PREPRINT

Author-formatted, not peer-reviewed document posted on 20/05/2021

DOI: https://doi.org/10.3897/arphapreprints.e68864

Occurrence data for the two cryptic species of *Cacopsylla pruni* (Hemiptera: Psylloidea)

Nicolas Sauvion, Jean Peccoud, Christine Meynard, David Ouvrard

ARPHA Preprints Author-formatted, not peer-reviewed document posted on 20/05/2021. DOI: https://doi.org/10.3897/arphapreprints.e68864

Occurrence data for the two cryptic species of *Cacopsylla pruni* (Hemiptera: Psylloidea)

Nicolas Sauvion[‡,§], Jean Peccoud[|], Christine Meynard[¶,‡], David Ouvrard[#]

‡ National Research Institute for Agriculture, Food and Environment (INRAE), Montpellier, France
§ PHIM, Univ Montpellier, INRAE, CIRAD, Montpellier SupAgro, Montpellier, France
| UMR CNRS 7267 Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, Université de Poitiers, Poitiers, France
¶ CBGP, Univ Montpellier, INRAE, CIRAD, IRD, Montpellier SupAgro, Montpellier, France
# ANSES-Laboratoire de la Santé des Végétaux, Montpellier, France

Corresponding author: Nicolas Sauvion (nicolas.sauvion@inrae.fr)

Abstract

Background

Cacopsylla pruni is a psyllid that has been known since 1998 as the vector of the bacterium ‘*Candidatus* Phytoplasma prunorum’, responsible for the European stone fruit yellows (ESFY), a disease that affects species of *Prunus*. This disease is one of the major limiting factors for the production of stone fruits, most notably apricot (*Prunus armeniaca*) and Japanese plum (*P. salicina*), in all EU stone fruit-growing areas. The psyllid vector is widespread in the Western Palearctic, and evidence for the presence of the phytoplasma that it transmits to species of *Prunus* has been found in 15 of the 27 EU countries.

Recent studies showed that *C. pruni* is actually composed of two cryptic species, which can be differentiated by molecular markers. A literature review on the distribution of *C. pruni* was published in 2012, but it only provided presence or absence information at the country level and without distinction between the two cryptic species.

Since 2012, numerous new records of the vector in several European countries have been published. We ourselves have acquired a large amount of data from sampling in France and other European countries. We have also carried out a thorough systematic literature review to find additional records, including all the original sources mentioning *C. pruni* (or its synonyms) since the first description by Scopoli in 1763. Our aim was to create an exhaustive georeferenced occurrence catalog, in particular in countries that are occasionnaly mentioned in the literature with little detail. Finally, for countries that seem suitable for the proliferation of *C. pruni* (USA, Canada, Japan, China, etc.), we digged deeper into the literature and reliable sources (e.g. checklist) to better subtanciate its current absence from those regions.

Information on the distribution ranges of these vector psyllids is of crucial interest in order to best predict the vulnerability of stone fruit producing countries to the ESFY threat in the foreseeable future.

New information

We give free access to a unique file of 1975 records of all occurrence data in our possession concerning *C. pruni,* which we have gathered through more than twenty years of sampling efforts in Europe or through intensive text mining.

We have made every effort to retrieve the source information for the records extracted from litterature (1201 records). Thus, we always give the title of the original reference, together with the page(s) citing *C. pruni* and, if possible, the year of sampling. To make the results of this survey publicly available, we give a URL to access the literature sources. In most cases, this link allows to freely download a PDF file.

We also give access to information extracted from GBIF (162 exploitable data points on 245 occurrences found in the database), which we thoroughly checked and often supplemented to make the information more easily exploitable.

We give access to our own unpublished georeferenced and genotyped record from 612 samples taken over the last 20 years in several European countries (Switzerland, Belgium, Netherlands, Spain, etc.). These include two countries (Portugal and North Macedonia) for which the presence of *C. pruni* had not been reported before. As our specimens have been genotyped (74 sites with species A solely, 202 with species B solely, and 310 with species A+B), our new data enable a better view of the geographical distribution of the two species at the Palaearctic scale.

