PALE-Blu work package 4, deliverable report (M54)

D4.4: Genetic structure maps and *Culicoides* microbiome composition (Partner 3, 6, 7, 12)

Background and deliverable description

One of the objective of the WP4 is to help disentangling the relative effects of environmental and anthropogenic factors in bluetongue virus (BTV) spread, by mapping interactions between the environment and *Culicoides* populations and their microbiomes.

Species distribution and population structure represent a complex interaction of ecological and evolutionary history, driven by abiotic conditions, species interactions, and dispersal ability. Whereas large data sets on *Culicoides* populations and regional distribution maps have been produced in the last decade, ecological factors determining distribution species in ecosystems alongside connectivity between populations and dispersal capacity remain poorly understood. Distribution, structure and dispersal of vectors are obvious key factors to understand the relative implication of *Culicoides* in BTV spread.

The project was then designed to collect samples and, after assessing the cryptic diversity within collected individuals, to establish genetic structure of *Culicoides* (task 4.3), but also to describe the microbiome composition of *Culicoides* populations (task 4.4). Genetic structure of *Culicoides* has been already used to infer colonization history of *Culicoides imicola* in the Mediterranean basin, concluding that the recent bluetongue emergence in the late 1990s were not linked to *C. imicola* colonization event, but rather to biological changes in the vector or the virus (Jacquet *et al.*, 2015).

The D4.4 activities were implemented as follows:

a. Collect *Culicoides obsoletus s.s.* and *C. imicola* samples from PALE-Blu partners and from collaborative networks;
b. Describe the cryptic diversity within the Obsoletus/Scoticus Complex, to properly identify individuals of the *Culicoides obsoletus s.s.* species;
c. Establish the genetic structure of *C. obsoletus s.s.* populations;
d. Describe the microbiome diversity (including bacteriome and virome) of *Culicoides* populations.

Publications related to this deliverable

- Mignotte A, Garros C, Gardès L, Balenghien T, Duhayon M, Rakotoarivony I, Tabourin L, Poujol L, Mathieu B, Ibañez-Justicia A, Deniz A, Cvetkovikj A, Purse BV, Ramilo DW, Stougiou D, Werner D, Pudar D, Petrić D, Veronesi E, Jacobs F, Kampen H, Pereira da Fonseca I, Lucientes J, Navarro J, de la Puente JM, Stefanovska J, Searle KR, Khallaayoune K, Culverwell CL, Larska M, Bourquia M, Goffredo M, Bisia M, England M, Robin M, Quaglia M, Miranda-Chueca MÁ, Bødker R, Estrada-Peña R, Carpenter S, Tchakarova S, Boutsini S, Sviland S, Schäfer SM, Ozoliņa Z, Segliņa Z, Vatansever Z, Huber K (2020). The tree that hides the forest: cryptic diversity and phylogenetic relationships in the Palaearctic vector Obsoletus/Scoticus Complex (Diptera: Ceratopogonidae) at the European level. Parasites & Vectors 13, 265, doi: 10.1186/s13071-020-04114-1.

- Aguilar-Vega C, Rivera B, Lucientes J, Gutiérrez-Boada I, Manuel Sánchez-Vizcaíno J (2021). A study of the composition of the Obsoletus complex and genetic diversity of *Culicoides obsoletus* populations in Spain. Parasites Vectors 14, 351: doi: 10.1186/s13071-021-04841-z.

Deliverable Activities

1. Cryptic diversity within the Obsoletus/Scoticus Complex

Culicoides obsoletus is an abundant and widely distributed Palearctic biting midge species, involved in the transmission of BTV and Schmallenberg virus (SBV) to wild and domestic ruminants. This vector species is reported jointly with morphologically very close species (especially for females): *C. scoticus* and *C. montanus*, forming the Obsoletus/Scoticus Complex. Recently, cryptic diversity within *C. obsoletus* species was reported in geographically distant sites. Clear delineation of species and characterization of genetic variability is mandatory to revise the taxonomic scheme and assess the vector role of each taxonomic entity.

The objectives were to characterize and to map the cryptic diversity within the Obsoletus/Scoticus Complex at the European level and along a North-South transect using a multi-marker methodology, and to revise the current taxonomic scheme using species delimitation models. Sequencing a portion of the COI mitochondrial gene of 3,789 individuals belonging to Obsoletus/Scoticus Complex from 20 countries by partner 3 revealed the presence, aside *C. obsoletus s.s.*, of at least three undescribed phylogenetic clades (*C. obsoletus* clade O2, *C. obsoletus* clade 'dark' and one not yet named and described) of *C. obsoletus s.l.* (Figure 1).

