DR. JOSE DE LA FUENTE (Orcid ID : 0000-0001-7383-9649)

Article type   : Original Article

Biotic and abiotic factors shape the microbiota of wild-caught populations of the arbovirus vector *Culicoides imicola*

Sandra Díaz-Sánchez[1&], Angélica Hernández-Jarguín[1&], Alessandra Torina[2&], Isabel G. Fernández de Mera[1], Agustín Estrada-Peña[3], Margarita Villar[1], Francesco La Russa[2], Valeria Blanda[2], Joaquín Vicente[1], Santo Caracappa[1], Christian Gortazar[1], José de la Fuente[1,4*]

[1]SaBio. Instituto de Investigación en Recursos Cinegéticos IREC-CSIC-UCLM-JCCM, Ronda de Toledo s/n, 13005 Ciudad Real, Spain.

[2]Intituto Zooprofilattico Sperimentale della Sicilia, Via G. Marinuzzi n°3, 90129 Palermo, Sicily, Italy.

[3]Facultad de Veterinaria, Universidad de Zaragoza, 50013 Zaragoza, Spain.

[4]Department of Veterinary Pathobiology, Center for Veterinary Health Sciences, Oklahoma State University, Stillwater, OK 74078, USA.

[&]The first three authors contributed equally to the work reported in this paper.

Running title: *Culicoides imicola* microbiota

* Corresponding author: José de la Fuente. SaBio. Instituto de Investigación de Recursos Cinegéticos, IREC-CSIC-UCLM-JCCM, 13005 Ciudad Real, Spain; email: jose_delafuente@yahoo.com

Abstract

Biting midges of the genus *Culicoides* are known vectors of arboviruses affecting human and animal health. However, little is known about *Culicoides imicola* microbiota and its influence on this insect's biology. In this study, the impact of biotic and abiotic factors on *C. imicola* microbiota was characterized using shotgun-metagenomic sequencing of whole body DNA samples. Wild-caught *C. imicola* adult nulliparous females were sampled in two locations from Sicily, Italy. The climatic variables of temperature and soil moisture from both localities were recorded together with potential host blood meal sources. Shared core microbiome among *C. imicola* populations included *Pseudomonas, Escherichia, Halomonas, Candidatus* Zinderia*, Propionibacterium*, and *Schizosaccharomyces*. Specific and unique taxa were also found in *C. imicola* from each location, highlighting similarities and differences in microbiome composition between both populations. DNA and protein identification showed differences in host preferences between both populations with *Homo sapiens* and *Canis lupus familiaris* L. being the preferred blood meal source in both locations. A principal component analysis showed that the combined effect of host preferences (*H. sapiens*) and local soil moisture factors shape the microbiome composition of wild-caught populations of *C. imicola*. These results contribute to characterizing the role of the microbiome in insect adaptation and its utility in predicting geographic expansion of *Culicoides* species with potential implications for the control of vector-borne diseases.

Introduction

Biting midges of the genus *Culicoides* Latreille, 1809 (Diptera: Ceratopogonidae) are abundant haematophagous insects and vectors of arbovirus affecting human (Carpenter *et al*., 2013) and animal (Mellor *et al*., 2000) health including bluetongue virus (BTV) and Oropouche virus (OROV). During blood feeding, *Culicoides* lacerate the skin to ingest the effusion into this injury containing blood, skin cells and lymph (Pagès *et al*., 2014). In particular, *Culicoides imicola* Kieffer has been associated with the transmission of BTV in small ruminants and African horse sickness virus, which affects equids such as horses, mules, and donkeys (Purse *et al*., 2005). *C. imicola's* distribution was restricted to Africa and occasionally found in European Mediterranean countries (Purse *et al*., 2005; Venter *et al*., 2006) but it has been hypothesized that climate change in the last decades has extended their prevalence to other regions of the world (Acevedo *et al*., 2010; Guichard *et al*., 2014; Jacquet *et al*., 2015). How *C. imicola* have adapted to spread so quickly in different areas of Europe, Asia and Africa is not fully understood (Guichard *et al*., 2014; Jacquet *et al*., 2015).

After mating in swarms and blood meal ingestion, *Culicoides* females lay eggs in muddy areas with abundant organic material. The development from eggs to larvae, pupae and adults usually takes about 15-25 days, but can be up to 7 months during overwintering and is affected by climatic conditions (Benelli et al., 2017). The four larval instars live as omnivores/detritivores, thus eating both animal and plant derived material (Conte *et al*., 2007).

Like other insects, *Culicoides* spp. harbor a complex microbiome essential to maintain their life cycle and fitness through interactions with microbial endosymbionts (Campbell *et al*., 2004; Lewis *et al*., 2014; Nayduch *et al*., 2015). Most of these microbial endosymbionts are part of the core microbiome that function by maintaining important physiological processes within the host (Franzenburg *et al*.,

2013; Douglas, 2015). Biotic (host, plants) and abiotic (temperature, soil moisture) factors play an important role in the microbiome composition favoring selection of microbial communities acquiring specific-attributed functions needed for vector survival, adaptation and pathogen acquisition and transmission (Zouache *et al*., 2011; Minard *et al*., 2013; Villegas and Pimenta, 2014; Douglas, 2015). Acevedo *et al*. (2010) modeled the importance of biotic and abiotic factors such as spatial, topoclimatic, host and soil factors on *C. imicola* distribution. The authors showed that the effect of host and topoclimate factors followed by soil explains the variation in the abundance of *C. imicola*.

Limited information is available on the biotic and abiotic factors that impact the microbiome composition in wild-caught *Culicoides* spp. Most of the studies with *C. imicola* are descriptions of the microbiome composition, or solely targeting specific bacterial endosymbionts (Campbell *et al*., 2004; Lewis *et al*., 2014). For example, Campbell *et al*. (2004) concluded that the main sources of bacteria for *C. sonoriensis* was strongly linked to their breeding sites, affected by factors such as soil composition, plants and dung from small ruminants.

Novel whole genome shotgun metagenomics can offers a better resolution of the microbiome composition by sequencing the genome of all microorganisms present in the sample, including bacteria, archaea, viruses and parasites. Herein, we hypothesized that *C. imicola* populations share a core metagenome, but differences in abiotic and biotic factors such as temperature, soil moisture and host preferences may shape the microbiome potentially affecting insect adaptation and vector competence. To address this hypothesis, in this study a whole genome shotgun metagenomic sequencing was used to characterize the composition of the microbiome of wild-caught *C. imicola* collected from two locations with differences in biotic and abiotic factors to characterize the impact of these variables on microbiome composition.

Results

Characterization of *C. imicola* microbiome revealed similarities and differences between different wild-caught populations.

An experimental approach was developed to characterize the impact of biotic and abiotic factors on *C. imicola* microbiome under natural conditions (Fig. 1). The microbiome in whole bodies from two field-collected *C. imicola* populations was characterized using a whole genome shotgun metagenomic sequencing approach, and validated by quantitative polymerase chain reaction (qPCR) using selected bacteria (Additional file 1: Figs. S1A-S1B and Additional file 2: Dataset S1). The metagenomic sequencing revealed that *C. imicola* harbor a microbiome including representatives from both Eukaryote and Bacteria phyla in addition to viruses (Fig. 2 and Additional file 1: Figs. S2A-2C, Table S1).

Bacteria were the most represented kingdom including three bacterium phyla, 17 families, 20 genera, and 28 species (Fig. 2 and Additional file 1: Figs. S2A-2C, Table S1). A shared core microbiome in *C. imicola* populations from Collesano and Trapani was composed by the genera *Pseudomonas*, *Escherichia, Halomonas*, *Candidatus* Zinderia, *Propionibacterium*, and *Schizosaccharomyces* (Fig. 3). No differences in relative abundance were detected on the shared core microbiome, except for *P. putida*, which relative abundance was significantly higher ($P = 0.02$) in Collesano (14.4%) when compared to Trapani (0.3%) (Fig. 3 and Additional file 1: Figs. S1A-S1B and Table S1). In the shared core microbiome, the kingdom Eukaryota was represented by the genera *Schizosaccharomyces* (Fig. 3). None of the viruses identified in this study were found in the shared core microbiome of *C. imicola* (Figs. 3 and 4).