Keywords

Hemiptera, psyllid, *Cacopsylla pruni*, vector-borne plant pathogen, phytoplasma, '*Candidatus* phytoplasma prunorum', European stone fruit yellows, species distribution, epidemiology

Introduction

Psyllids (Psylloidea), or jumping plant-lice, are plant sap-sucking hemipterans that could be considered as a minor group in terms of species diversity (3,573 described species according to Ouvrard 2021, compared to 104,165 hemipteran species according to Zhang 2013). However, a few psyllids are among the most devastating pests of annual and perennial crops due to their ability to transmit phytopathogenic bacteria causing significant agricultural losses. For example, *Bactericera cockerelli* (Šulc, 1909) is the vector of a liberibacter responsible of the Zebra chips (ZC), a disease that caused millions of dollars in losses to the potato industry in the United States, Mexico, Central America and New

Zealand, often leading to the abandonment of entire fields (Munyaneza 2012). The huanglongbing (HLB), the world's most devastating disease of trees of species of *Citrus*, is associated with two psyllid species, *Diaphorina citri* (Kuwayama, 1908) and *Trioza erytreae* (Del Guercio, 1918) (Bové 2006, Gottwald 2010, Khamis et al. 2017, Shimwela et al. 2016 , Rwomushana et al. 2017, Ajene et al. 2020). In 2014, *T. erytreae* was fortuitously discovered in Spain and Portugal (Arenas-Arenas et al. 2018). Although circum-Mediterranean species of *Citrus* have been thus far spared from the disease, the sporadic records of *T. erytreae* in these regions exposes them to a potential devastating epidemic (Cocuzza et al. 2017).

Other bacteria transmitted by psyllids to fruit trees have major economic impacts, in Europe in particular (Hadidi et al. 2011). These are phytoplasmas of trees of species of *Prunu*s, as well as apple and pear trees, transmitted by psyllids of the genus *Cacopsylla*. These respectively cause the European stone fruit yellows (ESFY), the Apple Proliferation (AP) and the Pear Decline (PD) (Jarausch et al. 2019). These bacteria and their vectors are native to Europe where they occur widely in orchard as well as wild habitats, preventing the eradication of the vectors and therefore containment of the diseases. The psyllid vectors are controlled mainly by insecticides, but the evolution of farming practices (e.g. reduction in the use of pesticides) and European regulations (i.e. pathogens removed from the list of quarantine organisms) could be the source of new emergences in the near future. In spite of great efforts from the European research community to better understand the biology and the ecology of the psyllid vectors of phytoplasmas (COST Action FA0807 2013, MacLeod et al. 2012), the presence of these insects in some part of Europe, and even in other parts of the world imapcted by these diseases, remains unclear (Steffek et al. 2012). Resolving this uncertainty would help to assess the risks posed by the fruit tree phytoplasmas (MacLeod et al. 2012) and to make decisions to manage these risks.

Dispersal of psyllid vectors poses a threat to food security across countries, stressing the need to anticipate the risks associated with introductions of new psyllids. Mapping the vector potential distributions under scenarios of introduction is crucial to an efficient pest risk assessment (PRA) framework (Venette et al. 2010). Occurrences representing the extent and variability within the current range of a given species are key to characterize and map its potential distribution under scenarios of introduction or climate change. Species distribution models (SDMs) have become the main predictive tool to achieve this goal (Elith and Leathwick 2009, Guisan et al. 2013). SDMs have proven their usefulness, inter alia, in invasion biology (e.g., Meynard et al. 2017, Syfert et al. 2017) and in conservation biology (e.g., Guisan et al. 2013, Muscatello et al. 2021). In plant pathology, SDMs are also increasingly used to predict the potential distributions of vector-borne plant pathogens (e.g., Benhadi-Marn et al. 2020, Narouei-Khandan et al. 2016, Shimwela et al. 2016). However, the reliability of these models heavily depends on the quality of the occurrence data that is used as input to map species distributions.

At least four criteria should be considered before using occurrence data as input for SDMs (Meynard et al. 2019): geographic and environmental representation and extent, quantity, accuracy of the georeferenced records and accuracy of the taxonomic identification. In short, occurrence data points should represent the full extent of biodiversity within the environments that the species is able to occupy, they should be numerous enough to allow its characterization, and geographic coordinates and taxonomic identification should be accurate, as these may otherwise introduce error in the modelled occurrence-environment correlations. High-quality data to properly map a species' distribution are often difficult to obtain, especially in insects. Indeed, collecting insects and information on their biology is often a time-consuming process that requires high taxonomic expertise. Insect species identification may necessitate painstaking morphological analyses or even the development of specific tools such as molecular markers (e.g., Peccoud et al. 2013). Recent studies have shown that different populations or genotypes within the same taxon can represent different risks, resulting in strikingly different SDM outputs (Meynard et al. 2017 Chardon et al. 2020). Genotypic information throughout the species range can therefore be crucial in the risk assessment process.