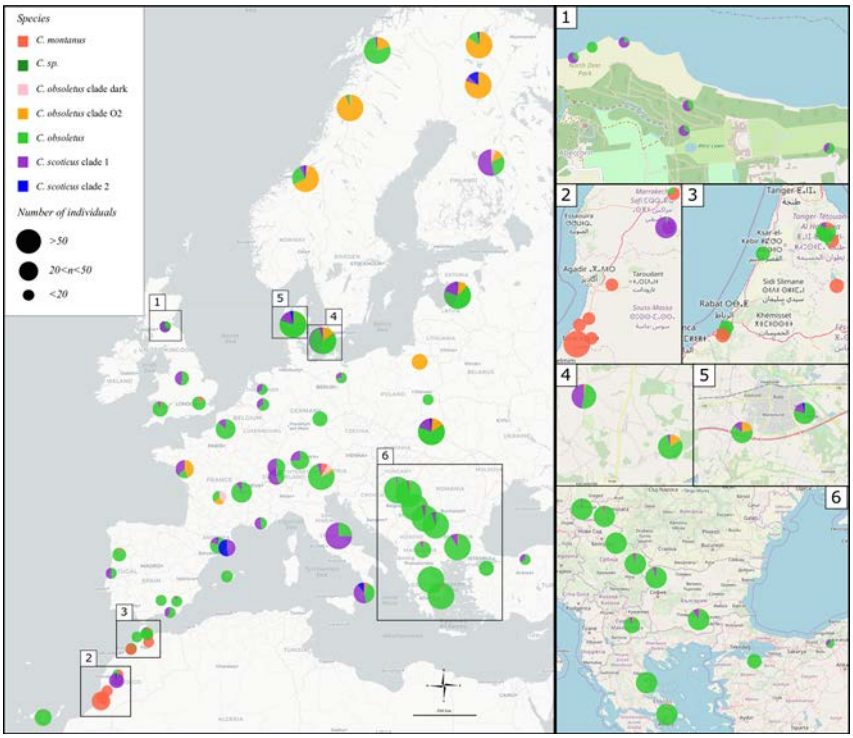

Figure 1. Relative distribution of inter-specific divergence and intra-specific variation in COI for all cryptic species of the Obsoletus/Scoticus Complex.

These analyses also revealed two clades within *C. scoticus* and raised questions about the taxonomic status of *C. montanus*. These results are supported by 16S rDNA mitochondrial gene sequences and a gene coding for ribosomal 28S rDNA, over the entire specific diversity resulting from COI barcoding (Figure 2). This provides for the first time a genetic characterization of the Obsoletus/Scoticus Complex on a large geographical scale and allows a revision of the current taxonomic classification for an important group of vector species of livestock viruses in the Palearctic region.

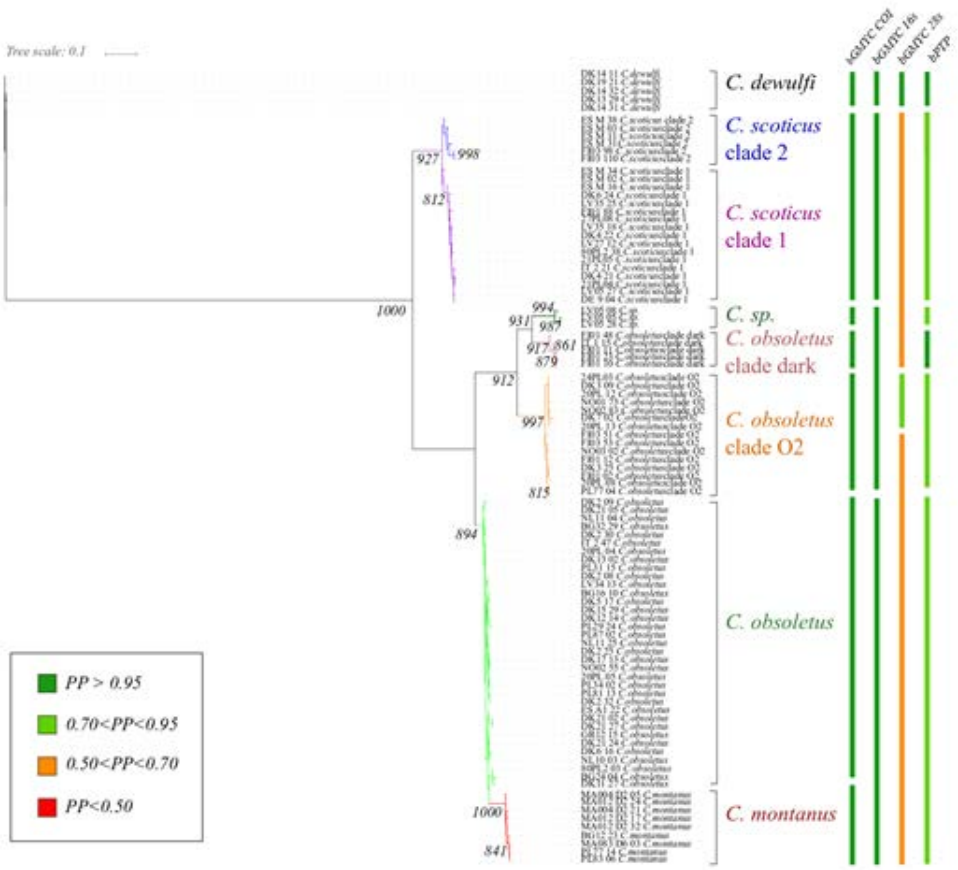

Figure 2. Maximum likelihood phylogenetic tree using concatenate genes (COI; rDNA16s, 28S rDNA) representing species delimitation and molecular relationships within the Obsoletus/Scoticus Complex.