Differences in the microbiome composition were also found in *C. imicola* populations that evidenced the presence of unique species for each location (Fig. 4). In *C. imicola* from Collesano, the most abundant species in the microbiome belonged to *Bacillus cereus* with a 66.0% relative abundance (Figs. 2 and 4). However, in *C. imicola* from Trapani viruses and particularly *Musca hytrovirus* was the most abundant microorganism with an 82.4% relative abundance (Figs. 2 and 4).

Characterization of host DNA and proteins in *C. imicola* revealed differences in host preferences between both populations.

Gene sequences from the mitochondrial Cytochrome b gene (*Cyt b*) were used to identify potential host blood meal sources at species level (Fig. 5 and Additional file 1: Table S2). All the *Cyt b* sequences from *C. imicola* collected in Collesano (*n* = 11) were assigned to *Homo sapiens*. However, in Trapani the *Cyt b* sequences (*n* = 6) were assigned to *Canis lupus familiaris* L. as the main host blood meal source for *C. imicola* (66%), followed by *Ovis aries* L. and *H. sapiens* (17% each), revealing significant differences between both populations ($P \leq 0.001$). Although the highest sequence identity for *C. lupus familiaris* corresponded to *Cyt b* psedogenes, these sequences showed >99% query cover with >80% identity (E-value > 7e-83) only to *Canis* spp. mitochondrial *Cyt b* sequences (Additional file 1: Table S2), thus providing support for species identification. In contrast, the proteomics analysis identified proteins (*n* = 494) belonging to all host species included in the analysis (*H. sapiens, Bos taurus* L*., Sus scrofa* L*., O. aries* L*., Capra aegagrus hircus* and *C. lupus familiaris*), without differences in their representation as host blood meal source between both populations of *C. imicola* (Fig. 5 and Additional file 3: Dataset S2). However, as in the DNA analysis, the proteomics analysis identified *H. sapiens* (38%) (Collesano) and *H. sapiens* (36%) and *C. lupus familiaris* (20%) (Trapani) as the main host blood meal sources for *C. imicola* in these locations (Fig. 5 and Additional file 3: Dataset S2).

Characterization of abiotic factors revealed differences between *C. imicola* locations

Climate data was collected in both locations for all four seasons (Additional file 1: Table S3). Trapani and Collesano showed climate differences in terms of temperature and normalized difference in vegetation index (NDVI) due to the island topography. Every season was warmer in Trapani than in Collesano with differences of 27% (spring), 20% (summer), 17% (fall) and 26% (winter), and yearly differences of 22% ($P \leq 0.01$; Additional file 1: Table S3). However, the total number of days in which total temperature or seasonal temperature was above zero was similar in both locations. In

Collesano, soil moisture observed in terms of NDVI was higher than in Trapani (up to a 20%; Additional file 1: Table S3). However, plant coverage was more variable every season in Trapani than in Collesano.

The biotic and abiotic factors impact on the microbiome composition of *C. imicola*

The influence of the combined effect of abiotic (temperature and soil moisture) and biotic (host DNA) factors on *C. imicola* microbiome was characterized by PCA. The results of the three PC showed that within the 20 variables included in the study (Additional file 1: Table S4), the variation in *C. imicola* microbiome composition was explained by two factors (Table 1). The first PC explained 75% of the variation and was composed by the variables NDVI and *H. sapiens* blood meal source (Table 1). These results demonstrated that under our experimental conditions, the variations in *C. imicola* microbiome composition are explained by the combined effect of host preferences (*H. sapiens*) and NDVI (as a proxy for local soil moisture) factors.

Discussion

In this study, the metagenome and the impact of biotic (host blood meal source) and abiotic (temperature and soil moisture) factors on microbiome composition were characterized in wild-caught *C. imicola* populations. Our results provided additional information on the *C. imicola* microbiome, and the impact of biotic and abiotic factors on microbiome composition under natural conditions. The microbiome of *C. imicola* showed low diversity taxa composition in both populations. Similar studies in other insects have also reported low diversity records in the microbiome (Douglas, 2011; Broderick and Lemaitre, 2012; Jones *et al*., 2013; Charan *et al*., 2016), and high relative abundance of selected microorganisms (Osei-Poku *et al*., 2012; Minard *et al*., 2015). However, site/sample comparisons of virus presence may be affected by the sampling methodology used, which is suitable for microbiome studies but sample storage at room temperature and in ethanol is not suitable for virome studies and the unbiased detection of RNA viruses.

It has been demonstrated that geographical adaptation of insects is strongly associated to the microbiome composition (Minard *et al.*, 2013), and the microbiome might respond differently to abiotic and biotic challenges (Engel and Moran, 2013; Minard *et al.*, 2015). Recently, Minard *et al.* (2015) reported a low diversity in the gut microbiota of invasive mosquito species colonizing new geographical areas. These results suggested that low diversity in *C. imicola* microbiome composition might reflect the recent expansion of these midges in the Mediterranean area (Jacquet *et al.*, 2016).

Despite the low diversity observed in the *C. imicola* microbiome, a shared core microbiome was identified that might play a role in insect adaptation to new geographical areas (Crotti *et al.*, 2010; Gusmão *et al.*, 2010). The shared core microbiome is the result of a positive selection over the constant invasion of transient microorganisms (Minard *et al.*, 2015). How insects select and acquire microorganisms that become part of a semi-stable microbiota is not well understood. However, the acquisition of bacteria is strongly linked to the bacterial load in the adult bloodmeal sources and microorganisms they are exposed to in the breeding sites (Broderick *et al.*, 2004; Chandler *et al.*, 2012; Schauer *et al.*, 2012; Jones *et al.*, 2013; Aharon *et al.*, 2013). Based on these results, we can speculate that the shared core microbiome community found in our study might be explained by similarities in the *C. imicola* habitat found in both locations, which may lead to the acquisition of similar microorganisms (Mellor *et al.*, 2000).

The shared core microbiome identified in *C. imicola* from both locations was predominantly composed by *Proteobacteria* and *Actinobacteria*, which have been previously reported in mosquitoes and diptera (Seitz *et al.*, 1987; Maudlin *et al.*, 1990; Vasanthi and Hoti, 1992; Dermaio *et al.*, 1996; Fouda *et al.*, 2001). In *C. imicola*, the relative abundance of *Pseudomonas* was significantly higher than other microorganisms and was present in the shared core microbiome with differences between both populations for *P. putida*. *Pseudomonas* is a common clade ubiquitous in *Culicoides* breeding sites but also found as a gut commensal (Parker *et al.*, 1977; Campbell *et al.*, 2004; Erham, 2016). *Pseudomonas* have been widely associated with water and humid environments, the preferred sites for *Culicoides* during larval development, where these bacteria are likely acquired and maintained in the gut until adults completely develop. Recently, Chavshin *et al.* (2015) showed the persistence of

Pseudomonas during the maturation of *Anopheles stephensi* under laboratory conditions. They observed that *Pseudomonas* is able to colonize the Malpighian tubules and persist during larval stage to adults. The members of this clade have been attributed with a great variety of benefits for the insect, as detoxifiers of polluted environments, protection of eggs against other bacteria, insect growth and habitat adaptation, and blood digestion in blood-feeding insects (Peck and Walton, 2006; Lam *et al.*, 2007; Alvarez *et al.*, 2012; Senderovich and Halpern, 2013; Chavshin *et al.*, 2015). Our results support a role for *Pseudomonas* in *C. imicola* as part of the shared core microbiome. However, the difference in the relative abundance of *Pseudomonas* between *Culicoides* populations collected in Collesano and Trapani may correlate with higher soil moisture (as indicated by an increased NDVI) registered in Collesano across all seasons.

