

RESEARCH

Open Access



# What complete mitochondrial genomes tell us about the evolutionary history of the black soldier fly, *Hermetia illucens*

J. Guilliet<sup>1,2,3\*</sup>, G. Baudouin<sup>2</sup>, N. Pollet<sup>1,3</sup> and J. Filée<sup>1,3\*</sup>

## Abstract

**Background:** The Black Soldier Fly (BSF) *Hermetia illucens* is a cosmopolitan fly massively used by industrial companies to reduce biowaste and produce protein and fat for poultry and aquaculture feed. However, the natural history and the genetic diversity of the BSF are poorly known. Here, we present a comprehensive phylogeny and time tree based on a large dataset of complete mitochondrial genomes better to understand the evolution and timing of the BSF.

**Results:** In this study, we analyzed 677 CO1 sequences derived from samples found all over the five continents, leading us to discover 52 haplotypes, including ten major haplotypes. This worldwide cryptic genetic and genomic diversity is mirrored at a local scale in France, in which we found five major haplotypes sometimes in sympatry. Phylogenetic analyses of 60 complete mitochondrial genomes robustly resolved the phylogeny of the major BSF haplotypes. We estimate the separation events of the different haplotypes at more than 2 million years for the oldest branches characterizing the ancestral split between present North American lineages and the other highly diverse south-central American clades, possibly the following radiation beyond the isthmus of Panama northwards. Our data confirm that this North American lineage ultimately gave birth to almost all commercial BSF stocks that participated in the worldwide BSF dissemination through farm escapements.

**Conclusions:** Our data resolve the phylogenetic relationships between the major lineages and give insights into the BSF's short and long-term evolution. Our results indicate that commercial BSF stock's genetic and genomic diversity is very low. These results call for a better understanding of the genomic diversity of the BSF to unravel possible specific adaptations of the different lineages for industrial needs and to initiate the selection process.

**Keywords:** *Hermetia illucens*, Phylogeny, Mitochondrial genome, Phylogeography

## Background

The black soldier fly (BSF) *Hermetia illucens* (Linnaeus, 1758) is a Diptera of the Stratiomyidae family. The native area of the BSF is considered to be Central America and

the northern part of South America, then spread to the entire continent over the last thousands of years [1, 2]. These geographical areas are now considered their natural area [1, 3]. The BSF is now a cosmopolitan fly found in tropical, subtropical, and tempered regions, from the 40th parallel north to the 45th parallel south [3, 4] where it is considered an exotic, non-invasive species, according to the current regulations [5–7]. This fly likely spread through shipping routes [8]. While studies based on naturalistic observations show that it was first discovered in Malta in 1926 [9], it has been suggested that the earliest

\*Correspondence: joseph.guilliet@universite-paris-saclay.fr; jonathan.filee@universite-paris-saclay.fr

<sup>3</sup> IDEEV - Laboratoire Evolution, Génomes Comportement, Ecologie CNRS Université Paris Sud UMR 9191, IRD UMR 247, 12 rue 128, 91198 Gif sur Yvette, France

Full list of author information is available at the end of the article



record in Europe was in Italy about 500 years ago [10]. The BSF appeared in the early 1950s in the South-East of France [11] before spreading through the South-West [12] along the Rhône valley [13] and finally along the Atlantic coast [14]. Currently, the BSF can be found on the whole European territory, with few exceptions. It is reasonable to assume that the rise of international traffic and the industrial use of the BSF may have led to a steady flow of introductions in France and worldwide [15, 16].

The BSF was initially used in forensic entomology to date the post-mortem time [17]. Industrial companies heavily use it to reduce biowastes [18, 19], to produce protein [19–21], and fat [19, 20, 22]. These products are used to make animal feed [23] or biodiesel [24, 25].

BSF industries face similar issues as other sectors: knowing how to manage their species and genetic aspects to get the best out of it. To this end, it is crucial to know the genetic diversity of BSF existing internationally, nationally, and locally. Many articles exist about the nutritional aspect of the BSF [22, 26–29] or its microbiota [30–32], but few have focused on the genetic diversity of the BSF. We have some evidence of BSF genetic diversity from previous studies. Ten haplotypes based on the CO1 gene were found in South Korea with 245 individuals [33]. Still, on the CO1 gene and a worldwide sampling, 56 haplotypes were found with a divergence rate up to 4.9%, and it has been shown that individuals belonging to divergent lineages are still interfertile [34]. Finally, based on 15 microsatellite markers, 16 genetic clusters were found with hot spots in South America [2]. Other works focused on more fundamental aspects of BSF and assembled a reference mitochondrial genome and a chromosomal-scale nuclear genome [35–37].

Within the framework of the BSF industry, phylogeography can allow on the one side to account for genetic diversity at different scales (global, national, and local) and, on the other side, to have keys to the geographical origin of the species present in a territory or used by companies. This study may suggest local adaptations or some phenotypic plasticity due to habituation to different biomes.

Our study quantified genetic diversity at different scales: at the global level by studying the variation within the CO1 gene and more locally at the French level. Our report is based on data obtained by sequencing 114 new individuals for the CO1 gene focusing specifically on France (90 French individuals were analyzed) and extracting 563 CO1 sequences from databases. We confirm the hidden genetic diversity within *Hermetia illucens* at all scales of observation, allowing us to find nine different haplotypes. We sequenced and assembled 56 mitochondrial genomes from four different continents and 40 different locations. In addition, we collected sequencing data from databases allowing us to assemble three additional mitochondrial genomes. We also collected a mitochondrial genome already assembled from databases, resulting in 60 mitochondrial genomes. These data allowed us to resolve the phylogenetic relationships between the major haplotypes robustly. In addition, we also provide here the date of the separation events of the different haplotypes found. This separation turns out to be more than 2 million years old for the most distant haplotypes.

## Materials and methods

### Sampling

We decided to conduct a global sampling followed by a specific focus on France. A detailed table of samples and associated metadata is provided in (Additional file 1: Table S1). The origin of all CO1 sequences used in our study is summarised in Table 1, and a summary of sample numbers by continent is provided in Table 2.

Regarding the samples, one adult individual was found in Mexico in the wild in 2016. Individuals coming from Asia (Taiwan), Africa (Ghana and Kenya), and South America (French Guiana) were collected during the spring and summer of 2019. They come either from the wild or breeding farms (see Additional file 1). In France, the collection of BSF was done with the help of the citizen compost network [38] during the summer of 2020. We recovered mainly L5 larvae exclusively from shared and personal composts and some nearby adults. Between

**Table 1** Origin of CO1 sequences and mitochondrial genomes

Origin	Number of individuals (CO1 sequences)	Number of individuals (mitochondrial genome)
Collection from the wild	91	43
Commercial origin	23	13
Mined from databases (GenBank and BOLD)	560	1
Assembled from SRA (raw sequences)	3	3
Total	677	60

**Table 2** Geographic origin of CO1 sequences and mitochondrial genomes

Geographic origin	Number of individuals (CO1 sequences)	Number of individuals (mitochondrial genome)
South America	55	4
North America	43	2
Asia	277	4
Oceania	106	0
Africa	78	11
Europe (including France)	118 (90)	39 (38)
Total	677	60

one and ten individuals were collected per area. In some French localities, collections in the wild have been made near BSF breeding areas (about 5 km). We obtained larvae in the areas of Blois, Bordeaux, Toulouse, Montpellier, Lyon, Paris. All these samples were stored in 100% ethanol at  $-20^{\circ}\text{C}$  prior to DNA extraction.

In the NCBI Sequence Read Archive (SRA), we gathered all the CO1 sequences on NCBI and Bold and whole-genome sequencing data and reconstructed the mitochondrial genome. We were surprised in the Bayonne area to have BSF-like larvae, which turned out to be *Exaireta spinigera* larvae. These larvae were analyzed in the same way as the BSF.

#### DNA extraction

We extracted DNA from larvae, pupae, and imago under the same conditions by grinding individuals in liquid nitrogen using a mortar and pestle. The ground samples were then processed according to the protocol of the Macherey–Nagel AXG 100 [39] kit for the isolation of genomic DNA from tissue with some modification: we added a second purification step with 3.5 ml of freshly prepared ethanol and centrifugation at 15,000 rpm at  $4^{\circ}\text{C}$  for 15 min just before the final re-suspension. The DNA pellet was finally re-suspended in 100  $\mu\text{l}$  of Tris 10 mM EDTA 1 mM pH 8.0. The purity was evaluated by spectrophotometry using a Nanodrop [40], and the concentration was quantified using the Qubit kit DNA Broad range [41]. DNA integrity was monitored using routine agarose gel electrophoresis.

#### CO1 sequencing

We designed specific CO1 primers from the reference mitochondrial DNA [35] (Additional file 2: Table S2), then a mix was realized using NEB  $5\times$  standard buffer One Taq. The mix is then amplified in ONETAQ HOT 45 PCR. We purified the PCR products using the Illustra

Exostar 1-step GE Healthcare kit before Sanger cycle sequencing. Cycle sequencing reactions were performed using the BigDye 3.1 chemistry, purified by precipitation, and sequenced using an ABI 3130 sequencer (Applied Biosystems). Base-calling was performed using the sequencing analysis software version 5.3 (Applied Biosystems). We assembled forward and reverse reads using the Geneious V R11 [42] assembler with the sensitivity set at medium/fast. The regions containing the primers were excised, and we kept only the sequences having more than 75% of high-quality base calls.