Historical data may also consitute a precious resource to help trace vector dispersion routes or simply to access specimens that can no longer be obtained (e.g., samples from an inaccessible locality). Many museums and academic institutions hold field notebooks and maintain first-rate collections that are rich of valuable information (e.g., collection date and locality) on insect specimens collected during scientific expeditions (Graham et al. 2004, Lister and Climate Change Research Group 2011, Suarez and Tsutsui 2004). Such data have proven useful in reconstructing the history of human or animal infectious diseases and in identifying their sources or reservoirs, in particular for mosquito-borne pathogens (e.g., West Nile virus, Suarez and Tsutsui 2004). To our knowledge however, this task has never been undertaken for vector-borne plant diseases, and historical records appear underexploited, even if they concern regions where such diseases have been endemic for tens to hundreds of years.

Cacopsylla (Thamnopsylla) pruni (Scopoli, 1763) is known since 1998 as the vector of a bacterium, '*Candidatus* Phytoplasma prunorum' responsible for ESFY (Carraro et al. 1998) and is currently listed as Regulated Non-Quarantine Pest (RNQP) in the Annex IV-Part D of the European Council Directive 2019/2072 (EUR-Lex 2020). This psyllid is widespread in the Western Palearctic (Ouvrard 2021) and the phytoplasma it transmits are reported in 15 of the 27 EU countries (Steffek et al. 2012). ESFY is one of the major factors limiting the production of stone fruits, most notably apricot (*Prunus armeniaca* L.) and Japanese plum (*Prunus salicina* Lindl.) in all EU stone fruit-growing areas. These aears include the three most important apricot producing countries, Spain, Italy, and France, which provided 73% of the EU apricot production in 2012 according to Eurostats. In the last twenty years, great efforts have been made to characterize the biology of the ESFY vector (Peccoud et al. 2013, Peccoud et al. 2018), the life cycle of the transmission (Thébaud et al. 2009), the genetic variability of the pathogen (Danet et al. 2011, Marie-Jeanne et al. 2020) and the risk factors of the disease (Marie-Jeanne et al. 2020, Thébaud et al. 2006). But despite these efforts and the rigorous sanitary control of fruit trees as part of the certification process, the disease continues to pose great problems to fruit growers in Europe, which raises the question of the origin of contaminations in orchards.

In their review, Steffek et al. (2012) pointed out important uncertainties that could undermine the management of ESFY. The rate of psyllid dispersal at various scales (i.e., a growing region, country, Europe or even larger), by natural means or human transportation, and the risk of introduction and establishment in new countries were two of the essential issues that remained unresolved. The presence of the vector in several countries from the southernmost part of Europe (Portugal, southern Spain, Greece, etc) which can be directly impacted by ESFY, as well as neighboring countries, remains undetermined. At the time of review by Steffek et al. (2012), preliminary studies had shown that *C. pruni* was composed of two genetic groups then called "biotypes" (Sauvion et al. 2007, Sauvion et al. 2009). However, no detailed data was available on the European distribution of these two biotypes, which were analyzed jointly in this review.

Establishing the geographic distribution of *C. pruni*, and possibly for each biotype, was therefore a priority. To this end, we developed molecular markers to easily identify the *C. pruni* biotypes (Peccoud et al. 2013), which allowed us to establish their species status (Peccoud et al. 2018). Numerous new surveys on the presence of *C. pruni* in several European countries have been published (e.g., Etropolska et al. 2015, Jarausch et al. 2019, Sabaté et al. 2016, Seljak 2020, Warabieda et al. 2018), sometimes with a distinction between the two species. In our own laboratory at INRAE-Montpellier, we obtained a large collection of samples through twenty years of surveys in France and other European countries (Portugal, Spain, Belgium, Switzerland, Italy, etc). Some of these samples have been used in publications, but the vast majority have not yet been released in a georeferenced format. We were also able to find unpublished and valuable information in GBIF (e.g., metadata from Natural History Museum of London). Recently, we conducted an extensive literature survey for the original sources mentioning *C. pruni,* as a mean of verification, but more importantly, to precisely locate the source of each specimen. This laborious work often resembled a treasure hunt with its typical pitfalls and puzzles, such as correctly translating Mogolian locality names from a text written in Russian and then georeferencing them (Fig. 1). Sparing others these obstacles was part of our motivation to make the results of this survey publicly available.