2. Analysis of the diversity and genetic structure of *C. obsoletus s.s.* in Europe

A classical population genetic analysis was carried out by partner 3 on 931 individuals from 22 countries (1-2 sites per country), using 11 microsatellite markers. All of these samples were genotyped and sequenced. A weak genetic structure was observed during analyses, however with a slight structure observed between Mediterranean Basin and Northern Europe (Figure 3).

A focused population genetic analysis was carried out by partner 12 on populations from 12 locations in Spanish mainland, and from 1 location in the Canary Islands (Figure 4). A higher genetic diversity in mainland Spain than in the Canary Islands was found (Figure 5). The low genetic diversity, inexistent genetic differentiation among the structure of *C. obsoletus s.s.* in all Canary Islands populations, and strong divergence with the other sampling sites are probably a consequence of their isolation, which implies inbreeding and low gene flow outside the islands.

Altogether these results suggest a high dispersal capacity of *C. obsoletus s.s.* leading to an important gene flow between even distant populations, with the exception of island populations.

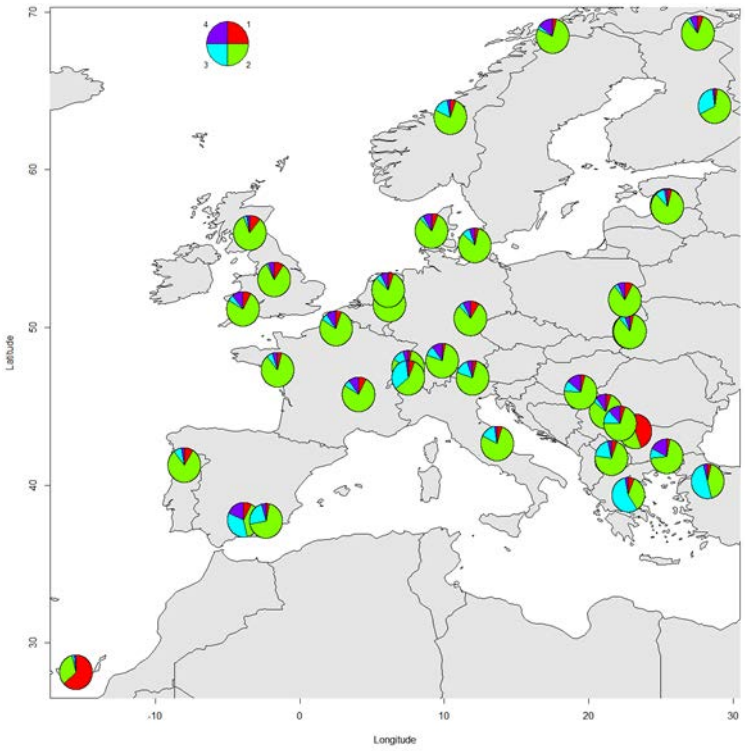

Figure 3. Distribution map of the populations of *Culicoides obsoletus s.s.* represented by their percentage of assignment to a genetic group.

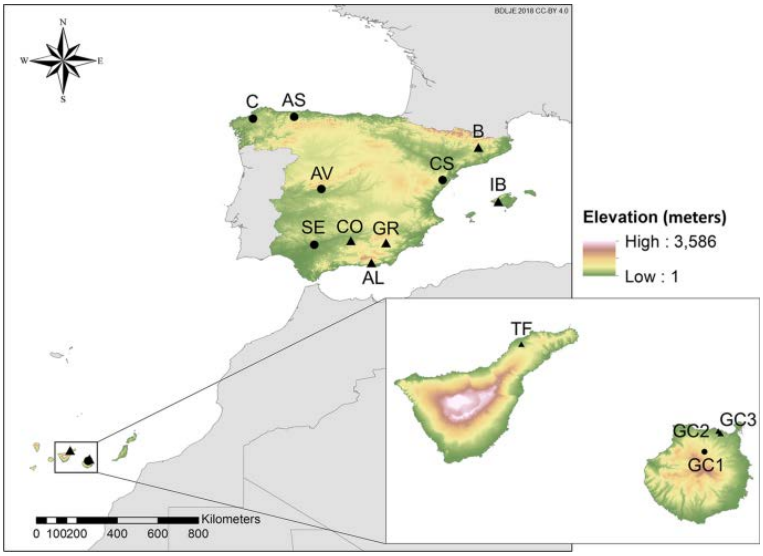

Figure 4. Location of the collection sites of the midges of the Obsoletus/Scoticus complex used in this study (black dots) and of the georeferenced sequences retrieved from GenBank (black triangles). Locations are coded based on the province's location: Barcelona (B), Mallorca (IB), Córdoba (CO), Granada (GR), Almería (AL), Tenerife (TF), Gran Canaria (GC2, GC3). Spanish administrative boundaries were provided by the Instituto Geográfico Nacional (ign.es) (BDDAE CC-BY 4.0).