Other microorganisms present in the *C. imicola* shared core microbiome have been identified in other insects with functional implications. *Candidatus* Zinderia, a symbiont commonly associated with the sap-feeding insect Arizona spittlebug (*Clastoptera arizonara*) that is able to synthesize three specific amino acids, tryptophan, methionine and histidine (McCutcheon and Moran, 2010) has been never reported before in *Culicoides* spp. *Propionibacterium* was previously identified in *C. sonorensis* Wirth and Jones and *C. variipenis* (Coquillette) (Campbell *et al.*, 2004). *Culicoides* can acquire *Propionobacterium* during blood feeding as it is a common commensal found in mammalian skin that synthesize attractive substances for mosquitoes, which explains the high infection rate by this bacterium (Verhulst *et al.*, 2010). Insect associations with yeast are very specific and play a symbiotic role (Vega and Dowd, 2005). Yeast has been frequently isolated from the external parts of mosquitoes and non-biting Diptera such as *Drosophila* (Frants and Mertvetsova, 1986; Coluccio *et al.*, 2008), but none of these associations have been previously reported in *C. imicola* in which *Schizosaccharomyces* was identified in the present study. Most of the insect-yeast symbiotic relationships are nutritional, as yeast provide specific enzymatic machinery components for digestion, and synthesis of amino acids, vitamins and sterols. In turn, yeasts are maintained in an adequate environment, transported and dispersed (Morais *et al.*, 1994). Furthermore, insect-yeast associations exert an important effect on

shaping the microbial community (Jones *et al*., 2013), which could also explain the low diversity of bacterial taxa in the microbiome of both *C. imicola* populations.

In addition to the shared core microbiome, *C. imicola* also harbored a population-specific microbiome in this study. This population specific-microbiome could reflect the impact of abiotic and biotic factors defining the habitat of *C. imicola* that modulate the microbial composition of the environment, and the presence of symbiotic, pathogenic, or commensal microorganisms in the microbiome (Minard *et al*., 2013; 2015) (Additional file 1: Table S5). Furthermore, the analysis reported here was conducted with whole body samples, and therefore the identified microbiome may contain microorganisms associated with different insect tissues such as midgut, salivary gland and exoskeleton. Some of these microorganisms may be pathogenic in plants (i.e. *Potyvirus*) and animals (i.e. *Alpharetrovirus*) for which *C. imicola* may serve as vector. However, despite the outbreaks of bluetongue reported in Sicily (Torina *et al*., 2004), BTV was not identified in *C. imicola* from Collesano and Trapani. The relatively low proportion of BTV-infected *C. imicola* in vector populations may affect these results, together with the factors discussed above of sampling methodology and sample preservation that may influence virus detection. Nevertheless, additional studies are required to characterize the role of symbiotic, commensal and pathogenic microorganisms in the *C. imicola* microbiome.

Host preferences for *C. imicola* were characterized using two complementary experimental approaches for the identification of host DNA and proteins. The results of this study showed that *H. sapiens* and *C. lupus familiaris* were the main host blood meal sources for *C. imicola*, and these hosts were present in both farms. Taken together, the results of these methods for the characterization of host blood meal sources identified all hosts described in both farms except horses, which were present in low numbers. In the farm located in Collesano they raise cattle, goats and sheep, and horses and pigs are also present. In Trapani, the farm is devoted to sheep, but cattle and pigs are also present. Dogs are present in both farms, but are more abundant in the farm located in Trapani. *C. imicola* is considered to be a mammalophilic opportunist, with preference for livestock species (Meiswinkel *et al*., 2004; Lassen *et al*., 2012). In Senegal, *C. imicola* was found to have preference for horses when

compared to sheep, and to be mostly nocturnal with peak activity after sunset (Fall *et al.*, 2015). In Tunisia, *C. imicola* samples collected near human habitats and analyzed for *Cyt b* and *prepronociceptin* (*PNOC*) genes showed feeding preferences for humans, goats and sheep (Slama *et al.*, 2015).

Although the results using both methods supported that *H. sapiens* and *C. lupus familiaris* were the main host blood meal sources for *C. imicola*, differences between both methods were evident. In general, the host diversity identified in *C. imicola* was higher at the protein than DNA levels. These results could be explained be several factors such as the higher resolution of mass spectrometry when compared to PCR (Niare *et al.*, 2016), and the longer stability for proteins than DNA in blood-feeding arthropods (Martinez-de la Puente *et al.*, 2013; Villar *et al.*, 2016). These results supported the combination of PCR and proteomics-based methods for a better characterization of host blood meal sources in blood-feeding insects.

The results showed that NDVI (as a proxy for soil moisture) and the presence of humans as blood meal source for *C. imicola* constitute strong factors to explain the variations in the microbiome composition. NDVI is an indicator of photosynthetic activity and thus a proxy of soil moisture that favors the presence of certain microbial communities as we observed for *Pseudomonas,* showing higher relative abundance in *C. imicola* from Collesano. Other studies have found differences in the gut microbial communities of insects associated to blood meal and specific environmental conditions (Toft and Andersson, 2010; Morag *et al.*, 2012; Osei-Poku *et al.*, 2012; Nayduch *et al.*, 2015; Charan *et al.*, 2016), which supports the impact of biotic and abiotic factors on the microbiome.

Conclusions

These results expanded the information available on *C. imicola* microbiome composition, including the identification of eukaryotic microorganisms and viruses. A shared core microbiome was characterized in *C. imicola* wild-caught populations from Collesano and Trapani. Additionally, *C. imicola* also harbored a population specific-microbiome, which could reflect the impact of abiotic and biotic factors defining the habitat of *C. imicola* that modulate the microbiome composition. Further

studies should be directed to characterize the functional role of symbiotic, commensal and pathogenic microorganisms in the *C. imicola* microbiome, and their effect on vector competence for pathogens. The results of this study also showed the impact of biotic (*H. sapiens* blood meal source) and abiotic (NDVI as a proxy for soil moisture) factors on the *C. imicola* microbiome, which constitutes important information to characterize insect adaptation and predict geographic expansion of these vectors with potential implications for the control of vector-borne diseases.

Experimental Procedures

Study design and sample collection

Culicoides imicola adult nulliparous females were sampled during September-October 2013 in two localities from Sicily, Italy (Fig. 1). The first farm is located in Collesano, a small village of 108 Km^2 and around 4000 inhabitants at 266 m above sea level in Palermo province (latitude 37.94547, longitude 13.88009). In this farm they raise cattle, goats and sheep, and horses and pigs are also present. The second farm is located in Trapani province, Locogrande district, near the Trapani-Birgi Airport (latitude 37.898199, longitude 12.510562) at 26 m above sea level. Sheep are the main livestock host at this farm, but cattle and pigs are also present. Dogs are present in both farms, but are more abundant in the farm located in Trapani. Both farms are located in agricultural areas, including heterogeneous areas with annual crops associated to permanent crops with olive groves and vineyards, and near natural grassland. In both cases, sampling was carried out using ultraviolet (UV) suction traps of Onderstepoort type (ARC-Institute for Agricultural Engineering, South Africa) (Venter and Meiswinkel, 1994; Venter *et al*., 2009). Traps were hung at a height of 1.5 m above ground. Distance to stables and paddock with animals was less than 20 m and traps were active from sunset (07:00 PM) to sunrise (08:00 AM). The insects, attracted by the UV light, were sucked into plastic containers containing 200-300 ml of water in which few drops of detergent (Hederol, Procter and Gamble Professional, UK) were added to reduce surface

tension. Collected insects were placed in vials with 70% ethanol and stored at room temperature. *Culicoides imicola* specimens were identified and separated from other insects by using a stereomicroscope to characterize wing patterns as previously described (Rawlings, 1996). Five hundred specimens of *C. imicola* from different trap catches were selected from each location and five biological replicates with 100-pooled specimens each were used for further analysis. Two replicates were used for DNA analyses and the remaining three replicates for host protein identification.

Whole genome shotgun metagenomic sequencing and analysis

Genomic DNA was extracted from whole bodies of two biological replicates of pooled samples with 100 specimens each that were grinded and pulverized in liquid nitrogen and homogenized using a glass homogenizer (20 strokes) in 4 ml buffer (0.25 sucrose, 1 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.4) supplemented with 4% SDS and complete mini protease inhibitor cocktail (Roche, Basel, Switzerland). DNA was extracted using the NucleoSpin TriPrep kit (Macherey-Nagel, Duren, Germany) according to the manufacturer's instructions and quantified using PicoGreen (Invitrogen, Carlsbad, CA, USA) for template ranges of 4 ng to 100 ng. An aliquot of each pooled replicate was adjusted to the same DNA concentration and fragmented using a BioRuptor (Diagenode, Inc, NJ, USA). Libraries were prepared using the Ultra DNA library preparation kit (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's instructions. Samples were indexed using a unique combination of primers provided by the manufacturer. After library preparation, DNA library was purified in 1.5% agarose gels to select fragments with sizes within 150-400 bp, quantified with a Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and then pooled in equimolar concentrations. Finally, 2x100 bp pair-end sequencing was performed in the HiSeq2000 platform (Illumina, San Diego, CA, USA) using a 2x100 program according to the manufacturer's instructions. Obtained reads were demultiplexed and filtered by quality, which finally rendered 250 million pair-end reads passing filter per sample.