#### Multiple sequence alignment and haplotype network reconstruction

We aligned CO1 sequences obtained in the laboratory and those recovered from the databases using the MAFFT [43] aligner included as a plug-in in the Geneious software [42]. We used the algorithm G-INS-i with a scoring matrix of 200 PAM/k=2, a gap open penalty at 1.53, and an offset value at 0.123. We exported the multiple sequence alignment in nexus format [44] for haplotype network reconstruction and visualization using POPART [45]. We modified the output file to add geographical origins to the sequences.

#### Shotgun sequencing and mitochondrial DNA assembly

A total of 56 BSF individuals were used for whole-genome shotgun sequencing using an Illumina method by Novogene UK [46–50]. The technique used is pair-end with a medium coverage rate of  $4.68\times$  (SD 0.50). The resulting sequences were checked for quality and counted for read parity. We take advantage of the fact that mtDNA is highly repeated in insect DNA. In insects, mtDNA represents on average 0.42% of the total DNA in the genome sequence project [51]. As stipulated, we used 5000 kb of data/sample ( $4\text{--}5\times$  coverage) for the mitochondrial assembly, which on average contains  $(5000 \times 0.42)/100 = 21,000$  kb of mtDNA. As the mitochondrial genome is about 15 kb, we obtain coverage of  $1400\times$  for the mitochondrial genome, which is mainly sufficient to obtain a complete genome in one contig. The sequences were assembled de novo with the MEGAHIT software [52] (based options); the mitochondrial genomes were thus recovered from the assemblies by blast [53] using the reference BSF mitochondrial genome sequence (NC\_035232). The mitochondrial DNAs were annotated using Mitos online software [54], and the D-Loop zone was removed. We added four complete mitochondrial sequences from databases or reconstructed them via reading sequences available on the SRA

[55] using the same assembly procedure and then aligned them under the same conditions.

### CO1 and whole mitochondrial genome phylogeny

We use the MEGAX [56] software to determine the best model and compute an ML Tree with 500 Bootstraps using a complete deletion parameter. We visualized the data with the online tool ITOL [57], and we rooted the tree with the Stratiomyidae *Exaireta spinigera*. We chose *Exaireta spinigera* because it was the closest species for which we have a complete mitochondrial genome (the blastn result of the complete mitochondrial genome of *Hermetia illucens* vs. *Exaireta spinigera* gives us 81.28% of similarity). To simplify the interpretation of the phylogeny, we have kept only a single identical sequence per haplotype.

### Time tree

We used the BEAST software [58] to estimate the divergence time of the different BSF groups. The analysis was performed with an expansion growth model and an uncorrelated lognormal relaxed clock with the proposed

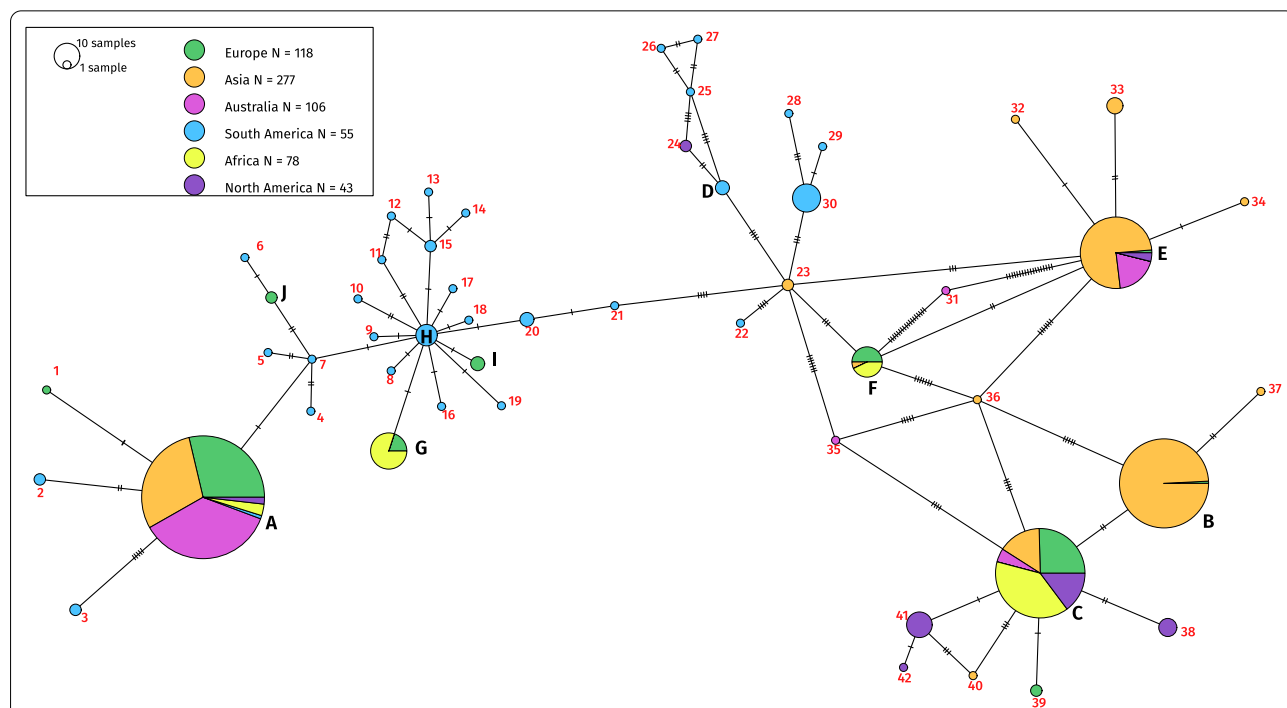
insect molecular clock. Ten million iterations were performed, and then with the TreeAnnotator [59], the obtained trees were clustered with a burn-in of 25% of the trees. The visualization of the final tree was done with the online software ITOL [57].

## Results

### Haplotype network

We reconstructed a haplotype network from the analysis of 677 sequences spanning 658 nucleotides of the CO1 mitochondrial gene. Using the minimum spanning method [60], 52 distinct haplotypes were delineated from the analysis of this CO1 fragment. These different haplotypes group up to 230 sequences (Fig. 1).

There are 31 haplotypes containing only one sequence and 11 haplotypes with between 2 and 4 sequences from the same continent. We have named these 42 haplotypes “minor haplotypes.” We observed ten haplotypes between five and 230 sequences, including six multi-continental haplotypes that we will refer to as major haplotypes. We found an important diversity both within and between continents using the CO1 phylogenetic mitochondrial gene. Minor haplotypes represent 8.5% of all sequences



**Fig. 1** CO1 haplotype network of *Hermetia illucens*. The haplotype network generated by PopArt after multiple alignments are based on the CO1 gene of *Hermetia illucens*. Analysis of the differences between the CO1 gene sequences of 677 individuals (we did not separate commercial and wild individuals). The pie charts represent the same haplotype, and their relative size indicates the number of individuals sharing this haplotype. The number of bars between each segment indicates the differences between the two sequences. Within a pie chart, the colours represent the geographical origin of the individuals. The major haplotypes include more than five sequences (A, B, C, E, F, G, H, 30, 38 & 41). We grouped individuals by continent, separating North America from Central and South America. The list of sequences and their exact origin is available as an Additional file 1

(58 sequences) and 80.8% of haplotypes. Major haplotypes represent 91.5% of the sequences (619 sequences) and 19.2% of the haplotypes.

In Europe, we found ten distinct haplotypes from 118 sequences, nine for Asia from 277 sequences, five for Oceania from 106 sequences, 30 for South America from 55 sequences, four in Africa from 78 sequences, and seven in North America from 43 sequences. Some major haplotypes are represented in a more restricted area. For example, haplotype B is mainly in Asia (Fig. 1). We observed a high haplotype diversity in South America with 28 minor haplotypes that contain between one and four sequences. The haplotype C (we choose C for “commercial”) contains all the sequences derived from farmed strains of the industry working on BSF.