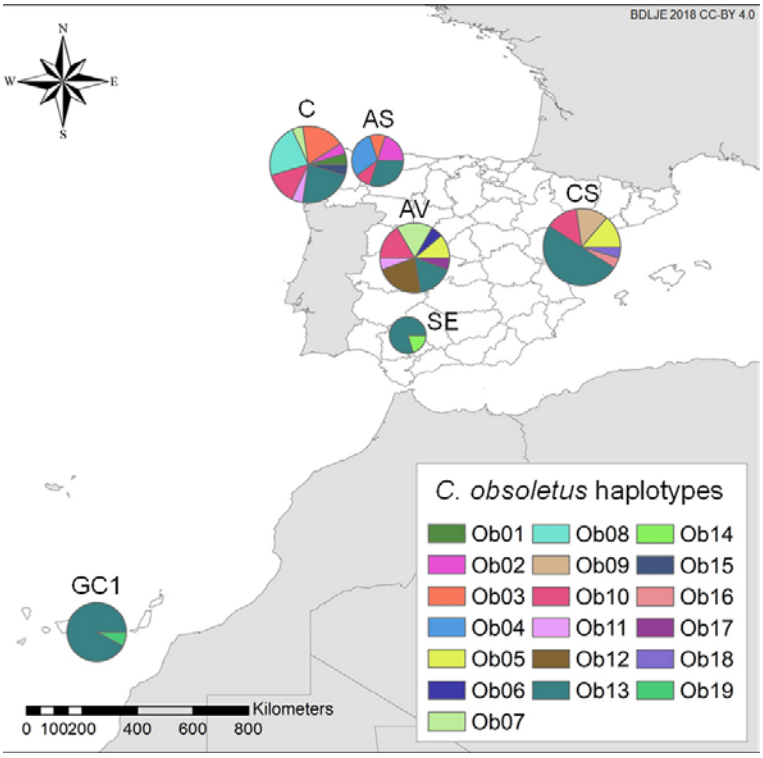

Figure 5. Distribution of the 19 different C. obsoletus haplotypes identified in this study. Differences in size of the pie charts represent the number of individuals sampled per site. Spanish administrative boundaries were provided by the Instituto Geográfico Nacional (ign.es) (BDDAE CC-BY 4.0). Abbreviations are those of the province: Castellón (CS), Asturias (AS), Ávila (AV), La Coruña (C), Seville (SE) and Las Palmas (GC1).

3. *Culicoides* bacteriome composition in Europe

The bacteriome of 88 C*. obsoletus s.s.* from 6 European countries (including Denmark, Poland, Latvia, Serbia, Bulgaria, and Greece, Figure 6) was investigated by Partner 3 using 16S rRNA gene amplicon high-throughput sequencing. For this purpose, were used the sample collections and the total genomic DNA extractions realized for phylogenetic and phylogeographic studies described above.

The V3-V4 fragment (ca. 450 bp) of the 16S rRNA gene was amplified and the amplified fragments were sequenced on an Illumina Miseq plateform from the University of Liege. 96 samples (with three PCR replicates per sample) including individuals originating from bovine or horse farms as well as negative controls were processed with identical DNA extraction protocols to identify and filtrate possible contaminants.

Reads were first clustered into Operational Taxonomic Units (OTUs) at 97% sequence identity level using the cluster-otus command from the Usearch algorithm (Edgar, 2010) to get an overview of the structure of bacterial communities. Data were then analyzed using the Swarm v.2.1.1 pipeline (Mahé et al., 2015) to investigate *Culicoides*-associated bacterial communities at finer resolution. In the latter, three PCR replicates per sample (technical replicates) were analyzed separately to examine putative stochastic DNA amplification from certain taxa. Results were consistent between both Operational Taxonomic units (OTUs) and the finer bacterial variant analyses; *C. obsoletus s.s.* specimen from the same country showed similar bacterial communities, despite some inter-individual variation and the presence of few outliers (e.g., specimen from Poland, Figure 7). Nevertheless, some overlaps were observed between the overall community composition of specimen from Denmark and Bulgaria, with some samples displaying a similar bacterial community.