Our objective is to give access, through a unique dataset, to all the data we have gathered on the two species of *C. pruni*. In this way, we hope to contribute to a better management of ESFY in countries affected by the disease, and to a better anticipation of the risk of introduction in countries not yet affected.

General description

Purpose: This dataset is a compilation that is meant to include all available information (literature, GBIF, INRAE unpublished data) on the geographical distribution of two cryptic species of the psyllid *Cacopsylla pruni* at the scale of the Palearctic (Fig. 2). We aimed to publish third-party data that can be otherwise hard to access and first-party data that are not yet published, and to ensure free, open access to that information.

Sampling methods

Study extent: The data contained in this dataset have three different origins: a systematic literature review, the Global Biodiversity Information Facility [GBIF] network, and field collections by researchers/students from INRAE-Montpellier. They cover several ecoregions of the Palaearctic (Fig. 2): the Euro-Siberian region, the Mediterranean Basin, the Western and East Asia (Northern parts). No data was found for Central Asia nor for the Nearctic, despite the known presence of trees of species of *Prunus* and conifers on which *C. pruni* could make its life cycle.

Sampling description:

Literature data

In order to extend upon the Steffek et al. (2012) review, we have undertaken a new systematic literature survey for articles/manuscripts/books using the keyword "*Cacopsylla pruni*", its previous combinations "*Chermes pruni" and "Psylla pruni",* or its synonym "*Psylla fumipennis".* To this end, we used the Google Scholar search engine (https://scholar.google.com/) and we explored several scientific databases (AGRICOLA, Agris, CAB Abstract, Web of Science), as well as other types of databases more or less specialized on the subject:

- Psyl'list (https://www.hemiptera-databases.org/psyllist/), an online database dedicated to jumping plant lice;
- National Inventory of Natural Heritage (https://inpn.mnhn.fr/accueil/index), the French portal for biodiversity and geodiversity;
- ISTEX (https://www.inist.fr/services/acceder/istex/) a platform offering the French higher education and research community access to more than 23 million articles from all scientific disciplines and which cover a very long period (from ~1400 to 2019);
- Collections of the Natural History Museum of London (https://www.nhm.ac.uk/our-science/collections.html);
- Gallica (https://gallica.bnf.fr), the digital library of the BNF (Bibliothèque Nationale de France);
- Biodiversity Heritage Library [BHL] (https://www.biodiversitylibrary.org/) the world's largest open access digital library for biodiversity literature and archives.

The searches were not restricted by language and were traced back to the first description of *C. pruni* (1763). Each line of the dataset that we make available (see section 'Data resources') corresponds to a reference. For almost all of them, we have retrieved the PDF file of the orignal publication (including old books) which allowed us to verify the information. The corresponding URL is given for each data in the dataset (DOI link or similar link generally giving direct access to the PDF). We systematically tried to specify the locality where the observation was made (see Quality control section). Whenever the information was available, we specified the cryptic species of *C. pruni* (A or B, according to

Peccoud et al. 2013) and the collection plant. In the end, we were able to exploit 1201 occurrence data from the literature survey (Fig. 3).

GBIF data

A search on the keywords "*Cacopsylla pruni*" returned 245 occurrences in GBIF (2020-12-07). Among these, we were able to extract the names of 45 localities with geographic coordinates. For 87 occurrences, for which only the name of the locality was given, we retrieved the geographic coordinates from Google Earth. The database also provided images of scanned slides from the NHM collection (https://www.gbif.org/fr/occurrence/gallery?taxon_key=2012955) from which we retrieved precise information about the sampling (date, location, host plant, collector)(Fig. 4), sometimes redundant with our own information (e.g. data from Iran). Finally, 28 occurrences were derived from information associated with DNA sequences deposited in iBOL (https://ibol.org), including 24 sequences deposited by us and already entered in our dataset (e.g., https://www.ncbi.nlm.nih.gov/nuccore/MH577786). In total, 162 occurrences data have been extracted from GBIF (Fig. 5).