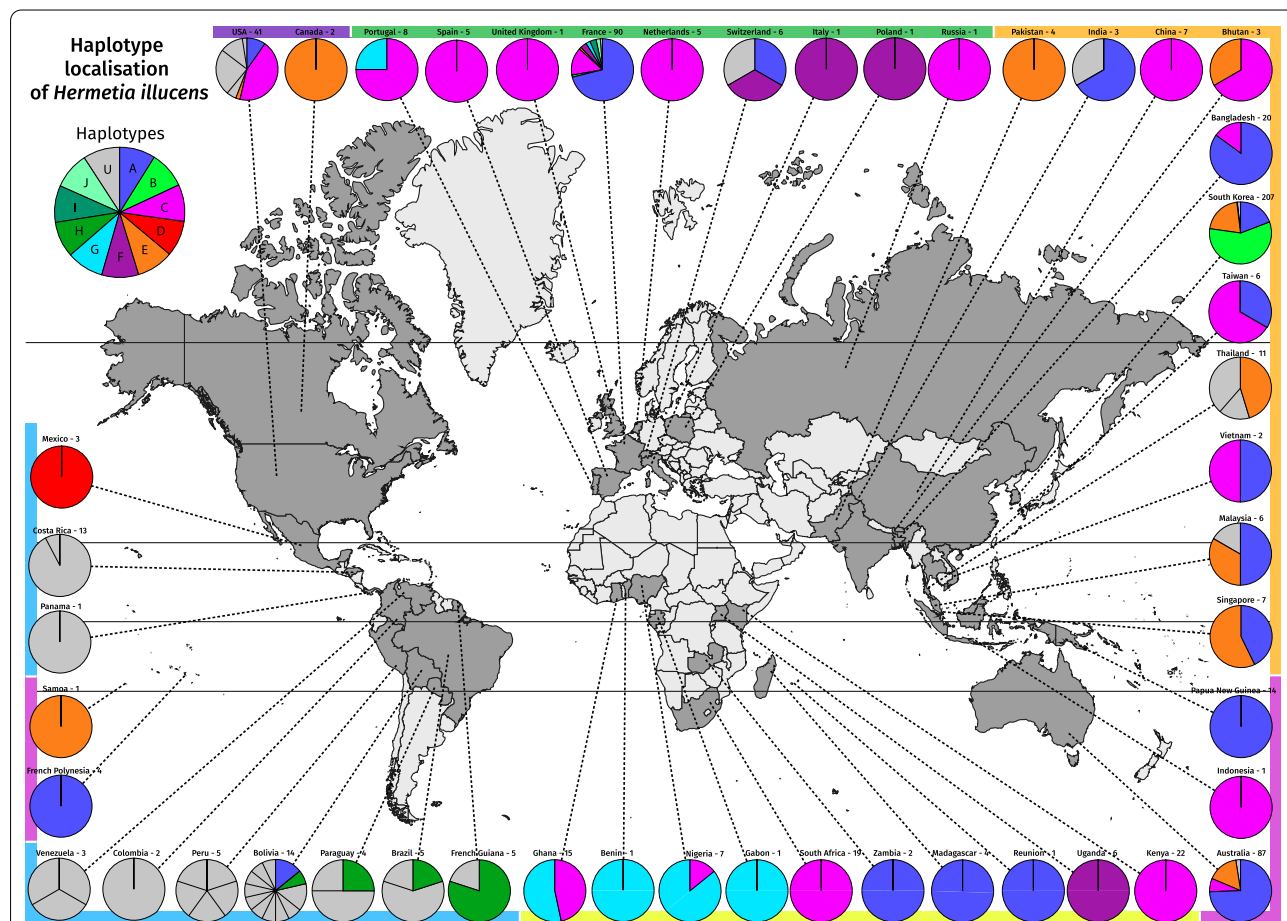
### Global distribution analysis

We visualized the actual distribution of the different haplotypes found previously on a distribution map (Fig. 2).

Each haplotype from A to J is now represented with a color (Fig. 2); haplotypes numbered in the haplotype network (Fig. 1) are now represented in grey.

We observed great variability in South America, highlighted on the map by the grey pie charts representing mostly minor haplotypes. We observe 22 singletons in South America out of the 52 haplotypes found, more than 40% of the total diversity. On the other hand, we found six out of ten major haplotypes in Europe.

Haplotype A and C were the most represented haplotypes in our sampling; Haplotype A was present on all continents and represented 34% of all sequences, and haplotype C was only absent from South America and represented 18% of all sequences. Some haplotypes



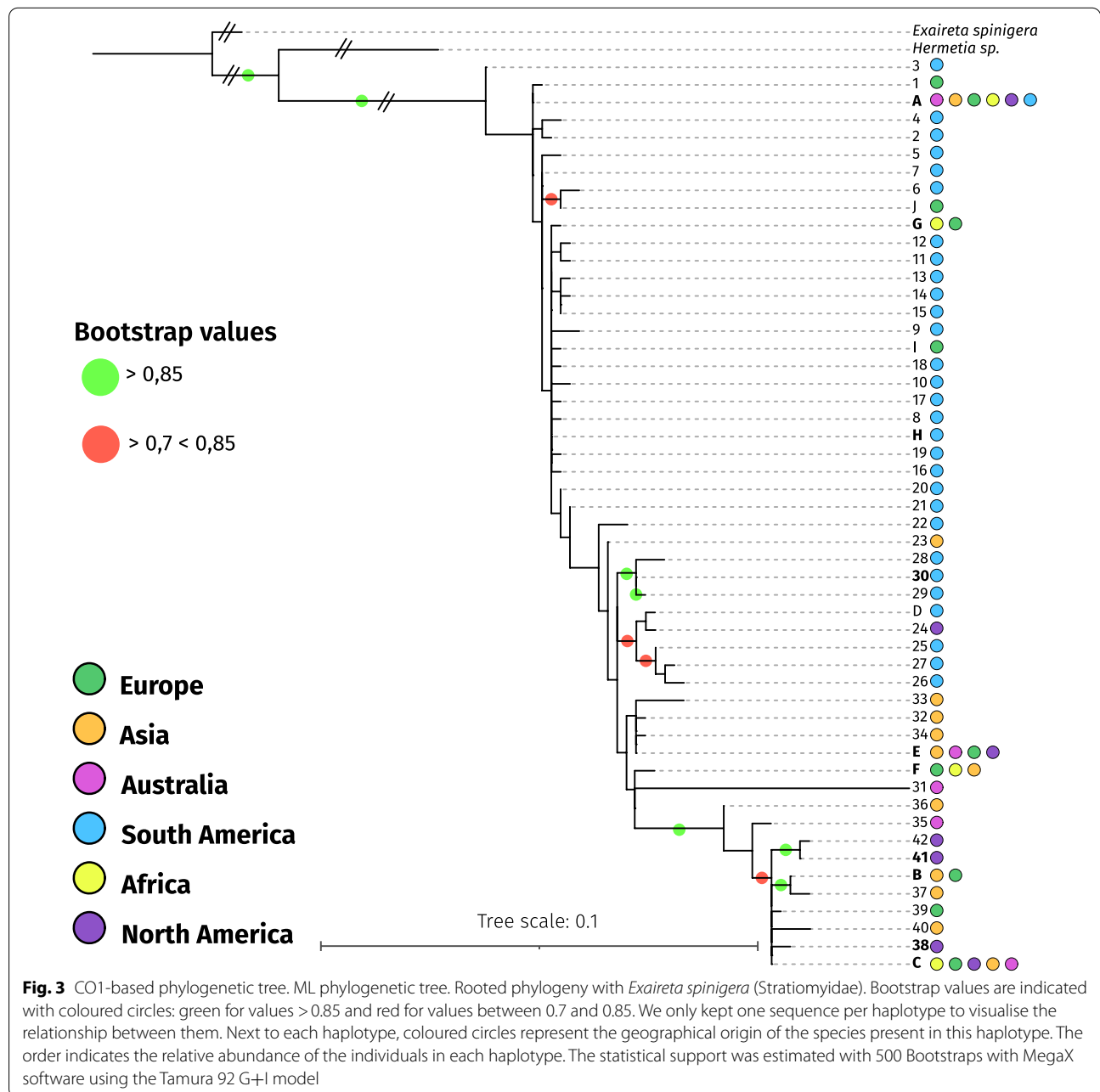
**Fig. 2** World CO1 haplotype distribution. Global distribution of the different *Hermetia illucens* haplotypes based on the CO1 gene. The pie charts represent the proportion of each haplotype within a country. The country's name is followed by the number of individuals analysed. A coloured band representing the continents is next to each pie chart. The colours represent the haplotypes determined in Fig. 1. The haplotypes represented with grey colours, indicated with the letter U, are represented with a number (1–42) in Fig. 1. Separation is made in the pie chart when the haplotypes change for the same country. For example, Venezuela contains three unique haplotypes, meaning that these three individuals are different. A summary table can be found in Additional file 1

seemed to be restricted to some world areas, such as haplotype G, which was present in all the localities in West Africa and surprisingly in a Portuguese sample, or the haplotype H in South America. Haplotype F was present in Europe, Uganda, and South Korea. Haplotype E was more present in Asia, Oceania, and North America. Finally, some haplotypes such as B, I, and J present rare sequences found in more restricted areas such as B in Korea or H and J in France (Fig. 2).

### CO1 phylogeny

During our CO1 gene analyses, some BSF-like larvae were identified as *Exaireta spinigera* larvae. The sequences were amplified and sequenced with the same primers as for BSF.

By taking only one sequence for each haplotype, we have reconstructed a phylogeny of the CO1 gene sequence. The CO1 phylogeny was rooted with the *Exaireta spinigera* (Fig. 3). A sample from Cameroon was revealed to belong to a closely unidentified species



**Fig. 3** CO1-based phylogenetic tree. ML phylogenetic tree. Rooted phylogeny with *Exaireta spinigera* (Stratiomyidae). Bootstrap values are indicated with coloured circles: green for values >0.85 and red for values between 0.7 and 0.85. We only kept one sequence per haplotype to visualise the relationship between them. Next to each haplotype, coloured circles represent the geographical origin of the species present in this haplotype. The order indicates the relative abundance of the individuals in each haplotype. The statistical support was estimated with 500 Bootstraps with MegaX software using the Tamura 92 G+I model

of BSF and was termed *Hermetia* sp. Cameroon in our subsequent analysis (Blastn score between *Hermetia* sp. from Cameroon and BSF gives us 85.01% identity with an E Value of  $2e-174$ ).

The CO1 gene phylogeny supports the analyses performed with the haplotype network (Fig. 1), showing remarkable sequence diversity among the CO1 gene. Interestingly, all the commercial strains are found in the C haplotype (a detailed phylogenetic tree is given as a supplementary with an asterisk placed on individuals from farms (Additional file 3: Fig. S3).

A group containing haplotypes 35 to 42 and haplotypes B and C forms a divergent group from the others, with strong statistical support (bootstrap value of 0.99). Some individuals captured in the wild have been found near BSF commercial breeding sites. For example, we have captured C haplotype individuals in Avignon close by companies working with the BSF.

