi AD, Szalanski AL, Strange JP (2018) Novel multiplex PCR reveals multiple trypanosomatid species infecting North American bumble bees (Hymenoptera: Apidae: *Bombus*). J Invertebr Pathol 153:147–155. https://doi.org/10.1016/j.jip.2018.03.009

17. Gabín-García LB, Bartolomé C, Guerra-Tort C, Rojas-Nossa SV, Llovo J, Maside X (2021) Identification of pathogens in the invasive hornet *Vespa velutina* and in native Hymenoptera (Apidae, Vespidae) from SW-Europe. Sci Rep 11:11233. https://doi.org/10.1038/s41598-021-90615-7

18. Graystock P, Goulson D, Hughes WO (2015) Parasites in bloom: flowers aid dispersal and transmission of pollinator parasites within and between bee species. Proc Biol Sci 282:20151371. https://doi.org/10.1098/rspb.2015.1371

19. Nanetti A, Bortolotti L, Cilia G (2021) Pathogens spillover from honey bees to other arthropods. Pathogens 10. https://doi.org/10.3390/pathogens10081044

20. Schmid-Hempel P, Puhr K, Krüger N, Reber C, Schmid-Hempel R (1999) Dynamic and genetic consequences of variation in horizontal transmission for a microparasitic infection. Evolution 53:426–434. https://doi.org/10.1111/j.1558-5646.1999.tb03778.x

21. Schmid-Hempel P, Reber Funk C (2004) The distribution of genotypes of the trypanosome parasite, *Crithidia bombi*, in populations of its host, *Bombus terrestris*. Parasitology 129:147–158. https://doi.org/10.1017/s0031182004005542

22. Popp M, Erler S, Lattorff HMG (2012) Seasonal variability of prevalence and occurrence of multiple infections shape the population structure of *Crithidia bombi*, an intestinal parasite of bumblebees (*Bombus* spp.). MicrobiologyOpen 1:362–372. https://doi.org/10.1002/mbo3.35

23. Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol 11:1591–1604. https://doi.org/10.1046/j.1365-294X.2002.01576.x

24. Putman AI, Carbone I (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecol Evol 4:4399–4428. https://doi.org/10.1002/ece3.1305

25. Cepero A, Ravoet J, Gómez-Moracho T, Bernal JL, Del Nozal MJ, Bartolomé C, Maside X, Meana A, González-Porto AV, de Graaf DC, Martín-Hernández R, Higes M (2014) Holistic screening of collapsing honey bee colonies in Spain: a case study. BMC Res Notes 7:649. https://doi.org/10.1186/1756-0500-7-649

26. Schmid-Hempel R, Tognazzo M (2010) Molecular divergence defines two distinct lineages of *Crithidia bombi* (Trypanosomatidae), parasites of bumblebees. J Eukaryot Microbiol 57:337–345. https://doi.org/10.1111/j.1550-7408.2010.00480.x

27. Gerasimov E, Zemp N, Schmid-Hempel R, Schmid-Hempel P, Yurchenko V (2019) Genomic variation among strains of *Crithidia bombi* and *C. expoeki*. mSphere 4:e00482–e00419. https://doi.org/10.1128/mSphere.00482-19

28. Hedrick PW (2001) Conservation genetics: where are we now? Trends Ecol Evol 16:629–636. https://doi.org/10.1016/S0169-5347(01)02282-0

29. Lowe WH, Kovach RP, Allendorf FW (2017) Population genetics and demography unite ecology and evolution. Trends Ecol Evol 32:141–152. https://doi.org/10.1016/j.tree.2016.12.002

30. Habel JC, Zachos FE, Dapporto L, Rödder D, Radespiel U, Tellier A, Schmitt T (2015) Population genetics revisited – towards a multidisciplinary research field. Biol J Linn Soc 115:1–12. https://doi.org/10.1111/bij.12481

31. Meeus I, de Graaf DC, Jans K, Smagghe G (2010) Multiplex PCR detection of slowly-evolving trypanosomatids and neogregarines in bumblebees using broad-range primers. J Appl Microbiol 109:107–115. https://doi.org/10.1111/j.1365-2672.2009.04635.x

32. Scriven JJ, Woodall LC, Goulson D (2013) Nondestructive DNA sampling from bumblebee faeces. Mol Ecol Resour 13:225–229. https://doi.org/10.1111/1755-0998.12036

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

34. Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

35. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276. https://doi.org/10.1016/0040-5809(75)90020-9

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

37. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595. https://doi.org/10.1093/genetics/123.3.585

38. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302. https://doi.org/10.1093/molbev/msx248

39. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351:652–654. https://doi.org/10.1038/351652a0

40. Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

41. Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

42. Leigh JW, Bryant D (2015) PopART: full-feature software for haplotype network construction. Methods Ecol Evol 6:1110–1116. https://doi.org/10.1111/2041-210x.12410

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

44. Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491. https://doi.org/10.1093/genetics/131.2.479

45. Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164. https://doi.org/10.1093/genetics/111.1.147

46. Myers SR, Griffiths RC (2003) Bounds on the minimum number of recombination events in a sample history. Genetics 163:375–394. https://doi.org/10.1093/genetics/163.1.375

47. McVean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 160:1231–1241. https://doi.org/10.1093/genetics/160.3.1231