Sampling data

For more than 20 years, researchers (Gérard Labonne, Gaël Thébaud, Jean Peccoud, Christian Cocquempot, Nicolas Sauvion) or students of INRAE-Montpellier have collected *C. pruni* individuals. Using a beating tray (80 cm x 80 cm), we collected essentially on *Prunus spinosa* L. (blackthorn) in spring, and the rest of the year on *Pinus nigra* J.F Arnold (Black Pine), *Picea abies* (L.) H.Karst. (Common Pitch-fir), and *Abies alba* Mill. (Common silver Fir). Other congeneric species where sometimes caught, but *C. pruni* individuals were easily recognized by the colour of the forewing, which is dark brown at the apex and brown in the remaining part. Soon after identification, samples were conserved in 96% ethanol until DNA extraction, and then genotyped (for species determination) according to the protocol described by Peccoud et al. (2013).

We recorded the GPS coordinates of all collected samples in their wild habitat, geolocalizing the bush, hedge or shrub sampled. For the few insects sampled in orchards, we attributed a unique GPS coordinate — corresponding to the centre of each plot — to all the corresponding samples. The name of the locality given in the dataset corresponds to the nearest locality to the sampled point. We sampled mainly France, without restriction to apricot-growing regions and focusing on Southern regions where species A and B live in sympatry or in strict allopatry. We also collected samples in Spain, Switzerland and Italy. The addition of these 612 new occurrence data improves the picture of the geographical distribution of the two species, hence it should be valuable for risk assessment, phylogeography or population genetics studies (Fig. 2, Fig. 6, Fig. 7)

Quality control:

We have a strong expertise in the taxonomy of psyllids (Ouvrard 2021). Over the last few years, we have accumulated a large number of references on these insects in an article database, including reference that are old and/or difficult to trace. As we had all these articles in PDF or paper, or other metadata (e.g., scanned images), we were able to retrieve and thoroughly verify all information concerning *C. pruni* or its synonyms and combinations.

All the specimens the we collected in the field were first carefully visually examined and then genotyped according to Peccoud et al. (2013), which effectively eliminates all risk of misidentification.

Wherever possible, geographic coordinates (in WGS-84 coordinate system) refer to specific localities. We used Google Earth to search and reference each locality name found in the literature or GBIF, being careful about homonymy and translation of names, and possible changes of country names. We consider the precision of these geographical coordinates to be a few kilometers, as authors rarely give very precise coordinates of their sampling points. Conversely, whenever we found geographical coordinates in GBIF, we plotted them on a Google Earth map to identify the closest locality and to check consistency with other information provided (name of the region, country, etc). When no locality name was given, precision may vary from city to province, region or country (e.g., "USSR: South European Part"). In this case, we specified that the “locality is not stated". For data points only specifying countries, we provided the GPS coordinates of the country centers extracted from Google Earth, for lack of a better option. We therefore included a column with the estimated precision for each records, stressing that some of these data should be used with caution depending on the level of precision required for analyses. Conversely, GPS coordinates of our own collected samples (see previous section) have an accuracy of a few meters. Each point was first geolocalised with a portable GPS and then checked on Google Map.

Step description: Most field names of the dataset were chosen according to Darwin Core format (Wieczorek et al. 2012), and the latest version of the list of core terms as of 2020-10-28 (http://rs.tdwg.org/dwc/version/terms/2020-10-28.htm): “catalogNumber”, “phylum”, “order”, “genus”, “acceptedNameUsage”, “Occurrence”, “country”, “countryCode”, “locationRemarks”, “locality”, “coordinateUncertaintyInMeters”, “decimalLatitude”, “decimalLongitude”, “ownerInstitutionCode”, “locationAccordingTo”, “dateIdentified”, “eventDate”, “associatedReferences”. We have added 11 columns with names not defined by Darwin Core: “suborder”, “superfamily”, “family“, “subfamily”, “speciesA”, “speciesB”, “hostPlantFamily”, “hostPlantLatinName”, “hostPlantVernacularName”, “sourceCategory”, “page”.

Geographic coverage

Description: The database covers the entire known geographic range of the two species of the psyllid *C. pruni,* from Morocco to Norway and from Portugal to Mongolia.

We have also extended our search to other countries where either species could potentially be found, in particular countries where different species of *Prunus* are described in wild or cultivated ecological compartments (e.g. Japan, China, USA, Canada), and where these psyllids could be phytoplasma vectors. Whenever possible, we relied on checklists from recognised taxonomists to ensure the veracity of the information before concluding to an "absence" (e.g., Inoue 2010, Maw et al. 2000).

Coordinates: 33.815458 and 65.59623333 Latitude; -8.383379 and 112.52588611 Longitude.