Although the global resolution of the deep branches of the tree is low, it is also possible to better understand the relationship between the haplotypes: haplotypes B and C are close to each other and well separated from other haplotypes. The other haplotypes are not very well resolved with the CO1 sequences. Finally, the samples coming from Latin America are scattered all over the tree, showing their great diversity and supporting the hypothesis of the geographical origin of the BSF.

#### France displays vast BSF genetic diversity at local scales

Among the ten major haplotypes, our sampling effort in France showed that five are present in France, and a unique haplotype is only found in France (Fig. 4) in La Rochelle. Haplotype A remains the dominant haplotype in France (71.1% of all sequences) in seven of the eight localities (excluding Beaufort en Vallée, where the flies come from a company). Regarding the haplotype C, BSF belonging to this group have been found in four locations in the regions of Poitiers, Bordeaux, Beaufort-En-Vallée and Montpellier. Except for Beaufort-En-Vallée, where the flies come from a company, the other BSF come from natural environments. BSF industries exist in all these localities. More surprisingly, different haplotypes coexist in some localities. For example, we found three different haplotypes in the region of Paris and four in the region of Montpellier. Finally, we found a singular haplotype in Colomiers (J) and La Rochelle (1).

#### Complete mitochondrial genome analyses

To better understand BSF phylogeny and estimate the timing of haplotype divergence, we analyzed the mitochondrial genomes of 60 individuals (56 that have been sequenced and assembled, three from which sequences have been recovered and assembled and one recovered),

including 3 *Exaireta spinigera* genomes (used as an outgroup). The 60 reconstructed mitochondrial genomes have the typical structure of insect mtDNA: 13 protein-coding genes, 22 tRNA, two rRNA-coding genes, and non-coding regions and control regions (an example is given as a Additional file 4: Fig. S4). The genetic differences are homogeneously distributed overall mtDNA and are not localized on specific genes (Fig. 5A). A great genetic diversity at the mitochondrial genome level is observed. Two very homogeneous groups, A and C, are well represented, and sequences of the C haplotype display many mutations all along the genomes from the others. The two sequences with the most differences are 55-F-Angouleme and 21-Ghana, with a similarity percentage of 96.373% (all results are in Additional file 5. Shorter alignment lengths result from comparisons with the outgroup sequences that split the alignments into two fragments.) The mitochondrial genomes of the three *Exaireta spinigera* were extremely close to each other (100% by BLAST—Additional file 5), so we kept only one to serve as an outgroup.

We also wanted to know if the mutations are equally (or not) distributed between the different genes. Therefore, we compared the similarity of the 13 mitochondrial genes of the 57 BSFs (Fig. 5B) to those of *Exaireta spinigera*. The results are presented as percent sequence similarity.

We grouped the individuals into three different categories:

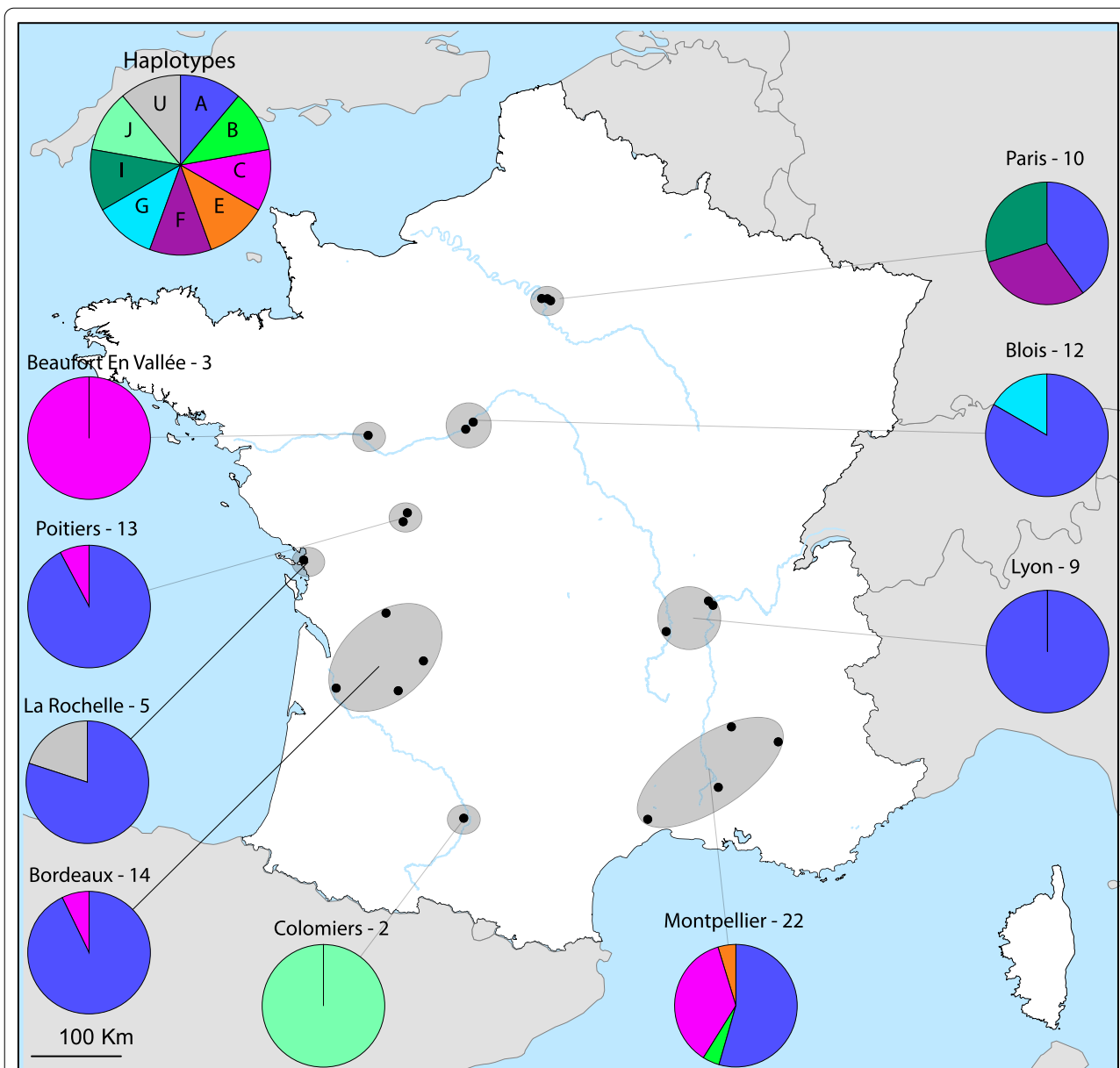
Haplotype A (Fig. 5B1) represents the percentages of differences of the 27 individuals of haplotype A compared with *Exaireta spinigera*. All individuals show the same level of gene difference as those of *Exaireta spinigera*. Only the NAD6 gene presents slight differences.

Haplotype C (Fig. 5B2) represents the percentages of differences of the 22 individuals of haplotype C compared to *Exaireta spinigera*. It shows that most of the mitochondrial genes are identical except for the ATP8 gene, which is different for four individuals (11 Angouleme; 14 Kenya; 19 Kenya, and 15 Kenya). Slight differences are also present in the NAD3 and NAD6 genes.

Haplotype D–F–G–H–I (Fig. 5B3) represents the percentages of differences of the eight other individuals of haplotype C compared with *Exaireta spinigera*. The profiles are slightly different, with a more substantial disparity for the ATP8, NAD1, and NAD5 genes, especially the Mexican one (57). The profiles show the same pattern except for the ATP8 gene of the individual from Mexico (57), which is more conserved.

#### Complete mitochondrial genome phylogeny

The complete mitochondrial genome phylogeny demonstrates the very high genetic divergence among BSF



**Fig. 4** France CO1 haplotype distribution. Distribution of the different haplotypes of *Hermetia illucens* in France. The dots represent the specific area of collection. The pie charts represent the distribution of haplotypes in the corresponding area. The names of the main cities in each area are indicated above, and the number of sequenced individuals. Paris includes Vincennes and Nogent Sur Marnes. Blois includes Valloire Sur Cisse. Lyon includes Saint Marcellin En Forez, Corbas, Champagne Au Mont d’Or and Dardilly. Montpellier includes Avignon, Trescléoux and Dieulefit. Bordeaux includes Lormont La Force and Périgueux. Poitiers includes Saint Georges Lès Baillargeaux

populations found with the CO1 gene analysis (Additional file 7: Fig. S7). The increased sensitivity of the analysis, due to the size of the sequences (13.96 kb, 13 genes), robustly resolves the relationships between different haplotypes identified previously (nodes > 95%). In addition, it allows us to distinguish subgroups within the previously determined haplotypes. Two major divisions emerge within the tree: group C (which includes all commercial

strains) and others. Two subgroups emerge within the C group (11—F—Angouleme; 14 Kenya; 19 Kenya; 15 Kenya and the others).

The other large group contains sequences of A, D, I, G, and H haplotypes. The I G and H haplotypes are related and close to each other. The very cosmopolitan A haplotype is well conserved with limited mitochondrial sequence divergences.



Moreover, although phylogenetically close, the I, G, and H haplotypes are found in very disparate areas (the G haplotype is mainly in Africa, the I in Europe, and the H in Latin America), supporting multiple worldwide introductions.