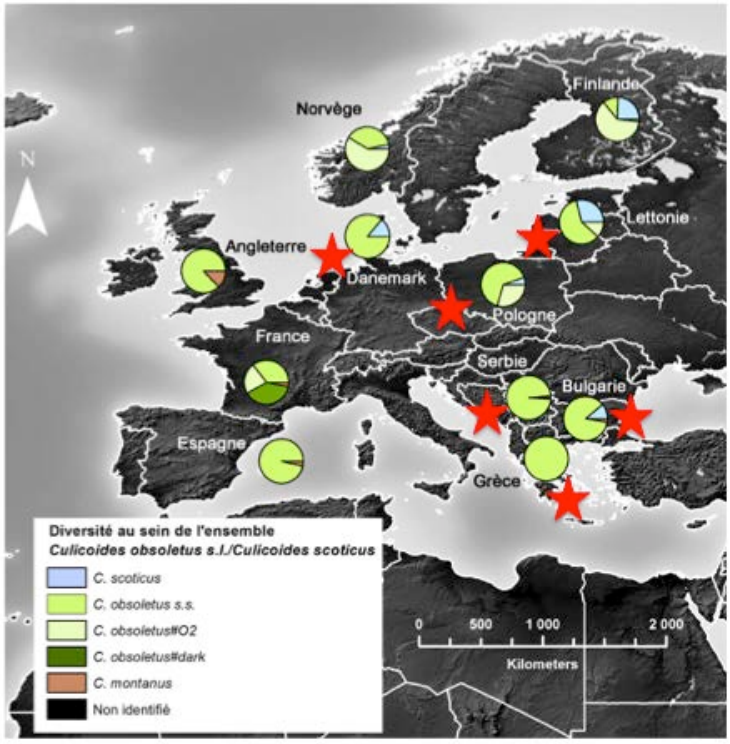

Figure 6. Distribution map of *Culicoides* species, with *C. obsoletus s.s.* sampled for bacteriome studies indicated by a red star.

Figure 7. NMDS plots for community structure of bacterial variants based on the Bray-Curtis index after Swarm clustering.

Alpha diversity analyses showed some higher diversity for Latvia, Poland and Serbia, as compared to Denmark, Bulgaria and Greece, which could partly explain the observed results. Finally, members of *C. obsoletus s.s.* and *C. obsoletus* clade O2 from Denmark showed distinct bacterial communities, highlighting that distinct species from the same country hold distinct bacterial communities.

These data suggest that genetic factor, in addition to the environment, likely influence the structure of bacterial communities.

4. *Culicoides* virome composition

To estimate virus exchange between vector populations, the objective was to compare the virome (i.e. the community of virus species associated to a host or environment) of *C. imicola* populations situated along the geographical distribution of this vector species using metagenomics. Geographical clustering of viromes will be then compared to the vector genetics (Jacquet *et al.* 2015).

Partner 3 has first generated a collection of samples from *C. imicola* populations derived from regions previously defined as part of the four genetic clusters in *C. imicola* observed in African and Europe (Figure 8). This collection required midge sampling campaigns and specimen identification to the species level in each of 13 participant countries (19 sampling sites in total). Moreover, for each country, Partner 3 carried out the legal and administrative procedures required for the exchange of insect samples (material transfer agreements and Nagoya protocol requirements). Partner 3 also organized sample transportation at the temperature conditions required to maintain sample integrity.

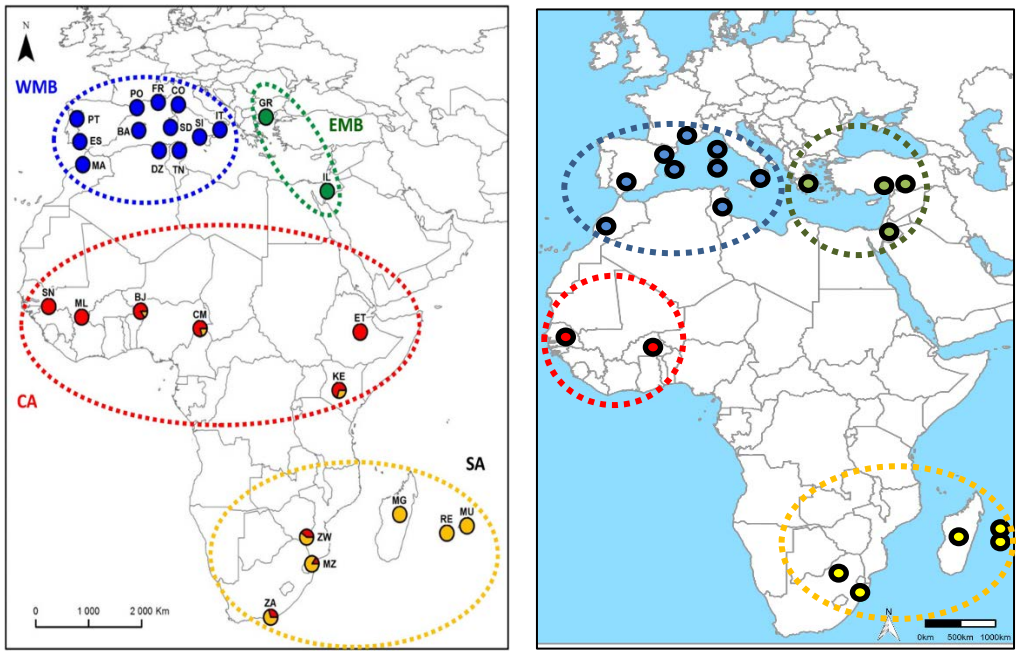

Figure 8. Comparaison of sampling areas in a previous study on *C. imicola* genetics and current sampling for the virome of *C. imicola*. Left panel: Genetic clustering of *C. imicola* populations obtained with spatial Bayesian clustering based on microsatellite data (from Jacquet et al. 2015; Fig. 2a in the article). Right panel: regions selected for sampling of *C. imicola* populations to study the virome under PALE-Blu. Sampling in Eastern Africa (Kenya) has been attempted but unsuccessfully due to operational problems.