For assembly of the metagenomic reads, raw data files from shotgun sequencing were de-multiplexed, filtered by quality and converted into fastq using Casava v.1.8.2 (Illumina). High quality reads were processed using a metagenomic *de novo* assembly approach. First, a filtering analysis was performed in order to remove insect reads. Raw insect reads were mapped using Bowtie software (http://bowtie-bio.sourceforge.net/index.shtml) (Langmead *et al*., 2009), against *Chironomus tentans* genome (NCBI accession number SAMEA3158483, https://www.ncbi.nlm.nih.gov/assembly/GCA_000786525.1/) and *C. sonorensis* annotated reference transcriptome PRJNA238338 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA238338/) accessed in April, 2016. All the unmapped reads (61 to 69 million reads per sample representing 24-28% of the total; Additional file 1: Table S6) were assembled using SPADES with a minimum contig length of 200 bp (http://bioinf.spbau.ru/spades). The quality of genome assemblies corresponding to the *C. imicola* metagenome was further evaluated using QUAST (http://bioinf.spbau.ru/quast). All the assembly metrics are provided in the Additional file 1: Table S6. Metagenome relative abundance at different taxonomic levels was obtained taking the number of count reads or identifications (IDs) assigned to each identified sequence, and normalized against the total number of IDs (Hernández-Jarguín *et al*., 2018) using MetaPhlAn (Segata *et al*., 2012; Truong *et al*., 2015), and coverage analysis with SAM tools (http://samtools.sourceforge.net) (Additional file 2: Dataset S1). MetaPhlAn was used to assign sequences to particular taxa using default parameters that included BLASTn default e-value threshold of 1×10^{-6}, UCLUST nucleotide identity threshold of 75%, and rejected sequences as core genes if its homology pattern in the clade deviated from the expected baseline missanotation error rate with confidence greater than 95%. The same procedure was applied to all taxonomic levels (from phyla to genera to species). Cladograms were displayed using the PhyloT (http://phylot.biobyte.de) and iTOL (http://itol.embl.de) open source software platforms. Finally, the metagenome comparative analysis between *C. imicola* populations was performed using Metastats (http://metastats.cbcb.umd.edu/detection.html; $P \leq 0.05$) (White *et al*., 2009; Paulson *et al*., 2011).

Metagenome dataset validation by qPCR

The same *C. imicola* DNA samples used for DNA sequencing were also analyzed by qPCR to validate the metagenomic data. Specific primers for *Acinetobacter* spp., *Pseudomonas* spp. and *P. putida* genes were used as previously described (Spilker *et al*., 2004; Minard *et al*., 2013) (Additional file 1: Table S7). The iScript One-Step was used to perform the qPCR with SYBR Green and the iQ5 thermal cycler (Bio-Rad, Hercules, CA, USA) following manufacturer's recommendations. The qPCR SYBR Green reactions were carried out in 20 µl volume with 1 µl DNA (15-18 ng/ µl), 10 µM of each of forward and reverse primers, and 10 µl of One-Step SYBR reaction mix. PCR conditions are shown in the Additional file 1: Table S7. For negative controls, genomic DNA was replaced by molecular grade water. A dissociation curve was run at the end of the reaction to ensure that only one amplicon was formed and that the amplicons denatured consistently in the same temperature range for every sample (Ririe *et al*., 1997). DNA levels were normalized against *C. imicola* Elongation factor 1b gene (*Ef1 b*), following the conditions reported previously by Anbazhagan *et al*. (2011). Normalization was performed using the ddCT method as implemented by Bio-Rad iQ5 Standard Edition, Version 2.0 (Pfaffl 2001; Livak and Schmittgen, 2001). Normalized Ct values for *C. imicola* collected in Collesano and Trapani were compared by Student's t-test for samples with unequal variance ($P \leq 0.05$).

Biotic factors: Characterization of *C. imicola* host preferences

Two experimental approaches were used to characterize the potential blood meal sources of collected *C. imicola* populations: (a) host-DNA identification by PCR and (b) host-protein identification by proteomics.

(a) Host-DNA identification by PCR: The same *C. imicola* DNA samples used for DNA sequencing were used for host-DNA identification by PCR. The PCR and sequence analysis of *Cyt b* gene was done as previously described (Slama *et al*., 2015). The PCR was performed in a volume of 50 µl using 1 µl of each primer *cyt b* 1: 5'-CCA TCC AAC ATC TCA GCA TGA TGA AA-3' and *cyt b* 2: 5'-GCC CCT CAG AAT GAT ATT TGT CCT CA-3' and the 5PRIME Hot Master Mix

(www.5Prime.com) following the manufacturer's instructions. The PCR conditions included a denaturation step at 95 °C for 10 min, 40 amplification cycles (94 °C for 30 sec, 52 °C for 30 sec, 72 °C for 45 sec) and a final incubation at 72 °C for 5 min. The PCR products corresponding to the *Cyt b* gene were cloned according to the manufacturers instructions using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA, USA) and sequenced by Secugen, Madrid, Spain. A total of 11 and 6 *Cytb b* nucleotide sequences from samples collected at Collesano and Trapani, respectively were edited using the SNAPgene viewer (http://www.snapgene.com/products/snapgene_viewer/; GSL Biotech, Chicago, IL, USA) and aligned to the GenBank DNA sequence database using the nucleotide-nucleotide Basic Local Alignment Search Tool (BLAST; https://blast.ncbi.nlm.nih.gov/Blast.cgi) to assign a vertebrate host species (Additional file 1: Table S2). Host species assignment was completed when a match of 95% or more was found between our sequences and those in the GenBank. The number of sequences assigned to each host was compared between the Collesano and Trapani populations by Chi^2-test ($P \leq 0.05$). All sequences with BLAST identity < 100% were deposited in the GenBank with accession numbers MG182861-MG182873 (Additional file 1: Table S2).

(b) Host-protein identification by proteomics: For protein extraction, three biological replicates of pooled samples with 100 specimens each were grinded and pulverized in liquid nitrogen and proteins were extracted using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen Inc., Valencia, CA, USA) according to manufacturers instructions. Precipitated proteins were resuspended in 20 mM Tris-HCl pH 7.5 with 4% SDS and protein concentration was determined using the BCA Protein Assay (Thermo Scientific, San Jose, CA, USA) using bovine serum albumin (BSA) as standard. Protein extracts (100 μg per sample) were on-gel concentrated by SDS-PAGE as previously described (Villar *et al.*, 2015). The unseparated protein bands were visualized by staining with GelCode Blue Stain Reagent (Thermo Scientific), excised, cut into 2 × 2 mm cubes and digested overnight at 37°C with 60 ng/μl sequencing grade trypsin (Promega, Madison, WI, USA) at 5:1 protein:trypsin (w/w) ratio in 50 mM ammonium bicarbonate, pH 8.8 containing 10% (v/v) acetonitrile (Shevchenko *et al.*, 2006). The resulting tryptic peptides from each band were extracted by 30 min incubation in 12 mM ammonium bicarbonate, pH 8.8. Trifluoroacetic acid was added to a final concentration of 1% and the