Strikingly the high level of the conservation of the ATP8 sequence described above in the Mexican sample (haplotype D) does not explain the basal position of the sequence in the phylogeny; indeed, we tested the phylogeny by removing the ATP8 gene, and no change was found (Additional file 6: Fig. S6).

The ATP8 gene alone is not sufficient to explain the appearance of the observed subgroups. However, unlike the CO1 gene, which varies only slightly within this haplotype, the NAD2, NAD3, NAD5, and NAD6 genes are more variable.

The average percentage of identity between haplotypes A and C is about 96.39%. Between haplotypes C and D, 96.54%. Between haplotype C and H, 96.40% and 96.43% between C and I. Between A and D, 97.82% (all results are in Additional file 5: Fig. S5).

#### The divergence time of the different groups

The time tree (Fig. 6) gives us information about the divergence time of the different mitochondrial genomes. The phylogenetic tree presented in Fig. 6 contains both the Bayesian and the ML trees. The two separate trees are presented in Additional file 7. We noted a relatively long divergence time between haplotype C and the others (2.2 million years). The others have diverged more recently (about 1.2 million years). Within the C haplotype, a divergence has been about 300,000 years, while in the other groups, the divergence times are much lower, about 20,000 years for the A haplotype.

The time of divergence with *Exaireta spinigera* has been estimated at 110 million years, confirmed by the literature that gives us 112 million years with the Timetree tool [61].

## Discussion

*Hermetia illucens* is now considered a species of major interest for the feed industry [62, 63]. Thus, it became urgent to study the genetic diversity within the different populations at different scales. Analyzing the different haplotypes in the same geographical area is a powerful tool for studying the recent history of BSF. Hence, an in-depth study of the genetic variations between the different haplotypes may allow us to resolve better the scenario of the arrival of BSF in a territory.

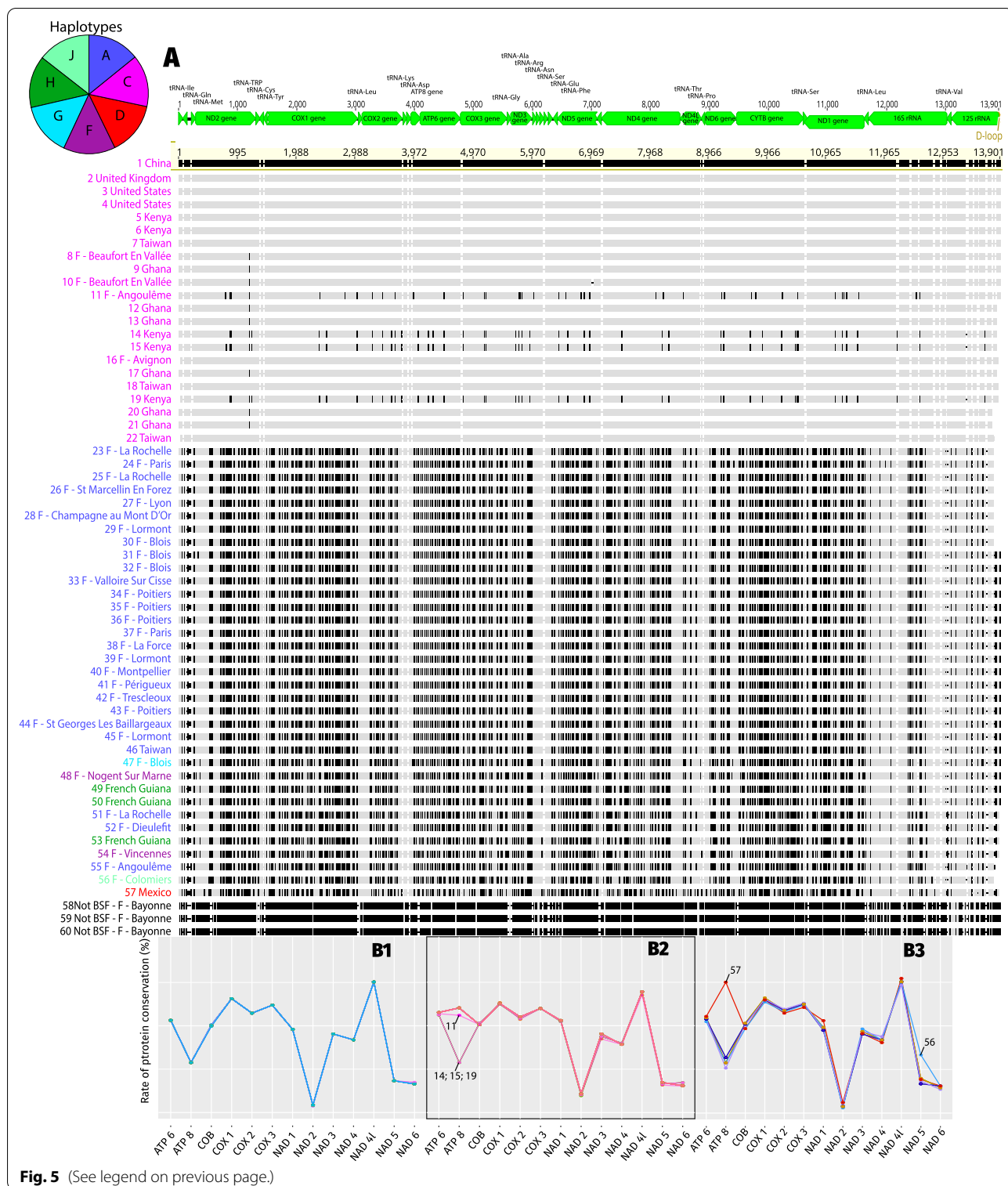
Our study uncovered a high genetic variability of mitogenomes within the species *Hermetia illucens*. We found that haplotype C is the most divergent. The dating of the division between the C haplotype and the others goes back more than 2 million years. We found a well-resolved separation within haplotype C (>95%); the other haplotypes are also well resolved (>95%).

Our study confirms the genetic diversity found in previous papers [2, 33, 34]. We found a comparable number of haplotypes, although we added many more sequences of diverse origin but mainly from non-indigenous areas. For example, Ståhls et al. [34] found 56 distinct haplotypes with 418 CO1 sequences, and we found a comparable number of haplotypes based on the analysis of 677 CO1 sequences. These results suggest we have now captured a major part of the extant BSF worldwide diversity outside of the genetic hotspots identified previously [2, 34], namely Latin America. Indeed, in South America, we found 30 haplotypes among the 55 sequences sampled that open up the possibility of a still unknown genetic diversity in this part of the world.

With the maximum diversity found in Latin America, the hypothesis of the geographical origin of BSF [1] is supported by our study. Analysis of the complete mitochondrial genome supports the hypothesis of several distinct introductions on all non-native continents. Considering the estimated divergence times between the different haplotypes, it seems impossible that this is due to a single introduction followed by diversification in the territory.

(See figure on next page.)

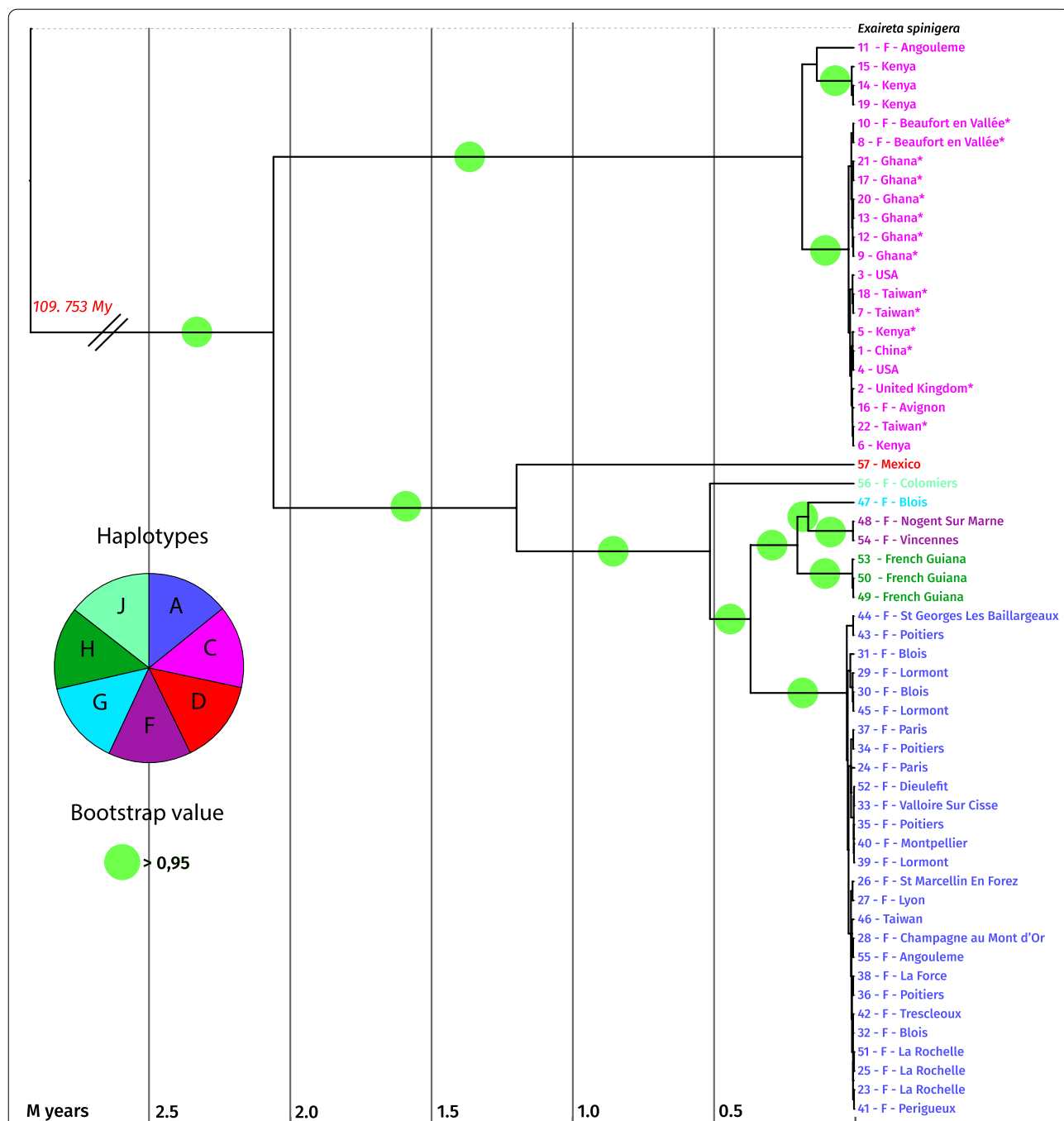
**Fig. 5** Mitochondrial DNA alignment and sequence conservation between *Hermetia illucens* and *Exaireta spinigera*. **A** Multiple alignments of mitochondrial DNA sequences from *Hermetia illucens*. Alignment was performed with the MAFFT aligner. The colours of the names come from their haplotypes determined with the CO1 sequences. Each black mark represents a difference with the reference (1—China). Above the alignment, there is a linear representation of the mitochondrial genes (a circular one can be seen in Additional file 4). The three last sequences are those of *Exaireta spinigera* analysed in the same way as *Hermetia illucens* (we kept one that we used as an outgroup because it was the closest complete mitochondrial sequence phylogenetically that we had). The sequences are numbered to be found in Fig. 6. **B** Sequence conservation of mitochondrial genes of *Hermetia illucens* to *Exaireta spinigera*. The sequences are grouped according to haplotypes. The A haplotype (**B1**), the C haplotype (**B2**), and the other haplotypes (**B3**). The sequences inside each graph are represented in the alignment (5A). Sequences from 23 to 46 plus sequences 51, 52, and 55 for haplotype A (**B1**); from 1 to 22 for haplotype C (**B2**); and from 47 to 57 (except sequences 51, 52, and 55) for the other haplotypes (**B3**)