All these activities have been severely hampered by the COVID-19 pandemic. Nevertheless, all samples, but the samples from Burkina Faso, have been transported to Partner 3 facilities (Table 1). Moreover, material transfer agreements have been obtained or are in the signature step for all participants (Table 1). For each country, the national bureau in charge of the Nagoya protocol has been contacted to validate that the sample exchange followed the national guidelines (Table 1). Thus, a collection from 12 countries can be analyzed with metagenomics.

Table 1. The *Culicoides imicola* collection. Sampling in Eastern Africa (Kenya) has been attempted but unsuccessfully due to operational problems. Obtained: already shipped to and stored in CIRAD. NR: not required

Region	Country	Organism	Sites	Collection status	MTA	Nagoya P.
Western Mediterranean	France	CIRAD	2	obtained	NR	NR
Western Mediterranean	Italy	IZS	2	obtained	done	_
Western Mediterranean	Morocco	IAV	1	obtained	done	_
Western Mediterranean	Tunisia	Institut Pasteur de Tunis	2	obtained	in progress	_
Western Mediterranean	Spain	CSIC (Andalusia)	1	obtained	done	in progress
Western Mediterranean	Spain	Univ. de Baleares (Balearic islands)	1	obtained	done	in progress
West Africa	Senegal	ISRA	1	obtained	done	done
West Africa	Burkina Faso	CIRAD	1	in progress	done	done
South Africa	South Africa	ARC-OVR	1	obtained	done	done
Indian Ocean	La Réunion (France)	CIRAD	2	obtained	NR	NR
Indian Ocean	Madagascar	Institut Pasteur de Madagascar	2	obtained	in progress	done
Eastern Mediterranean	Greece	Greece Institute for Infectious and Parasitic Diseases	1	obtained	done	_
Eastern Mediterranean	Turkey	KAFKAS UNIVERSITESI	2	obtained	done	_
Eastern Mediterranean	Israel	KIMRON VETERINARY INSTITUTE	1	obtained	done	_

In parallel to collection generation, a metagenomic analysis of the virome in available samples has been carried out. This analysis included *C. imicola* populations from Corsica (France), Sardinia (Italy) and Calabria (Italy). All these sites are separated by the Mediterranean Sea and belong to Western Mediterranean cluster of the *C. imicola* genetics. Although limited in geographical extent, this analysis represents the first in-depth characterization of the *C. imicola* and provides key data to improve the experimental design in the analysis of the full collection.

For each site, 13 200 *C. imicola* adults caught in 2016 and 2017 were pooled in groups of fifty individuals (264 pools; between 83 and 92 pools per site). Each pool was processed with our virus metagenomics protocol to generate a library for Illumina sequencing (Gil *et al.* 2021). Moreover, eight additional control libraries were included. Library quality was first validated through a low-depth sequencing run in a MiSeq instrument (Nano cartridge; Illumina). Then, the 272 libraries were sequenced at high depth in a full sequencing run in a HiSeq2500 instrument (rapid mode; 250 base-pair reads in paired-end mode). Read output was 609 million reads, providing a mean read coverage per library of 2 million reads as expected. Reads were then processed with an in-house bioinformatics pipeline (Gil *et al.* 2021). The pipeline involves several steps including quality filtering, de novo assembly and a homology search of viral sequences.

Figure 9 shows the output of reads, contigs and OTUs for main steps in the pipeline.

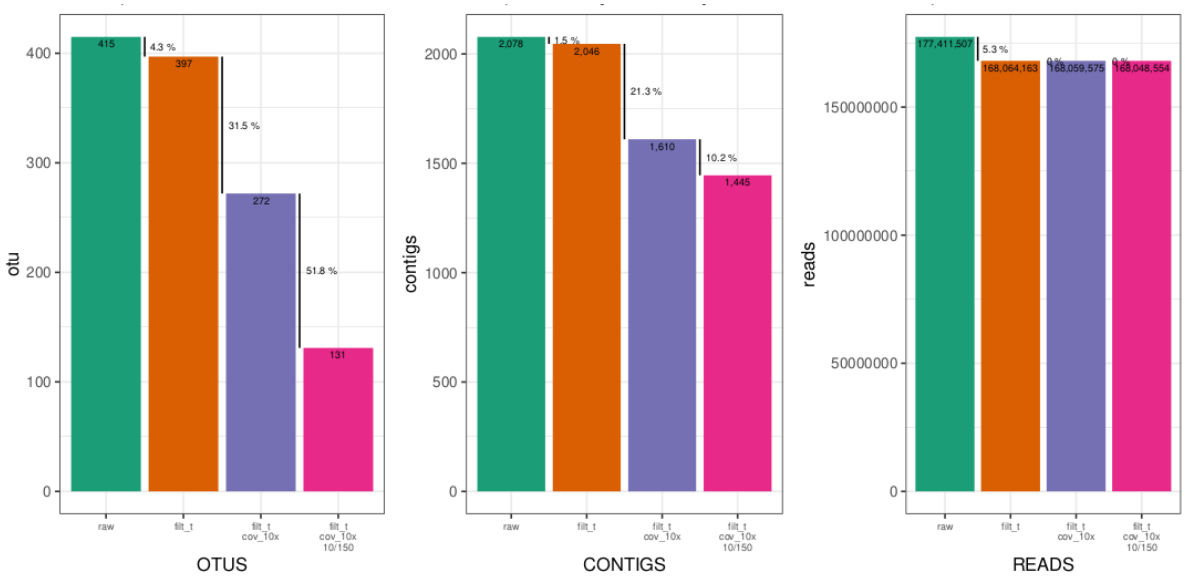

Figure 9. OTU, contig and viral read output. Each bar provides output at each step, from the sequencing through the different filters of the bioinformatics pipeline.