Taxonomic coverage

Description: The data paper focuses on two cryptic species of *Cacopsylla (Thamnopsylla) pruni* (Scopoli, 1763), currently referred to as A and B. species of *Cacopsylla pruni* show clear genetic differences despite being morphologically and ecologically indistinguishable (Peccoud et al. 2013, Peccoud et al. 2018). These psyllids are sternorrhynchans of the order Hemiptera, belonging to the superfamily Psylloidea, family Psyllidae, and subfamily Psyllinae according the classification by Burckhardt et al. (2021).

Temporal coverage

Living time period: 1763-2020.

Notes: Litterature data cover 1763 to 2020.

INRAE data cover 1998 to 2020.

Usage licence

Usage licence: Open Data Commons Open Database License (ODbL)

Data resources

Data package title: Compilation of occurrence data for two psyllid species of the *Cacopsylla pruni* complex (Hemiptera: Psylloidea)

Resource link: https://doi.org/10.15454/VC9UR5

Number of data sets: 1

Data set name: Cacopsylla pruni_occurrences_v29.csv

Character set: text/tab-separated-values

Download URL: https://data.inrae.fr/dataset.xhtml?persistentId=doi:10.15454/VC9UR5

Data format: Darwin Core Archive

Data format version: 10

Column label	Column description
catalogNumber	An identifier which assigns a unique code to each of the 1975 records (NS0001 to NS1975).
phylum	The full scientific name of the phylum in which the taxon is classified.
class	The full scientific name of the class in which the taxon is classified.
order	The full scientific name of the order in which the taxon is classified.
suborder	The full scientific name of the suborder in which the taxon is classified.
superfamily	The full scientific name of the ssuperfamily in which the taxon is classified.
family	The full scientific name of the family in which the taxon is classified.
subfamily	The full scientific name of the subfamily in which the taxon is classified.
genus	The full scientific name of the genus in which the taxon is classified.
acceptedNameUsage	The full name, with authorship and date information of the currently valid (zoological) taxon.
Occurrence	An existence of an Organism (sensu http://rs.tdwg.org/dwc/terms/Organism) at a particular place at a particular time. Here, five modalities: "insufficient data" (i.e., insufficient information to determine presence or absence); "probable absence" (i.e., no presence data yet found in records); "probable presence" (i.e., presence very likely but not yet confirmed); "confirmed presence".
speciesA	Information concerning the assignment of the specimens of a population (i.e. caught on the same day in the same locality on the same host plant) to species A of *C. pruni*. Three modalities: "not genotyped"; "not species A" (i.e., no individual of genotype A was found in the population analysed, but individuals of species B); "species A" (i.e. at least one individual of genotype A found in the population analysed). Genotyping was based on Peccoud et al. 2013.
speciesB	Information concerning the assignment of the specimens of a population (i.e. caught on the same day in the same locality on the same host plant) to species B of *C. pruni*. Three modalities: "not genotyped"; "not species B" (i.e., no individual of genotype B was found in the population analysed, but individuals of species A); "species B" (i.e., at least one individual of genotype B found in the population analysed). Genotyping was based on Peccoud et al. 2013.
country	Names of the countries where the individual(s) attributed to *C. pruni* have been recoprded according the universally applicable code ISO 3166-2:2013
countryCode	Two-letter country codes defined in ISO 3166-1, part of the ISO 3166 standard to represent countries where species have been described
locationRemarks	Comments or notes about the location.