**Fig. 5** (See legend on previous page.)

On the other hand, according to Kaya et al. [2], an alternative but complementary scenario is also possible: bridgehead populations, consisting of a mixture of haplotypes, have been introduced on different continents.

Indeed, taking Europe as an example, the eight haplotypes found in Europe can be explained by the admixture scenario proposed by Kaya et al. [2], who argue that central European BSF populations originated from



**Fig. 6** Mitochondrial DNA phylogenetic and time tree. Bayesian time tree obtained with BEAST of 57 mitochondrial genomes of *Hermetia illucens* deprived of D-Loop regions, aligned with MAFFT and rooted with *Exaireta spinigera*. The scale is in millions of years. A break in scale is made at the root level. The colours of the names correspond to the haplotypes determined with the CO1 sequences (Fig. 1). Commercial individuals are marked with an asterisk. The green circles come from the ML phylogenetic tree of the 57 mitochondrial DNA sequences of *Hermetia illucens*, deprived of D-Loop regions. MAFFT was used to align the sequences, 500 Bootstrap replicas were performed. The outgroup used is *Exaireta spinigera* (Stratiomyidae). Bootstrap values are indicated by a green circle when they are > 0.95 (the separate trees can be seen in Additional file 7)

admixture between Australian and Southeast African populations, which themselves can be traced back to earlier independent admixture events. Furthermore, Kaya

et al. [2] suggested some gene flows with Mediterranean (altogether originating from Latin America) and particularly East African populations, which interestingly seems

to be corroborated in our study, as exemplified by the F haplotype. It would be extremely useful in the future to combine nuclear data and mitochondrial data to better resolve the potentially diverging patterns of the different markers and to precisely estimate relative impacts of historic genetic admixtures versus recent introductions on regional population structures. France contains on its territory a great diversity of haplotypes, whose origin can be explained either by the historical introduction around the year 1950 [11] in addition to more recent and repetitive introductions. The haplotype A, which is by far the most common haplotype in wild BSF found in France (71.1%) and in Europe, may correspond to the initial introduction. On the other hand, our data suggest that the presence of haplotype C in France may be due to a more recent introduction caused by BSF industries.

As suggested in a recent article on the structure and demography of the BSF based on microsatellite data [2], the captive populations used for breeding in Europe and North America derive from a common, genetically related strain. Our analysis based on CO1 and complete mitochondrial genome sequences confirm this result: the haplotype we named C contains all the sequences of industrial origin with very little genetic diversity. Interestingly, this commercial group does not have any South American relatives. Based on nuclear genetic data, Kaya et al. report wild North American populations being most closely related to the most prevalent managed strains, which we infer to be dominated by haplotype C. This result reinforces the hypothesis of a North American origin of the BSF commercial stock.

Moreover, specific individuals in the wild belonging to this group were collected in the vicinity (around 5 km) of industries breeding BSF. For example, if we focus on the commercial haplotype analysis in France, the wild BSF belonging to haplotype C may have escaped from industries in Poitiers, Bordeaux, Avignon, and Montpellier. This last point raises the question of the biosecurity of BSF farms worldwide. We can assume that the priority of companies involved in BSF farming is not biosecurity and escape control. Indeed, costs in terms of security to prevent insects from escaping are high, and the regulation is late at this level. In that case, the continuous flow of escaping BSF of industrial origin in large numbers could either result in an establishment in territories where it was not yet present or trigger losses of genetic integrity of locally resident wild populations via continuous one-directional introgression. This would lead inexorably to a loss of the local genetic diversity that may prevent future adaptations to changing environmental conditions.

However, it is interesting to note that in Ståhls et al. [34], in contrast to our study, some genetic diversity in the breeding strains in Europe and North America has

been evidenced (10 haplotypes). This may be due to a broader sampling effort of commercial strains. Indeed, Ståhls [34] had the opportunity to sample many small farms leading to 292 BSF sequences from rearing cultures. This allowed them to access a broader genetic diversity not found in our study. It is also possible that smaller farms did not obtain BSF from the dealers who constituted most of the more prominent industries but directly from the wild. Nevertheless, our results confirm that the industrial actors work with the same and unique haplotype.

We also evidence that the mitochondrial genome is ideally suited to resolve long-term BSF phylogenetic relationships. It allows us to date the natural history of the BSF, bringing a sum of details that were impossible to distinguish based only on the CO1. For example, we report puzzling variations in the level of conservation of the ATP8 genes for some samples (haplotypes C and D). Interestingly, these variations have already been found in other insect species (for example, in Coleoptera, see Zhang et al. [63], and in grasshoppers, see Li et al. [64]). The ATP8 gene is small (160 bp on average in insects) and encodes for a subunit of the complex V of the ATP synthase. In *Hermetia illucens*, the initiator codon of the ATP8 gene is ATT, is 167 bp long, and it overlaps with the ATP6 gene. It has been assumed in other species that the modification of the sequence of the ATP8 gene can affect the structure of mitochondrial complex V and thus its function.

Furthermore, it has been shown that changes within ATP synthase can be explained by adaptation to different ecological environments [65–67]. This indicates that some local adaptation driven by the ATP8 genes may exist in BSF, as it has already been shown in other insects. This can lead to an adaptation to survive in high-altitude environments, i.e., at a lower oxygen level and low temperature [68]. Even if these variations did not change the phylogeny of the BSF, it could be a clue to some local adaptation signal. The fact that all commercial strains (plus USA plus Mexico) show a high level of conservation of the ATP8 gene compared to all other wild strains may suggest some adaptation to rearing.

Complete mitochondrial genome data allows us to make a robust and well-resolved time tree of the species. The divergence times between the different haplotypes are surprisingly long (more than 2 million years for the most divergent lineages), and the differences observed between the sequences prove that there is still unsuspected and hidden genomic diversity. Comparing it to *Drosophila*, 2 million years correspond to many speciation events leading to different reproductively isolated species [69]. By contrast, Ståhls et al. [34] indicated that BSF belonging to phylogenetically divergent haplotypes

can still cross and generate offspring, indicating no apparent reproductive isolation in laboratory conditions. Interestingly, our data evidence that the cohabitation of different haplotypes at the same locality could lead to crosses between haplotypes. According to our results, the 2 million years of divergence between the C haplotypes and the other haplotypes could correspond to the North American lineages separating from the ancestral BSF population in South America upon their northwards range expansion beyond the Isthmus of Panama, which appears 2.8 million years ago [70]. This seems especially true since the C haplotype has not been found in South America but only in North America. This ancestral and natural colonization of North America more than 2 million years ago may also explain why the C haplotype group displays a higher degree of mitochondrial genetic diversity than the other groups, a process already observed in other insects as ants for example [71]. Further investigations are thus needed to understand the possible genetic barriers between highly divergent BSF groups but also the possible patterns of introgression in natural BSF populations. This phenomenon might considerably confuse the elucidation of the population structure and the natural history of these species.