After read processing, the proportion of viral reads was 29% (Figure 9), a figure one fold higher than those usually observed in the literature. Those viral reads were distributed over 131 operational taxonomic units (species-like taxonomic level) which, in turn, represented a large taxonomic diversity encompassing 38 clusters (family-like taxonomic level) (Figure 10). Viromes were largely dominated by one OTU close to Hubei chipolycivirus, a recently-discovered virus in the Polycipiviridae, a family of arthropod viruses. This OTU represented above 60% of the reads in most libraries, no matter the site (Figure 11).

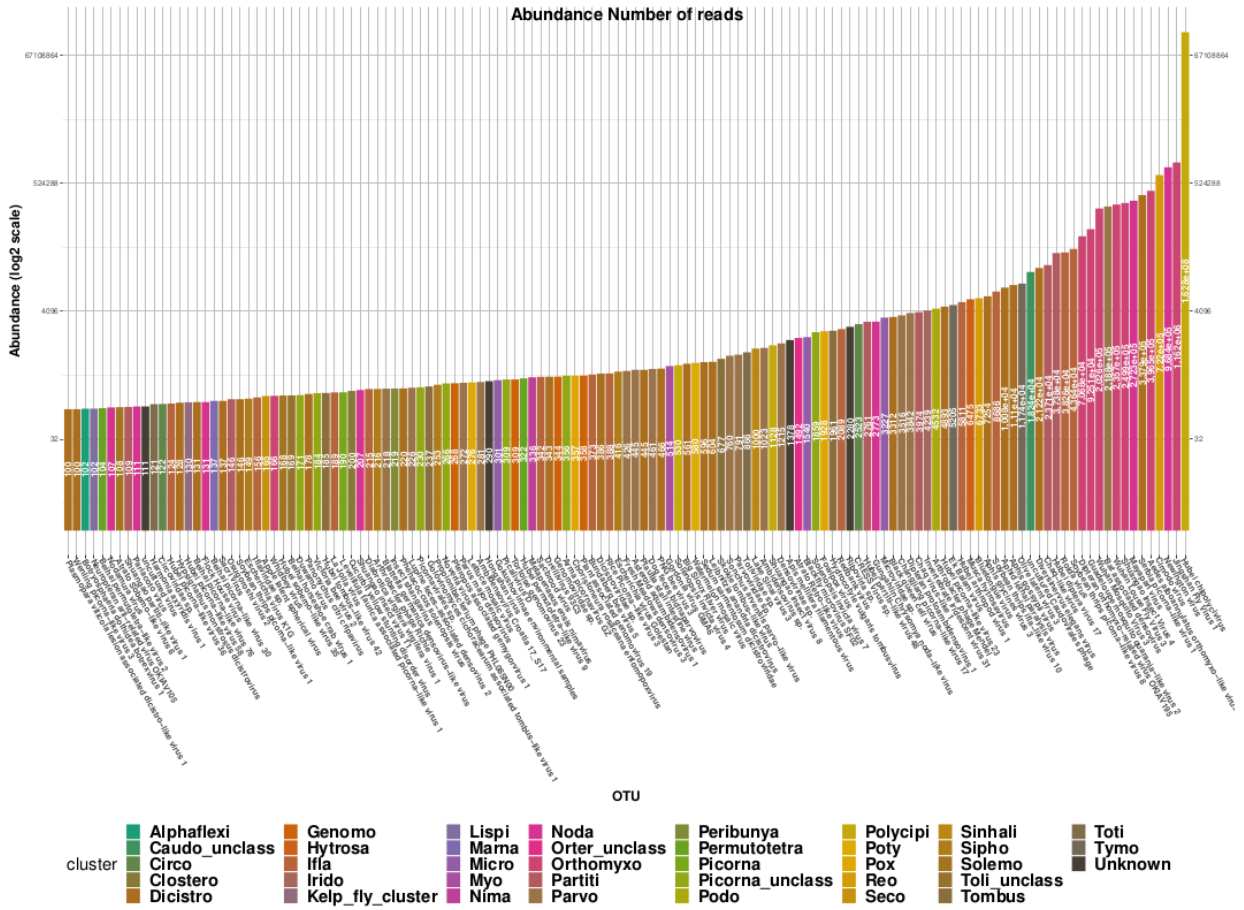

Figure 10. Read number of the 131 OTUs. Bar colours indicate the cluster or family-like taxonomic level of each OTU. The read number per OTU is provided on each bar.