peptides were finally desalted onto OMIX Pipette tips C18 (Agilent Technologies, Santa Clara, CA, USA), dried-down and stored at -20°C until mass spectrometry (MS) analysis. The desalted protein digests was resuspended in 0.1% formic acid and analyzed by reverse phase-liquid chromatography MS/MS (RP-LC-MS/MS) using an Easy-nLC II system coupled to a linear ion trap (LTQ) mass spectrometer (Thermo Scientific). The peptides were concentrated (on-line) by reverse phase chromatography using a 0.1×20 mm C18 RP precolumn (Thermo Scientific), and then separated using a 0.075×100 mm C18 RP column (Thermo Scientific) operating at 0.3 ml/min. Peptides were eluted using a 120-min gradient from 5% to 40% solvent B in solvent A (solvent A: 0.1% formic acid in water, solvent B: 0.1% formic acid in acetonitrile). Electrospray ionization was done using a Fused-silica Pico-Tip Emitter ID 10 mm (New Objective, Woburn, MA, USA) interface. Peptides were detected in survey scans from 400 to 1600 amu (1 mscan), followed by fifteen data dependent MS/MS scans (Top 15), using an isolation width of 2 mass-to-charge ratio units, normalized collision energy of 35%, and dynamic exclusion applied during 30 sec periods. Three technical replicates per sample were analyzed. The MS/MS raw files were searched against a compiled database containing all the proteins for *B. taurus*, *O. aries*, *S. scrofa*, *C. lupus familiaris, H. sapiens*, *C. aegagrus hircus* and *Culicoides* spp. Uniprot (http://www.uniprot.org) entries (18348, 23112, 26104, 25491, 70946, 3101, and 1909 entries in March 2016, respectively) using the SEQUEST algorithm (Proteome Discoverer 1.4, Thermo Scientific). The following constraints were used for the searches: tryptic cleavage after Arg and Lys, up to two missed cleavage sites, and tolerances of 1 Da for precursor ions and 0.8 Da for MS/MS fragment ions and the searches were performed allowing optional Met oxidation and Cys carbamidomethylation. A false discovery rate (FDR) ≤ 0.01 was considered as condition for successful peptide assignments and considering that at least two peptides per protein were required for protein identification. Finally, a total of 494 proteins were assigned to vertebrate hosts and grouped by host species (Additional file 3: Dataset S2)*,* and the average number of Peptide Spectrum Matches (PSMs) for each species were normalized against the average number of PSMs of *Culicoides* spp. proteins ($n = 3$) and compared between the Collesano and Trapani populations by Chi^2-test ($P ≤ 0.05$).

Abiotic factors: Characterization of temperature and soil moisture in *C. imicola* collection sites. The study area is under a Mediterranean climate, mainly featured by hot and dry summers, warm and rainy winters. However, Trapani and Collesano showed climate differences in terms of temperature and soil moisture due to the island topography (Additional file 1: Table S3). Climate data included in models were the averaged land surface temperature (LSTD) and the normalized difference in vegetation index (NDVI), a proxy of photosynthetic activity of the plant canopy, which relates to the type of vegetation and soil moisture (Carlson *et al*., 2009). Both datasets were obtained at a resolution of 0.05º from the MODIS web site (https://modis.gsfc.nasa.gov/data/dataprod/) (Remer *et al*., 2005) and compared between locations by analysis of variance (ANOVA test, $P \leq 0.05$) (Estrada-Peña and de la Fuente, 2016). Other abiotic factors included were the land use and land cover, using the standard European classifications of CORINE3 (https://www.eea.europa.eu/data-and-maps/data/clc-2000-vector-6).

Characterization of the impact of biotic and abiotic factors on *C. imicola* microbiome.

The combined effect of biotic and abiotic factors on the *C. imicola* microbiome was characterized using principal component analysis (PCA) (Additional file 1: Table S4). Linear combinations of the biotic and abiotic variables were used to build the PCA to finally get the new variable that best explains variations in the microbiome. The Statistical Package for the Social Science (SPSS; IBM Analytics, Armonk, NY, USA; https://www.ibm.com/analytics/us/en/) was used for analysis including the biotic and abiotic variables and the microbiome components detected at the species level at a relative abundance higher than 1% (Additional file 1: Table S4).

Availability of data.

Raw metagenomic reads for the *C. imicola* populations collected at Collesano and Trapani are deposited in the [Dryad repository pending of article acceptance]. Proteomics data on host blood meal identification are deposited in the [Dryad repository pending of article acceptance].

Acknowledgements

References

Acevedo, P., Ruiz-Fons, F., Estrada, R., Márquez, A.L., Miranda, M.A., Gortázar, C. and Lucientes, J. (2010) A broad assessment of factors determining *Culicoides imicola* abundance: Modelling the present and forecasting its future in climate change scenarios. *PLoS One* 12: e14236.

Aharon, Y., Pasternak, Z., Ben Yosef, M., Behar, A., Lauzon, C., Yuval, B. and Jurkevitch, E. (2013) Phylogenetic, metabolic, and taxonomic diversities shape Mediterranean fruit fly microbiotas during ontogeny. *Appl Environ Microbiol* 79: 303-313.

Alvarez, C., Kukutla, P., Jiang, J., Yu, W. and Xu, J. (2012) Draft genome sequence of *Pseudomonas* sp. strain Ag1 isolated from the midgut of the malaria mosquito *Anopheles gambiae*. *J Bacteriol* 194: 5449.

Anbazhagan, D., Mui, W.S., Mansor, M., Siok Yan, G.O., Yusof, M. and Sekaran, S.D. (2011) Development of conventional and real-time multiplex PCR assays for the detection of nosocomial pathogens. *Braz J Microbiol* 42: 448-458.

Benelli, G., Buttazzoni, L., Canale, A., D'Andrea, A., Del Serrone, P., Delrio, G., Foxi, C., Mariani, S., Savini, G., Vadivalagan, C., Murugan, K., Toniolo, C., Nicoletti, M. and Serafini, M. (2017) Bluetongue outbreaks: Looking for effective control strategies against *Culicoides* vectors. *Res Vet Sci* 115: 263-270.

Broderick, N.A. and Lemaitre, B. (2012) Gut-associated microbes of *Drosophila melanogaster*. *Gut*

Microbes 3: 307-321.

Broderick, N.A., Raffa, K.F., Goodman, R.M. and Handelsman, J. (2004) Census of the bacterial community of the gypsy moth larval midgut by using culturing and culture independent methods. *Appl Environ Microbiol* 70: 293-300.

Campbell, C.L., Mummey, D.L., Schmidtmann, T. and Wilson, W.C. (2004) Culture-independent analysis of midgut microbiota in the arbovirus vector *Culicoides sonorensis* (*Diptera*: *Ceratopogonidae*). *J Med Entomol* 41: 340-348.

Carlson, T.N., Gillies, R.R. and Perry, E.M. (2009) A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover. *Remote Sensing Reviews* 9: 161-173.

Carpenter, S., Grouschup, M.H., Garros, C., Felippe-Bauer, M.L. and Purse, B.V. (2013) *Culicoides* biting midges, arboviruses and public health in Europe. *Antivir Res*,100: 102-113.

Chandler, J.A., Eisen, J.A. and Kopp, A. (2012) Yeast communities of diverse *Drosophila* species: Comparison of two symbiont groups in the same hosts. *Appl Environ Microbiol* 78: 7327-7336.

Charan, S.S., Pawar, K.D., Gavhale, S.D., Tickhe, C.V., Charan, N.S., Angel, B., Joshi, V., Patole, M.S. and Shouche, Y.S. (2016) Comparative analysis of midgut bacterial communities in three aedine mosquito species from dengue-endemic and non-endemic areas of Rajastgan, India. *Med Vet Entomol* 30: 264-277.

Chavshin, A.E., Oshaghi, M.A., Vatandoost, H., Yakhchali, B., Zarenejad, F. and Terenius, O. (2015) Malphigian tubules are important determinants of *Pseudomona*s transstadial transmission and longtime persistence in *Anopheles stephensi. Parasite Vectors* 8: 36.

Coluccio, A.E., Rodriguez, R.K., Kernan, M.J. and Neiman, A.M. (2008) The yeast spore wall enables spores to survive passage through the digestive tract of *Drosophila. PLoS One* 3: e2873.

Conte, A., Goffredo, M., Ippoliti, C. and Meiswinkel, R. (2007) Influence of biotic and abiotic factors on the distribution and abundance of *Culicoides imicola* and the Obsoletus Complex in Italy. *Vet*

Parasitol 150: 333-344.

Crotti, E., Rizzi, A., Chouaia, B., Ricci, I., Favia, G., Alma, A., Sacchi, L., Bourtzis, K., Mandrioli, M., Cherif, A., Bandi, C. and Daffonchio D. (2010) Acetic acid bacteria, newly emerging symbionts of insects. *Appl Environ Microbiol* 76: 6963-6970.