Finally, it is interesting to recall that, if indeed, most industrial actors use the same haplotype for the industry, the other haplotypes present in France are highly divergent from this one. Strikingly, we found in one compost bin in the region of Paris flies belonging to two different haplotypes (A and F). Thus, crossbreeding between these two individuals belonging to these two haplotypes would bring a solid genetic mix that could benefit breeding. Associated with this significant genetic diversity, BSF may also have some significant phenotypic diversity that might be useful for specific industrial usage and selection processes to increase performances or ameliorate peculiar traits or behaviors [72].

## Conclusion

Through the analysis of 60 mitochondrial genomes during this study, we proposed a robust and well-resolved time tree of the BSF that confirms the vast genetic diversity within *Hermetia illucens*. Our result elucidates the phylogenetic relationship and time divergence between the BSF lineages. By adding almost 20% of CO1 sequences, mainly from non-native areas, we did not significantly increase the number of haplotypes previously found [34]. This indicates that a significant part of the diversity from non-native areas has already been found and that we should now concentrate our efforts on native areas. Furthermore, the combination of our results with the nuclear analysis already performed [2] gives an interesting insight into the invasion routes of the BSF and the

divergence times between the different groups. We have also shown that many BSF industrial companies work with flies derived from a unique haplotype. We would also like to recall that the apparent biosecurity deficiency within some BSF industrial companies probably led to a continuous flow of escapement and possible local introduction/replacement of natural populations of BSF. Finally, our data indicated that it becomes fascinating to look at the nuclear genomes' evolution of the BSF to better understand the specific adaptations of the different lineages that might be useful for industrial needs and initiate the selection process for peculiar traits.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-022-02025-6>.

- Additional file 1.** List of all individuals used during the study. Origin and accession number.
- Additional file 2.** CO1 gene sequencing protocol. Primers used and Tm.
- Additional file 3.** Phylogenetic tree of the CO1 gene.
- Additional file 4.** Annotated mitochondrial genome of *Hermetia illucens*.
- Additional file 5.** Paired blasts of the mitochondrial genomes used in the study.
- Additional file 6.** Phylogeny based on the complete mitochondrial genome of *Hermetia illucens* with the ATP8 gene sequence removed.
- Additional file 7.** Mitochondrial DNA phylogenetic tree and mitochondrial DNA time tree.

## Acknowledgements

We want to thank Floran Laville and Frederic Marion-Poll for their continuous support and help during this project. We warmly thank the members of the Réseau Compost Citoyen for their precious help during the sampling in France. We also thank Matan Shelomi (National Taiwan University), Meng-Kun Wu, Shu-Min Chen, Jing-Jiun Huang (Yi Mi Community College, Taiwan) for flies from Taiwan. We thank the International Centre of Insect Physiology and Ecology (ICIPE) from Nairobi City, Kenya. We thank Mathieu Chouteau from CNRS LEESIA, in French Guiana. We finally thank Benoit Gilles for his help at the beginning of this project and the two anonymous reviewers for their constructive comments.

## Author contributions

JG carried out the sampling, performed data curation, and wrote the original draft. NP performed data curation and reviewed the manuscript. GB helped with the data curation and reviewed the manuscript. JF conceptualized and supervised the study, performed data curation, and reviewed and finalized the manuscript. All authors read and approved the final manuscript.

## Funding

This work was funded in the framework of a partnership between the CNRS and the CycleFarms company. A CIFRE PhD supports JG. Grant (CIFRE 2018/1095) from the Agence Nationale pour la Recherche et la Technologie (ANRT).

## Availability of data and materials

The dataset generated during the study has been deposited in the European Nucleotide Archive database under Accession Numbers PRJEB48031 Bioproject. Alignment and raw data can be downloaded following this link: [https://figshare.com/articles/dataset/sequences\\_and\\_trees\\_zip/19589335](https://figshare.com/articles/dataset/sequences_and_trees_zip/19589335). The supplementary can be downloaded following this link: [https://figshare.com/articles/dataset/Supplementary\\_zip/19589332](https://figshare.com/articles/dataset/Supplementary_zip/19589332).

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The private company Cycle Farms partially granted this project. The company has not influenced the design and the analysis of the results and has not influenced the writing of the manuscript and its conclusions. No commercial products have been developed based on the outputs of this article. None of the authors has any competing interests.

### Author details

<sup>1</sup>EGCE, CNRS, 12 rue 128, 91190 Gif sur Yvette, France. <sup>2</sup>Cycle Farms, 6 boulevard de entrepreneurs, 49250 Beaufort En Anjou, France. <sup>3</sup>IDEEV - Laboratoire Evolution, Génomes Comportement, Ecologie CNRS Université Paris Sud UMR 9191, IRD UMR 247, 12 rue 128, 91198 Gif sur Yvette, France.

Received: 11 January 2022 Accepted: 13 May 2022

Published online: 01 June 2022

## References

- Hauser M, Woodley M. The historical spread of the black soldier fly, *Hermetia illucens* (L.) (Diptera: Stratiomyidae, Hermetiinae), and its establishment in Canada. *J Kansas Entomol Soc.* 2015;146:51–4.
- Kaya C, Generalovic TN, Ståhls G, Hauser M, Samayoa AC, Nunes-Silva CG, et al. Global population genetic structure and demographic trajectories of the black soldier fly, *Hermetia illucens*. *BMC Biol.* 2021;19(1):94.
- Rozkosný R. A biosystematic study of the European Stratiomyidae (Diptera): volume 2—Clitellariinae, Hermediinae, Pachygasterinae and bibliography. Amsterdam: Springer Netherlands; 1983. (Series entomologica). <https://www.springer.com/gp/book/9789061931355>. Accessed 29 Dec 2020.
- Mason F, Rozkošný R, Hauser M. A review of the soldier flies (Diptera: Stratiomyidae) of Sardinia. *Zootaxa.* 2009;2318(1):507–30. <https://tinyurl.com/4f6nbe2t>.
- Wang Y-S, Shelomi M. Review of black soldier fly (*Hermetia illucens*) as animal feed and human food. *Foods.* 2017;6(10):91. <https://www.cbd.int/invasive/WhatareIAS.shtml>.
- Picker M, Griffiths C, Weaving A. Field guide to insects of South Africa. Cape Town: Penguin Random House South Africa; 2002. p. 452.
- Lindner E. Die amerikanische *Hermetia illucens* L. im Mittelmeergebiet: (Stratiomyiidae, Dipt.). Akadem. Verlagsges; 1936. p. 2.
- Benelli G, Canale A, Raspi A, Fornaciari G. The death scenario of an Italian Renaissance princess can shed light on a zoological dilemma: did the black soldier fly reach Europe with Columbus? *J Archaeol Sci.* 2014;49:203–5.
- Barbier J. Introduction en France d'un Diptère Stratiomyide américain. *Bull Soc Entomol France.* 1952;57(7):108.
- Dauphin P. Présence de *Hermetia illucens* (LINNÉ, 1758) dans le sud-ouest de la France (Diptera Stratiomyiidae). p. 3.
- Phillippe R. Sur la présence d'*Hermetia illucens* (Linnaeus, 1758) (Diptère Stratiomyiidae) dans la région lyonnaise/présence of *Hermetia illucens* (Linnaeus, 1758) (Diptera Stratiomyiidae) in the region of Lyons. *Bull mensuel de la Société linnéenne de Lyon.* 2009;78:137–8.
- Pierre-Olivier M, Denis R, Willems J. First record of the black soldier fly, *Hermetia illucens*, in the Western regions of France (Vendée, Loire-Atlantique, Ille-et-Vilaine) with notes on its worldwide repartition (Diptera, Stratiomyiidae). *Bull Soc Entomol France.* 2020;125:13–8.
- Leclercq M. Dispersion and transport of noxious insects: *Hermetia illucens* L. (Diptera Stratiomyidae). *Bulletin des recherches agronomiques de Gembloux.* 1969. [https://scholar.google.com/scholar\\_lookup?title=Dispersion+and+transport+of+noxious+insects%3A+Hermetia+illucens+L.%28Diptera+Stratiomyidae%29&author=Leclercq%2C+M.&publication\\_year=1969](https://scholar.google.com/scholar_lookup?title=Dispersion+and+transport+of+noxious+insects%3A+Hermetia+illucens+L.%28Diptera+Stratiomyidae%29&author=Leclercq%2C+M.&publication_year=1969). Accessed 9 Nov 2021.
- Craig Sheppard D, Larry Newton G, Thompson SA, Savage S. A value added manure management system using the black soldier fly. *Bioresour Technol.* 1994;50(3):275–9.
- Lord WD, Goff ML, Adkins TR, Haskell NH. The black soldier fly *Hermetia illucens* (Diptera: Stratiomyidae) as a potential measure of human post-mortem interval: observations and case histories. *JFS.* 1994;39(1):215–22.
- Gold M, Tomberlin JK, Diener S, Zurbrügg C, Mathys A. Decomposition of biowaste macronutrients, microbes, and chemicals in black soldier fly larval treatment: a review. *Waste Manag.* 2018;82:302–18.
- Hoc B, Genva M, Fauconnier M-L, Lognay G, Francis F, Caparros MR. About lipid metabolism in *Hermetia illucens* (L. 1758): on the origin of fatty acids in prepupae. *Sci Rep.* 2020;10(1):11916.
- Ewald N, Vidakovic A, Langeland M, Kiessling A, Sampels S, Lalander C. Fatty acid composition of black soldier fly larvae (*Hermetia illucens*)—possibilities and limitations for modification through diet. *Waste Manag.* 2020;102:40–7.
- Giannetto A, Oliva S, Ceccon Lanes CF, de AraújoPedron F, Savastano D, Baviera C, et al. *Hermetia illucens* (Diptera: Stratiomyidae) larvae and prepupae: biomass production, fatty acid profile and expression of key genes involved in lipid metabolism. *J Biotechnol.* 2019;307:44–54.
- Spranghers T, Ottoboni M, Klootwijk C, Obyn A, Deboosere S, De Meulenaer B, et al. Nutritional composition of black soldier fly (*Hermetia illucens*) prepupae reared on different organic waste substrates. *J Sci Food Agric.* 2017;97(8):2594–600.
- van Huis A. Edible insects: future prospects for food and feed security (FAO forestry paper). Rome: Food and Agriculture Organization of the United Nations; 2013. p. 187.
- Leong SY, Kutty SRM, Malakahmad A, Tan CK. Feasibility study of biodiesel production using lipids of *Hermetia illucens* larva fed with organic waste. *Waste Manag.* 2016;47(Pt A):84–90.
- Surendra KC, Tomberlin JK, van Huis A, Cammack JA, Heckmann LHL, Khanal SK. Rethinking organic wastes bioconversion: evaluating the potential of the black soldier fly (*Hermetia illucens* (L.)) (Diptera: Stratiomyidae) (BSF). *Waste Manag.* 2020;117:58–80.
- Heuel M, Sandrock C, Leiber F, Mathys A, Gold M, Zurbrügg C, et al. Black soldier fly larvae meal and fat can completely replace soybean cake and oil in diets for laying hens. *Poult Sci.* 2021;100(4): 101034.
- Schiavone A, Dabbou S, Petracci M, Zampiga M, Sirri F, Biasato I, et al. Black soldier fly defatted meal as a dietary protein source for broiler chickens: effects on carcass traits, breast meat quality and safety. *Animal.* 2019;13(10):2397–405.
- Neumann C, Velten S, Liebert F. N balance studies emphasise the superior protein quality of pig diets at high inclusion level of algae meal (*Spirulina platensis*) or insect meal (*Hermetia illucens*) when adequate amino acid supplementation is ensured. *Animals.* 2018;8(10):172.
- Biasato I, Renna M, Gai F, Dabbou S, Meneguz M, Perona G, et al. Partially defatted black soldier fly larva meal inclusion in piglet diets: effects on the growth performance, nutrient digestibility, blood profile, gut morphology and histological features. *J Anim Sci Biotechnol.* 2019;10:12.
- Khamis FM, Ombura FLO, Akutse KS, Subramanian S, Mohamed SA, Fiaboe KKM, et al. Insights in the global genetics and gut microbiome of black soldier fly, *Hermetia illucens*: implications for animal feed safety control. *Front Microbiol.* 2020. <https://doi.org/10.3389/fmicb.2020.01538/full>.
- Erickson MC, Islam M, Sheppard C, Liao J, Doyle MP. Reduction of *Escherichia coli* O157:H7 and *Salmonella enterica* serovar Enteritidis in chicken manure by larvae of the black soldier fly. *J Food Prot.* 2004;67(4):685–90.
- Liu Q, Tomberlin JK, Brady JA, Sanford MR, Yu Z. Black soldier fly (Diptera: Stratiomyidae) larvae reduce *Escherichia coli* in dairy manure. *Environ Entomol.* 2008;37(6):1525–30.
- Park S, Choi H, Choi JY, Jeong G. Population structure of the exotic black soldier fly, *Hermetia illucens* (Diptera: Stratiomyidae) in Korea. *Korean J Environ Ecol.* 2017;31(6):520–8.
- Ståhls G, Meier R, Sandrock C, Hauser M, ŠašićZorić L, Laiho E, et al. The puzzling mitochondrial phylogeography of the black soldier fly (*Hermetia illucens*), the commercially most important insect protein species. *BMC Evol Biol.* 2020;20(1):60.
- Qi Y, Xu J, Tian X, Bai Y, Gu X. The complete mitochondrial genome of *Hermetia illucens* (Diptera: Stratiomyidae). *Mitochondrial DNA Part B.* 2017;2(1):189–90.