locality	The specific description of the place. The locality is given as accurately as possible (precise address, village, town), but may sometimes be imprecise (e.g. mountain, region) or even absent (NA="locality not stated"). see column "coordinateUncertaintyInMeters" for more details on uncertainty.
coordinateUncertaintyInMeters	The horizontal distance (in meters) from the given decimalLatitude and decimalLongitude describing the smallest circle containing the whole of the Location. Leave the value empty if the uncertainty is unknown, cannot be estimated, or is not applicable (because there are no coordinates). Zero is not a valid value for this term. e.g., 30 m = margin of error in the measurement of coordinates using a GPS navigator; 1000 or 10000 m = uncertainty attributed to most locality names in the literature, in the absence of more precise information; 50 000 m = uncertainty when only the name of the region/province is known.
decimalLatitude	The geographic latitude (in decimal degrees according the geodetic coordinate reference system EPSG 4326) of the geographic centre of a location. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive.
decimalLongitude	The geographic longitude (in decimal degrees according the geodetic coordinate reference system EPSG 4326) of the geographic centre of a location. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive.
hostPlantFamily	Six modalities: "Fabaceae"; "Pinaceae"; "Rosaceae"; "Salicaceae"; "unknown" (specimens collected by sweeping or Malaise trap); "unspecified species". Here "host plant" is taken in the broadest sense, i.e. plants on which a psyllid species completes its immature to adult life cycle, or shelter plant (plants on which adult psyllids overwinter and on which they may feed), or casual plant (plants on which adult psyllids land but do not feed).
hostPlantLatinName	Latin name of the host plant species (i.e. host plant sensu stricto, shelter plant or casual plant) according to the International Code of Nomenclature for algae, fungi, and plants (https://www.iaptglobal.org/). e.g. *Picea abies* (L.) H.Karst., *Prunus spinosa* L., etc.
hostPlantVernacularName	Vernacular English name of the host plant species.
sourceCategory	The three different sources of information used to compile the dataset: "GBIF" (i.e., data from the Global Biodiversity Information Facility); "literature" (i.e., any data resulting from a text-mining from different sources - manuscript, book, article, etc - accessible or not on the web); "INRAE" (i.e., data from collections by INRAE Montpellier, not published to date).
ownerInstitutionCode	The name (or acronym) in use by the institution having ownership of the object(s) or information referred to in the record.

locationAccordingTo	Information about the source of this Location information. Could be a publication (gazetteer), institution, or team ofindividuals. Here, detailed title of the original reference associated with the locality; "no data" (i.e. no information found for a particular country, e.g. Kyrgyzstan, Malta).
dateIdentified	The date on which the subject was determined as representing the Taxon. Here, year of publication of the reference cited in the "locationAccordingTo" column.
page	Page where the original information about the locality can be found in the reference cited in the "locationAccordingTo" column
eventDate	The date-time or interval during which an Event occurred. For occurrences, this is the date-time when the event was recorded. Here, year(s) or date of sampling or observation in the locality according the information in the "locationAccordingTo" column.'1996' (some time in the year 1996). '2010-06' (some time in June 2010). '2010-02-12' (some time during 12 February 2010). '2007/2010' (some time during the interval between the beginning of the year 2007 and the end of the year 2010).
associatedReferences	A list (concatenated and separated) of identifiers (publication, bibliographic reference, global unique identifier, URI) of literature associated with the occurrence. Here, URL by which the original information can be retrieved (downloadable PDF file in open access, link to the publisher of a non-open access reference, direct link to the original GBIF occurrence, etc.).

Acknowledgements

We are very grateful to Josiane Peyre for her valuable technical assistance for the genotyping of thousands of psyllids, and the following individuals for their contributions to the collection of psyllid or plant samples: N. Courtieu & J.-M. Broquaire (SICA Centrex), N. Galabert & J. Delnatte (SICA L'Edelweiss), B. Rouillé (SRPV-PACA), E. Navarro (Terroir de Crau), P. Delon (CA-Gard), E. Falezan (GIE-Tain l'Hermitage), P. Exbrayat (CA-Drôme), M. Léon-Chapoux and V. Delaunay (SEFRA), G. Devènes (Agroscope). Many students also took part in the collection of the psyllids, for which they are warmly thanked: Ghislaine Sagna, Léa Merlet, Piroska Czibulyás, Elise Découvreur, Clara Bouchet, Clara Sauvion, Zo-Norosoa Andrianjaka-Camps, Florent Décugis, Olivier Lachenaud.

Part of this work benefitted from a postdoctoral grant to NS funded by an INRA-CIRAD SDIPS grant (Speciation and molecular Diagnosis of Insect Pest Species complexes). Field and molecular work for this study were supported by several projects during 15 years:

2005-07: ECOGER "Ecologie et adaptation des insectes phytophages en gestion de leurs populations" founded by le Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation-France;

2007-08: SEE-ERA.NET, network 'Phytoplasma epidemiology', funded by the 6th EU Frame-work Programme for Research and Development (contract number ERA-CT-2004-515805)

2009-11: SDIPS "Mechanisms of Speciation & Molecular Diagnosis of Insect Pest Species complexes" founded by INRA-France;

2010-12: SPEED@ID “Accurate SPEciEs Delimitation and IDentification of Eukaryotic biodiversity using DNA markers”. A project proposed by F-BoL, the French Barcode of Life initative - Genoscope Evry-France;