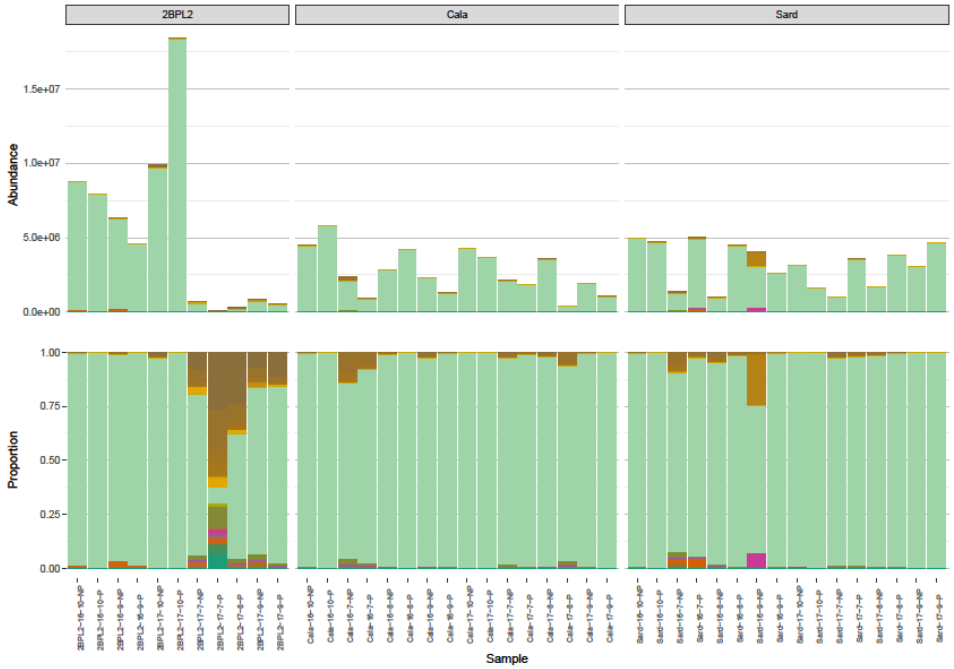

Figure 11. Read number (abundance) and proportion in each group of pools from the same site and period. The predominant OTU in the Polycipi cluster is shown in light green colour. 2BPL2: Corsica, Cala: Calabria, Sard: Sardinia.

No clear clustering of viromes depending on site was observed (Figure 12); a situation suggesting limited barriers to virus exchange among the populations in the study. Overall, these results strongly suggest the suitability of the methodology used to infer virome structure in *C. imicola*. Future analyses will include the full collection.

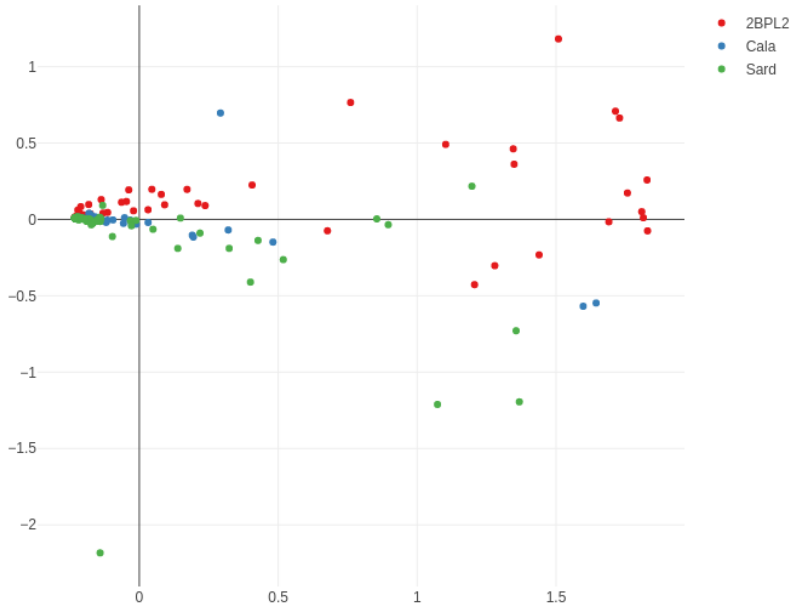

Figure 12. Non-parametric multidimensional scaling analysis (NMDS) of Bray-Curtis dissimilarities of the viromes in *C. imicola* populations. 2BPL2: Corsica, Cala: Calabria, Sard: Sardinia.

References

Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26(19): 2460-61. doi: 10.1093/bioinformatics/btq461.

Jacquet S, Garros C, Lombaert E, Walton C, Restrepo J, Allene X, *et al.* (2015). Colonization of the Mediterranean basin by the vector biting midge species *Culicoides imicola*: an old story. Molecular Ecology, 24(22): 5707-25. doi: 10.1111/mec.13422.

Gil P, Dupuy V, Koual R, Exbrayat A, Loire E, Fall AG, *et al.* (2021). A library preparation optimized for metagenomics of RNA viruses. Molecular Ecology Resources, 21(6): 1788-807. doi: 10.1111/1755-0998.13378.

Mahé F, Rognes T, Quince C, De Vargas C, Dunthorn M (2015). Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ 3: e1420. doi: 10.7717/peerj.1420.