36. Generalovic TN, McCarthy SA, Warren IA, Wood JMD, Torrance J, Sims Y, et al. A high-quality, chromosome-level genome assembly of the black soldier fly (*Hermetia illucens* L.). bioRxiv. 2020. <https://doi.org/10.1093/g3journal/jkab085>.
37. Zhan S, Fang G, Cai M, Kou Z, Xu J, Cao Y, et al. Genomic landscape and genetic manipulation of the black soldier fly *Hermetia illucens*, a natural waste recycler. *Cell Res*. 2020;30(1):50–60. <https://reseaucompst.org/>.
38. <https://tinyurl.com/46a7xejr>.
39. Desjardins P, Conklin D. NanoDrop microvolume quantitation of nucleic acids. *J Vis Exp*. 2010;45:2565. <https://tinyurl.com/evaah44c>.
40. <https://www.geneious.com>.
41. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
42. Maddison DR, Swofford DL, Maddison WP. NEXUS: an extensible file format for systematic information. *Syst Biol*. 1997;46(4):590–621.
43. Leigh JW, Bryant D. popart: full-feature software for haplotype network construction. *Methods Ecol Evol*. 2015;6(9):1110–6.
44. Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*. 2010;38(6):1767–71.
45. Hansen KD, Brenner SE, Dudoit S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res*. 2010;38(12): e131.
46. Erlich Y, Mitra PP, Delabastide M, McCombie WR, Hannon GJ. Alta-cyclic: a self-optimising base caller for next-generation sequencing. *Nat Methods*. 2008;5(8):679–82.
47. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, et al. Synthetic spike-in standards for RNA-seq experiments. *Genome Res*. 2011;21(9):1543–51.
48. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol*. 2013;20(9):1131–9.
49. Meng G, Li Y, Yang C, Liu S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualisation. *Nucleic Acids Res*. 2019;47(11): e63.
50. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31(10):1674–6.
51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
52. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, et al. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 2013;69(2):313–9.
53. Leinonen R, Sugawara H, Shumway M. International nucleotide sequence database collaboration. The sequence read archive. *Nucleic Acids Res*. 2011;39(Database issue):D19–21.
54. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35(6):1547–9.
55. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49(W1):W293–6.
56. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol*. 2018. <https://doi.org/10.1093/ve/vey016>.
57. Abrami G, Mehler A, Lücking A, Rieb E, Helfrich P. TextAnnotator: a flexible framework for semantic annotations. In: Proceedings of the 15th joint ACL-ISO workshop on interoperable semantic annotation (ISA-15). Association for Computational Linguistics, London; 2019. p. 1–12.
58. Paradis E. Analysis of haplotype networks: the randomised minimum spanning tree method. *Methods Ecol Evol*. 2018;9(5):1308–17.
59. Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol*. 2015;32(4):835–45.
60. Tomberlin JK, van Huis A. Black soldier fly from pest to 'crown jewel' of the insects as feed industry: an historical perspective. *J Insects Food Feed*. 2020;6(1):1–4.
61. Zhang S, Shu J, Wang Y, Liu Y, Peng H, Zhang W, et al. The complete mitochondrial genomes of two sibling species of camellia weevils (Coleoptera: Curculionidae) and patterns of Curculionini speciation. *Sci Rep*. 2019;9(1):3412.
62. Li X-D, Jiang G-F, Yan L-Y, Li R, Mu Y, Deng W-A. Positive selection drove the adaptation of mitochondrial genes to the demands of flight and high-altitude environments in grasshoppers. *Front Genet*. 2018;9:605.
63. Peng Q, Tang L, Tan S, Li Z, Wang J, Zou F. Mitogenomic analysis of the genus *Pseudois*: evidence of adaptive evolution of morphological variation in the ATP synthase genes. *Mitochondrion*. 2012;12(5):500–5.
64. Gu P, Liu W, Yao Y, Ni Q, Zhang M, Li D, et al. Evidence of adaptive evolution of alpine pheasants to high-altitude environment from mitogenomic perspective. *Mitochondrial DNA Part A*. 2016;27(1):455–62.
65. Zhang Q-L, Zhang L, Zhao T-X, Wang J, Zhu Q-H, Chen J-Y, et al. Gene sequence variations and expression patterns of mitochondrial genes are associated with the adaptive evolution of two *Gynaephora* species (Lepidoptera: Lymantriinae) living in different high-elevation environments. *Gene*. 2017;610:148–55.
66. Luo Y, Yang X, Gao Y. Mitochondrial DNA response to high altitude: a new perspective on high-altitude adaptation. *Mitochondrial DNA*. 2013;24(4):313–9.
67. Russo CAM, Mello B, Frazão A, Voloch CM. Phylogenetic analysis and a time tree for a large drosophilid data set (Diptera: Drosophilidae). *Zool J Linn Soc*. 2013;169(4):765–75.
68. O'Dea A, Lessios HA, Coates AG, Eytan RI, Restrepo-Moreno SA, Cione AL, et al. Formation of the isthmus of Panama. *Sci Adv*. 2016;2(8): e1600883.
69. Winston ME, Kronauer DJC, Moreau CS. Early and dynamic colonization of Central America drives speciation in Neotropical army ants. *Mol Ecol*. 2017;26(3):859–70.
70. Zhou F, Tomberlin JK, Zheng L, Yu Z, Zhang J. Developmental and waste reduction plasticity of three black soldier fly strains (Diptera: Stratiomyidae) raised on different livestock manures. *J Med Entomol*. 2013;50(6):1224–30.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