Demaio, J., Pumpuni, C.B., Kent, M. and Beier, J.C. (1996) The midgut bacterial flora of wild *Aedes triseriatus, Culex pipiens,* and *Psorophora columbiae* mosquitoes. *Am J Trop Med Hyg* 54: 219-223.

Douglas, A.E. (2011) Lessons from studying insect symbioses. *Cell Host Microbe* 10: 359-367.

Douglas, A.E. (2015) Multiorganismal insects: Diversity and function of resident microorganisms. *Ann Rev Entomol* 60: 17-34.

Engel, P. and Moran, N.A. (2013) The gut microbiota of insects-diversity in structure and function. *FEMS Microbiol Rev* 37: 699-735.

Erham, D. (2016) Characterization of the larval habitat of *Culicoides sonorensis* (Diptera: Ceratopogonidae) with emphasis on the significance of animal manure and the associated bacterial community. Thesis Dissertation. Kansas State University, Department of Entomology, College of Agriculture.

Estrada-Peña, A. and de la Fuente, J. (2016) Species interactions in occurrence data for a community of tick-transmitted pathogens. *Sci Data* 3: 160056.

Fall, M., Fall, A.G., Seck, M.T., Bouyer, J., Diarra, M., Lancelot, R., Gimonneau, G., Garros, C., Bakhoum, M.T., Faye, O., Baldet, T. and Balenghien T. (2015) Host preferences and circadian rhythm of *Culicoides* (Diptera: Ceratopogonidae), vectors of African horse sickness and bluetongue viruses in Senegal. *Acta Trop* 149: 239-245.

Fouda, M.A., Hassan, M.I., Al-Daly, A.G. and Hammad, K.M. (2001) Effect of midgut bacteria of *Culex pipiens* L. on digestion and reproduction. *J Egypt Soc Parasitol* 31: 767-780.

Frants, T.G. and Mertvetsova, O.A. (1986) Yeast associations with mosquitoes of the genus *Aedes* Mg. (Diptera, Culicidae) in the Tom-Ob river region. *Nauchnye Doki Vyss Shkoly Biol Nauki* 4: 94-

98.

Franzenburg, S., Walter, J., Künzel, S., Wang, J., Baines, J.F., Bosch, T.C. and Fraune, S. (2013) Distinct antimicrobial peptide expression determines host species-specific bacterial associations. *Proc Natl Acad Sci U.S.A.* 110: 3730-3738.

Guichard, S., Guis, H., Tran, A., Garros, C., Balenghien, T. and Kriticos, D.J. (2014) Worldwide niche and future potential distribution of *Culicoides imicola*, a major vector of bluetongue and African horse sickness viruses. *PLoS One* 12: e112491.

Gusmão, D.S., Santos, A.V., Marini, D.C., Bacci, M. Jr., Berbert-Molina, M.A. and Lemos, F.J.A. (2010) Culture-dependent and culture-independent characterization of microorganisms associated with *Aedes aegypti* (*Diptera: Culicidae*) (L.) and dynamics of bacterial colonization in the midgut. *Acta Trop* 115: 275-281.

Hernández-Jarguín, A., Díaz-Sánchez, S., Villar, M. and de la Fuente, J. (2018) Integrated metatranscriptomics and metaproteomics for the characterization of bacterial microbiota in unfed *Ixodes ricinus*. *Ticks Tick Borne Dis* pii: S1877-959X(18)30034-7. doi: 10.1016/j.ttbdis.2018.04.020.

Jacquet, S., Garros, C., Lombaert, E., Walton, C., Restrepo, J., Allene, X., Baldet, T., Cetre-Sossah, C., Chaskopoulou, A., Delecolle, J.C., Desvars, A., Djerbal, M., Fall, M., Gardes, L., de Garine-Wichatitsky, M., Goffredo, M., Gottlieb, Y., Gueye Fall, A., Kasina, M., Labuschagne, K., Lhor, Y., Lucientes, J., Martin, T., Mathieu, B., Miranda, M., Pages, N., Pereira da Fonseca, I., Ramilo, D.W., Segard, A., Setier-Rio, M.L., Stachurski, F., Tabbabi, A., Talla Seck, M., Venter, G., Zimba, M., Balenghien, T., Guis, H., Chevillon, C., Bouyer, J. and Huber, K. (2015) Colonization of the Mediterranean basin by the vector biting midge species *Culicoides imicola*: an old story. *Mol Ecol* 24: 5707-5725.

Jacquet, S., Huber, K., Pagès, N., Talavera, S., Burgin, L.E., Carpenter, S., Sanders, C., Dicko, A.H., Djerbal, M., Goffredo, M., Lhor, Y., Lucientes, J., Miranda-Chueca, M.A., Pereira Da Fonseca, I., Ramilo, D.W., Setier-Rio, M.L., Bouyer, J., Chevillon, C., Balenghien, T., Guis, H. and Garros, C. (2016) Range expansion of the Bluetongue vector, *Culicoides imicola*, in continental France likely

due to rare wind-transport events. *Sci Rep* 6: 27247.

Jones, R.T., Gonzales-Sanchez, L. and Fierer, N. (2013) A cross-taxon analysis of insect-associated bacterial diversity. *PLoS One* 8: e61218.

Lam, K., Babor, D., Duthie, B.E., Babor, M., Moore, M. and Gries, G. (2007) Proliferating bacterial symbionts on house fly eggs affect oviposition behaviour of adult flies. *Animal Behav* 74: 81-92.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: 25.

Lassen, S.B., Nielsen, S.A. and Kristensen, M. (2012) Identity and diversity of blood meal hosts of biting midges (Diptera: Ceratopogonidae: *Culicoides* Latreille) in Denmark. *Parasite Vectors* 5: 143.

Lewis, S.E., Rice, A., Hurst, G.D.D. and Baylis, M. (2014) First detection of endosymbiotic bacteria in biting midges *Culicoides pulicaris* and *Culicoides punctatus*, important Palaeartic vectors of bluetongue virus. *Med Vet Entomol* 28: 453-456.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta}CT$ method. *Methods* 25: 402-408.

Martinez-de la Puente, J., Ruiz, S., Soriguer, R. and Figuerola, J. (2013) Effect of blood meal digestion and DNA extraction protocol on the success of blood meal source determination in the malaria vector *Anopheles atroparvus*. *Malaria J* 12: 109.

Maudlin, I., Welburn, S.C. and Mehlitz, D. (1990) The relationship between rickettsia-like-organisms and trypanosome infections in natural populations of tsetse in Liberia. *Trop Med Parasitol* 41: 265-277.

McCutcheon, J.P. and Moran, N.A. (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol Evol* 2: 708-718.

Meiswinkel, R., Venter, G.J. and Nevill, E.M. (2004) Vectors: *Culicoides* spp. *Infectious Diseases of Livestock* (eds. J.A.W. Coetzer & R.C. Tustin) pp. 93-136. Oxford University Press, Cape Town.

Mellor, P.S., Boorman, J. and Baylis, M. (2000) *Culicoides* biting midges: their role as arbovirus

vectors. *Ann Rev Entomol* 45: 307-340.

Minard, G., Tran, F.H., Raharimalala, F.N., Hellard, E., Ravelonandro, P., Mavingui, P. and Valiente Moro, C. (2013) Prevalence, genomic and metabolic profiles of Acinetobacter and Asaia associated with field-caught *Aedes albopictus* from Madagascar. *FEMS Microbiol Ecol* 83: 63-73.

Minard, G., Tran, F.H., Van, V.T., Goubert, C., Bellet, C., Lambert, G., Kim, K.L., Thuy, T.H., Mavingui, P. and Valiente Moro, C. (2015) French invasive Asian tiger mosquito populations harbor reduced bacterial microbiota and genetic diversity compared to Vietnamese autochthonous relatives. *Front Microbiol* 6: 970.

Morag, N., Klement, E., Saroya, Y., Lensky, I. and Gottlieb, Y. (2012) Prevalence of the symbiont *Cardinium* in *Culicoides* (*Diptera: Ceratopogonidae*) vector species is associated with land surface temperature. *FASEB J* 26: 4025-4034.