48. Hudson RR (2001) Two-locus sampling distributions and their application. Genetics 159:1805–1817. https://doi.org/10.1093/genetics/159.4.1805

49. Charlesworth J, Eyre-Walker A (2008) The McDonald-Kreitman test and slightly deleterious mutations. Mol Biol Evol 25:1007–1015. https://doi.org/10.1093/molbev/msn005

50. Yang B, Peng G, Li T, Kadowaki T (2013) Molecular and phylogenetic characterization of honey bee viruses, *Nosema* microsporidia, protozoan parasites, and parasitic mites in China. Ecol Evol 3:298–311. https://doi.org/10.1002/ece3.464

51. Li J, Chen W, Wu J, Peng W, An J, Schmid-Hempel P, Schmid-Hempel R (2012) Diversity of *Nosema* associated with bumblebees (*Bombus* spp.) from China. Int J Parasitol 42:49–61. https://doi.org/10.1016/j.ijpara.2011.10.005

52. Rangel DA, Lisboa CV, Novaes RLM, Silva BA, Souza RF, Jansen AM, Moratelli R, Roque ALR (2019) Isolation and characterization of trypanosomatids, including *Crithidia mellificae*, in bats from the Atlantic forest of Rio de Janeiro, Brazil. PLoS Neglect Trop D 13:e0007527. https://doi.org/10.1371/journal.pntd.0007527

53. Erler S, Popp M, Wolf S, Lattorff HMG (2012) Sex, horizontal transmission, and multiple hosts prevent local adaptation of *Crithidia bombi*, a parasite of bumblebees (*Bombus* spp.). Ecol Evol 2:930–940. https://doi.org/10.1002/ece3.250

54. Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, Holderegger R, Widmer A (2017) Estimating genomic diversity and population differentiation - an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. BMC Genomics 18:69–69. https://doi.org/10.1186/s12864-016-3459-7

55. Tajima F (1989) The effect of change in population size on DNA polymorphism. Genetics 123:597–601. https://doi.org/10.1093/genetics/123.3.597

56. Hines HM (2008) Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: *Bombus*). Syst Biol 57:58–75. https://doi.org/10.1080/10635150801898912

57. Fay JC, Wu CI (1999) A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol Biol Evol 16:1003–1005. https://doi.org/10.1093/oxfordjournals.molbev.a026175

58. Ohta T (1973) Slightly deleterious mutant substitutions in evolution. Nature 246:96–98. https://doi.org/10.1038/246096a0

59. Hurst LD (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet 18:486. https://doi.org/10.1016/s0168-9525(02)02722-1

60. Egea R, Casillas S, Barbadilla A (2008) Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic Acids Res 36:W157–W162. https://doi.org/10.1093/nar/gkn337

61. Campos JL, Halligan DL, Haddrill PR, Charlesworth B (2014) The relation between recombination rate and patterns of molecular evolution and variation in *Drosophila melanogaster*. Mol Biol Evol 31:1010–1028. https://doi.org/10.1093/molbev/msu056

62. Prugnolle F, De Meeus T (2010) Apparent high recombination rates in clonal parasitic organisms due to inappropriate sampling design. Heredity 104:135–140. https://doi.org/10.1038/hdy.2009.128

63. Betancourt AJ, Welch JJ, Charlesworth B (2009) Reduced effectiveness of selection caused by a lack of recombination. Curr Biol 19:655–660. https://doi.org/10.1016/j.cub.2009.02.039

64. Schmid-Hempel R, Salathé R, Tognazzo M, Schmid-Hempel P (2011) Genetic exchange and emergence of novel strains in directly transmitted trypanosomatids. Infect Genet Evol 11:564–571. https://doi.org/10.1016/j.meegid.2011.01.002

65. MacLeod A, Tweedie A, McLellan S, Taylor S, Cooper A, Sweeney L, Turner CMR, Tait A (2005) Allelic segregation and independent assortment in T. brucei crosses: proof that the genetic system is mendelian and involves meiosis. Mol Biochem Parasitol 143:12–19. https://doi.org/10.1016/j.molbiopara.2005.04.009

66. Kraeva N, Butenko A, Hlaváčová J, Kostygov AY, Myšková J, Grybchuk D, Leštinová T, Votýpka J, Volf P, Opperdoes FR, Flegontov P, Lukeš J, Yurchenko V (2015) *Leptomonas seymouri:* adaptations to the dixenous life cycle analyzed by genome sequencing, transcriptome profiling and co-infection with *Leishmania donovani*. PLoS Pathog 11: e1005127. 10.1371/journal.ppat.1005127

67. Flegontov P, Butenko A, Firsov S, Kraeva N, Eliáš M, Field MC, Filatov D, Flegontova O, Gerasimov ES, Hlaváčová J, Ishemgulova A, Jackson AP, Kelly S, Kostygov AY, Logacheva MD, Maslov DA, Opperdoes FR, O'Reilly A, Sádlová J et al (2016) Genome of *Leptomonas pyrrhocoris:* a high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. Sci Rep 6:23704. https://doi.org/10.1038/srep23704

68. Schmid-Hempel P, Aebi M, Barribeau S, Kitajima T, du Plessis L, Schmid-Hempel R, Zoller S (2018) The genomes of *Crithidia bombi* and *C. expoeki,* common parasites of bumblebees. PLoS One 13:e0189738. https://doi.org/10.1371/journal.pone.0189738