Morais, P.B., Rosa, C.A., Hagler, A.N. and Mendonça-Hagler, L.C. (1994) Yeast communities of the cactus *Pilosocereus arrabidae* as resources for larval and adult stages of *Drosophila serido*. *Antonie van Leeuwenhoek* 66: 313-317.

Nayduch, D., Erram, D., Lee, M.B., Zurek, L. and Saski, C.A. (2015) Impact of the blood meal on humoral immunity and microbiota in the gut of female *Culicoides sonorensis*. *Vet Ital* 51: 385-392.

Niare, S., Berenger, J.M., Dieme, C., Doumbo, O., Raoult, D., Parola, P. and Almeras, L. (2016) Identification of blood-meal sources in the main African malaria mosquito vector by MALDI-TOF MS. *Malaria J* 15: 87.

Osei-Poku, J., Mbogo, C.M., Palmer, W.J. and Jiggins, F.M. (2012) Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. *Mol Ecol* 21: 5138-5150.

Pagès, N., Bréard, E., Urien, C., Talavera, S., Viarouge, C., Lorca-Oro, C., Jouneau, L., Charley, B., Zientara, S., Bensaid, A., Solanes, D., Pujols, J. and Schwartz-Cornil, I. (2014) *Culicoides* midge bites modulate the host response and impact on bluetongue virus infection in sheep. *PLoS One* 9: e83683.

Parker, M.D., Akey, D.H. and Lauerman, L.H. (1977) Microbial flora associated with colonized and

wild populations of the biting gnat *Culicoides variipennis*. *Entomol Exp Appl* 21: 130-136.

Paulson, J.N., Pop, M. and Corrada Bravo, H. (2011) Metastats: an improved statistical method for analysis of metagenomic data. *Genome Biol* 12: 17.

Peck, G.W. and Walton, W.E. (2006) Effect of bacterial quality and density on growth and whole body stoichiometry of *Culex quinquefasciatus* and *Culex tarsalis* (*Diptera: Culicidae*). *J Med Entomol* 43: 25-33.

Pfaffl MW. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29: e45.

Purse, B.V., Mellor, P.S., Rogers, D.J., Samuel, A.R., Mertens, P.P. and Baylis, M. (2005) Climate change and the recent emergence of bluetongue in Europe. *Nat Rev Microbiol* 3: 171-181.

Rawlings, P. (1996) A key, based on wing patterns of biting midges (Genus: *Culicoides* Latreille - Diptera: Ceratopogonidae) in the Iberian Peninsula, for use in epidemiological studies. *Graellsia* 52: 57-71.

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A. and Martins, J. V. (2005) The MODIS aerosol algorithm, products, and validation *J Atmos Sci* 62: 947-973.

Ririe, K.M., Rasmussen, R.P. and Wittwer, C.T. (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem* 245: 154-160.

Schauer, C., Thompson, C.L. and Brune, A. (2012) The bacterial community in the gut of the cockroach *Shelfordella lateralis* reflects the close evolutionary relatedness of cockroaches and termites. *Appl Environ Microbiol* 78: 2758-2767.

Segata, N., Waldron, L., Ballarini, A., Narasimham, V., Jousson, O. and Huttenhower, C. (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 9: 811-814.

Seitz, H.M., Maier, W.A., Rottok, M. and Becker-Feldmann, H. (1987) Concomitant infections of *Anopheles stephensi* with *Plasmodium berghei* and *Serratia marcescens*: additive detrimental effects. *Zentralbl Bakteriol Mikrobiol Hyg* 266: 155-166.

Senderovich, Y. and Halpern, M. (2013) The protective role of endogenous bacterial communities in chironomid egg masses and larvae. *ISME J* 7: 2147-2158.

Slama, D., Haouas, N., Mezhoud, H., Babba, H. and Chaker, E. (2015) Blood meal analysis of *Culicoides* (Diptera: Ceratopogonidae) in Central Tunisia. *PLoS One* 10: e0120528.

Spilker, T., Coenye, T., Vandamme, P. and LiPuma, J.J. (2004) PCR-Based assay for differentiation of *Pseudomonas aeruginosa* from other *Pseudomonas* species recovered from Cystic Fibrosis patients. *J Clin Microbiol* 42: 2074-2079.

Toft, C. and Andersson, S.G. (2010) Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat Rev Genet* 11: 465-475.

Torina, A., Caracappa, S., Mellor, P.S., Baylis, M. and Purse, B.V. (2004) Spatial distribution of bluetongue virus and its *Culicoides* vectors in Sicily. *Med Vet Entomol* 18: 81-89.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C. and Segata, N. (2015) MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 12: 902-903.

Vasanthi, V. and Hoti, S.L. (1992) Microbial flora in gut of *Culex quinquefasciatus* breeding in cess pits. *Southeast Asian J Trop Med Public Health* 23: 312-317.

Vega, F.E. and Dowd, M. (2005) Insect-Fungal Associations. Chapter 9. The role of yeasts as insect endosymbiont. In: Ecology and Evolution. Oxford University Press, 211-243.

Venter, G.J. and Meiswinkel, R. (1994) The virtual absence of *Culicoides imicola* (*Diptera: Ceratopogonidae*) in a light-trap survey of the colder, high-lying area of the eastern Orange Free State, South Africa, and implications for the transmission of arboviruses. *Onderstepoort J Vet Res* 61: 327-340.

Venter, G.J., Koekemoer, J.J. and Paweska, J.T. (2006) Investigations on outbreaks of African horse sickness in the surveillance zone of South Africa. *Rev Sci Tech* 25: 1097-1119.

Venter, G.J., Labushagne, K., Hermanides, K.G., Boikanyo, S.N.B., Majatladi, D.M. and Morey, L. (2009) Comparison of the efficiency of five suction light traps under field conditions in South Africa for the collection of *Culicodes* species. *Vet Parasitol* 166: 299-307.

Verhulst, N.O., Takken, W., Dicke, M., Schraa, G. and Smallegange, R.C. (2010) Chemical ecology of interactions between human skin microbiota and mosquitoes. *FEMS Microbiol Ecol* 74: 1-9.

Villar, M., Ayllón, N., Alberdi, P., Moreno, A., Moreno, M., Tobes, R., Mateos-Hernández, L., Weisheit, S., Bell-Sakyi, L. and de la Fuente, J. (2015) Integrated metabolomics, transcriptomics and proteomics identifies metabolic pathways affected by *Anaplasma phagocytophilum* infection in tick cells. *Mol Cell Proteomics* 14: 3154-3172.

Villar, M., López, V., Ayllón, N., Cabezas-Cruz, A., López, J.A., Vázquez, J., Alberdi, P. and de la Fuente, J. (2016) The intracellular bacterium *Anaplasma phagocytophilum* selectively manipulates the levels of vertebrate host proteins in the tick vector *Ixodes scapularis*. *Parasite Vectors* 9: 467.

Villegas, L.M. and Pimenta, P.F. (2014) Metagenomics, paratrangenesis and the *Anopheles* microbiome: a portrait of the geographical distribution of the anophelina based on meta-analysis of reported taxa. *Mem Inst Oswaldo Cruz* 109: 672-684.

White, J.R., Nagarajan, N. and Pop, M. (2009) Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* 5: e1000352.

Zouache, K., Raharimalala, F.N., Raquin, V., Tran-Van, V., Raveloson, L.H., Ravelonandro, P. and mavingui, P. (2011) Bacterial diversity of field-caught mosquitoes, *Aedes albopictus* and *Aedes aegypti*, from different geographic regions of Madagascar. *FEMS Microbiol Ecol* 75: 377-389.

Figure legends

Figure 1. Experimental design and workflow followed for the characterization of the impact of biotic and abiotic factors on the microbiome composition of *C. imicola* collected in Collesano and Trapani. Maps were constructed using the Esri ArcMap 9.3 software. The localization of the study areas in Trapani and Collesano, Sicily is shown. The digital elevation model was processed through the interpolation of level curves values of the Sicilian region, obtaining the elevations of study sites. The land use of the areas near to the farms was obtained from Corine Land Cover 2006 processed by the European Environmental Agency describing the coverage and, in part, the use of the soil in Europe. Spatial selection allowed deriving the different levels of the land use classes that affect the areas where the farms are placed. Land use classes: 1.1, urban fabric; 2.1.1, non-irrigated arable land; 2.2.1, vineyards; 2.2.2, fruit trees and berry plantations; 2.4.3, land principally occupied by agriculture; 3.2.1, natural grassland; 3.2.3, Sclerophyllous vegetation.

Figure 2. Relative abundance of microorganisms in *C. imicola* populations collected from Trapani and Collesano by (A) phylum, (B) family and (C) genus. Both populations were characterized by a microbiome composed of Virus, Eukaryota and Bacteria. Metagenome relative abundance at different taxonomic levels was obtained taking the number of count reads or identifications (IDs) assigned to each identified sequence, and normalized against the total number of IDs using MetaPhlAn. The mean between the two replicates was use to calculate the average relative abundance shown in the figure.

Figure 3. Venn diagram showing the shared core metagenome and the specific metagenome of *C. imicola* collected in Collesano and Trapani. The metagenome comparative analysis between *C. imicola* populations was performed using Metastats (http://metastats.cbcb.umd.edu/detection.html; *P* = 0.05). Taxa within the shared core metagenome that are significantly abundant are indicated (*$P < 0.05$).

Figure 4. Phylogenic tree and heatmap showing the relative abundance of *C. imicola* microbiome at species level on each location. Cladograms were displayed using PhyloT (http://phylot.biobyte.de) and iTOL (http://itol.embl.de), and the metagenome comparative analysis between *C. imicola* populations was performed using Metastats (http://metastats.cbcb.umd.edu/detection.html; *$P < 0.05$).

Figure 5. Distribution of host bloodmeal sources in *C. imicola* collected from Collesano and Trapani. Data was obtained by two complimentary experimental approaches, host-DNA identification by PCR and sequencing of the *Cyt b* gene and host-protein identification by proteomics.

Table 1. PCA to characterize the influence of the combined effect of abiotic (temperature and soil moisture) and biotic (host DNA) factors on *C. imicola* microbiome.

Variable	Component matrix		
	PC1	PC2	PC3
LSTD (temperature)	-0.995	-0.097	-0.017
NDVI (soil moisture)	0.995*	0.097	0.017
H. sapiens blood source	0.995*	0.097	0.017
C. lupus familiaris blood source	-0.995	-0.097	-0.017
C. aegagrus hircus blood source	-0.995	-0.097	-0.017
Propionibacterium acnes	-0.508	0.758	0.410
Bacillus cereus thurigiensis	0.987	0.156	-0.043
Bacillus megaterium	0.994	0.111	0.003
Bacillus mojavensis	0.840	-0.332	0.43
Bacillus pumilus	0.920	0.324	-0.22
Bacillus subtilis	0.993	0.117	-0.003
Candidatus Zinderia insecticola	0.167	0.953	-0.255
Aeromonas unclassified	-0.648	0.515	0.560
Enterobacter cloacae	-0.988	-0.141	-0.060
Escherichia coli	-0.814	0.385	0.434
Escherichia unclassified	0.565	-0.801	0.200
Pseudomonas putida	0.991	0.129	-0.043
Pseudomonas unclassified	-0.821	-0.481	-0.308
Propionibacterium phage P100D	0.621	-0.516	0.590
Musca hytrovirus	-0.995	-0.102	-0.021

Total of variance explained	PC1	PC2	PC3
Total eigenvalues	15.146	3.394	1.46
Percent variance	75.732	16.968	7.30
Percent cumulative variance	75.732	92.700	100.0

The table shows variable loadings on the first three principal components (PC) of the PCA to characterize the influence of the combined effect of abiotic (temperature and soil moisture) and biotic (host DNA) factors on *C. imicola* metagenome. Eigenvalues, percent variance and percent cumulative variance extracted from the full database for each principal component is shown. *Highest loading components.

Supporting Information

Additional file 1: Figure S1. Metagenome dataset validation by qPCR. (A) Bacterial DNA levels were determined by *Cyt b* qPCR and normalized against *Culicoides Ef1 b* and shown as average + S.D. normalized Ct values. Normalized Ct values for *C. imicola* collected in Collesano and Trapani were compared by Student's t-test for samples with unequal variance (*$P < 0.05$; $n = 2$ biological replicates). (B) The ratio between Collesano and Trapani *C. imicola* population values was calculated for qPCR and metagenomics data using the average normalized Ct values and percent relative abundance, respectively. Significant differences between both populations were obtained only for *P. putida* using both methods (*$P < 0.05$). **Figure S2.** Relative abundance of microorganisms in *C. imicola* populations collected from Trapani and Collesano by most abundant (A) phyla, (B) families and (C) genera. **Table S1.** Taxonomic composition of the microbiome in two different populations of *C. imicola*. **Table S2.** BLAST identity obtained for the host DNA sequenced clones of *Cyt b* from *C. imicola* collected in Collesano and Trapani. **Table S3.** Climate data of the averaged land surface temperature (LSTD) and the normalized difference vegetation index (NDVI), an indicator of

photosynthetic activity and thus a proxy of soil moisture. **Table S4.** Variables included in the Principal Component Analysis (PCA). **Table S5.** Characterization of selected microorganisms found in the *C. imicola* population-specific microbiome. **Table S6.** Assembly metrics for the *C. imicola* metagenome. **Table S7.** qPCR oligonucleotide primers and conditions used to validate the *C. imicola* metagenome dataset.

Additional file 2: Dataset S1. Metagenome relative abundance at different taxonomic levels was obtained using MetaPhlAn.

Additional file 3: Dataset S2. Proteomics analysis of vector and host derived proteins in *C. imicola* collected from Collesano and Trapani.

Collection of abiotic (temperature and humidity) factors

(A) TRAPANI (B) COLLESANO

Culicoides imicola females

DNA

PROTEIN

Whole genome shotgun metagenomics

Host-DNA identification by PCR and sequencing

Host-protein identification by proteomics

Metagenome dataset validation by real-time PCR

Host blood meal source characterization

Microbiome characterization

Principal component analysis for the characterization of the impact of biotic and abiotic factors on *C. imicola* microbiome

A — Phyla (Trapani, Collesano): Ascomycota, Actinobacteria, Proteobacteria, Viruses, Firmicutes. Average relative abundance (%)

B — Families (Trapani, Collesano): Retroviridae, Hytrosaviridae, Debaryomycetaceae, Xanthomonadaceae, Moraxellaceae, Aeromonadaceae, Comamonadaceae, Acetobacteraceae, Promicromonosporaceae, Dermabacteraceae, Brevibacteriaceae, Potyviridae, Schizosaccharomycetaceae, Corynebacteriaceae, Alcaligenaceae, Oxalobacteraceae, Burkholderiaceae, Halomonadaceae, Propionibacteriaceae, Enterobacteriaceae, Siphoviridae, Pseudomonadaceae, Bacillaceae. Average relative abundance (%)

C — Trapani Genera (Trapani): Other, Enterobacter, Escherichia, Propionibacterium, Pseudomonas, Muscavirus. Average relative abundance (%)

Collesano Genera (Collesano): Other, Propionibacterium, Escherichia, Siphoviridae, Pseudomonas, Bacillus. Average relative abundance (%)

Collesano

Trapani

Shared core microbiome

Bacillus
Corynebacterium
Pusilimonas
Burkholderia
Siphoviridae
Potyvirus

Pseudomonas*
Escherichia
Halomonas
Candidatus Zinderia
Propionibacterium
Schizosaccharomyces

Brevibacterium
Promicromonosporaceae
Brachybacterium
Acetobacter
Klebsiella
Pantoea
Comamonas
Aeromonas
Enterobacter
Acinetobacter
Stenotrophomonas
Debaryomycetaceae
Muscavirus
Alpharetrovirus

Collesano

Trapani

Host-DNA identification

H. sapiens (100%)

H. sapiens (17%)

O. aries (17%)

C. lupus familiaris (66%)

Host-protein identification

C. lupus familiaris (17%)

C. aegagrus hircus (0.2%)

O. aries (16%)

S. scrofa (10%)

B. taurus (19%)

H. sapiens (38%)

C. lupus familiaris (20%)

C. aegagrus hircus (0.2%)

O. aries (16%)

S. scrofa (10%)

B. taurus (17%)

H. sapiens (36%)