2010-12: PRIMA PHACIE “Pest risk assessment for the European Community plant health: A comparative approach with case studies”, founded by European Food Safety Authority (EFSA), grant agreement CFP/EFSA/PLH/2009/01;

2010-12: Bilateral project PIA BOSPHORUS between TUBITAK-Turkey and le Ministère des Affaires étrangères-France "Role of the vectors (psyllids) in the dissemination of the diseases due to phytoplasma on fruit trees";

2011-13: PHYLOPSYL from the project “Bibliothèque du vivant” (BdV) funded by three French institutions (the CNRS, INRA and MNHN);

2015-2018: E-SPACE project number 1504-004, Improving epidemiosurveillance of Mediterranean and tropical plant diseases, French Agropolis Foundation.

Author contributions

NS contributed to text mining, sampling and characterization of the insects, georeferencing, development of the dataset, map making and writing of the paper; DO provided easier access to scattered taxonomic data through its extensive expertise on psyllids, contributed to species validation and writing of the paper; JP contributed to sampling, molecular characterization of the insects and writing of the paper; CNM contributed to the development of the dataset and writing of the paper.

References

- Ajene IJ, Khamis F, Ballo S, Pietersen G, van Asch B, Seid N, Azerefegne F, Ekesi S, Mohamed S (2020) Detection of Asian citrus psyllid (Hemiptera: Psyllidae) in Ethiopia: a new haplotype and its implication to the proliferation of Huanglongbing. Journal of Economic Entomology 113: 1640-1647. https://doi.org/10.1093/jee/toaa113
- Arenas-Arenas FJ, Duran-Vila N, Quinto J, Hervalejo A (2018) Is the presence of *Trioza erytreae*, vector of huanglongbing disease, endangering the Mediterranean citrus industry? Survey of its population density and geographical spread over the last years. Journal of Plant Pathology 100: 567-574. URL: https://doi.org/10.1007/s42161-018-0109-8

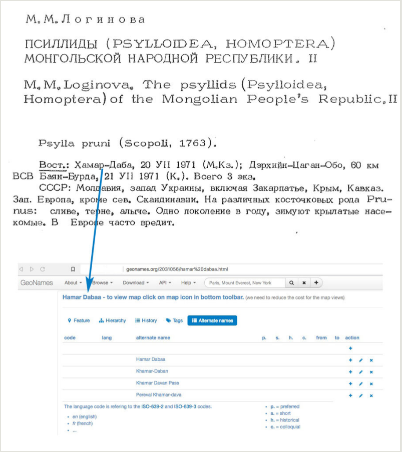
- Benhadi-Marn J, Fereres A, Pereira JA (2020) A model to predict the expansion of *Trioza erytreae* throughout the Iberian Peninsula using a pest risk analysis approach. Insects 11: 576. URL: https://doi.org/10.3390/insects11090576
- Bové JM (2006) Huanglongbing: a destructive, newly emerging, century-old disease of citrus. Journal of Plant Pathology 88: 7-37. URL: https://www.jstor.org/stable/41998278
- Burckhardt D, Ouvrard D, Percy D (2021) An updated classification of the jumping plant-lice (Hemiptera: Psylloidea) integrating molecular and morphological evidence. European Journal of Taxonomy 736: 137-182. https://doi.org/10.5852/ejt.2021.736.1257
- Carraro L, Osler R, Loi N, Ermacora P, Refatti E (1998) Transmission of European stone fruit yellows phytoplasma by *Cacopsylla pruni*. Journal of Plant Pathology 80: 233-239. URL: https://www.jstor.org/stable/41997929
- Chardon NI, Pironon S, Peterson ML, Doak DF (2020) Incorporating intraspecific variation into species distribution models improves distribution predictions, but cannot predict species traits for a wide-spread plant species. Ecography 43: 60-74. https://doi.org/10.1111/ecog.04630
- Cocuzza GE, Alberto U, Hernandez-Surez E, Siverio F, Di Silvestro S, Tena A, Carmelo R (2017) A review on *Trioza erytreae* (African citrus psyllid), now in mainland Europe, and its potential risk as vector of huanglongbing (HLB) in citrus. Journal of Pest Science 90: 1-17. https://doi.org/10.1007/s10340-016-0804-1
- COST Action FA0807 (2013) Integrated management of phytoplasma epidemics in different crop systems. URL: http://www.costphytoplasma.ipwgnet.org
- Danet J-L, Balakishiyeva G, Cimerman A, Sauvion N, Marie-Jeanne V, Labonne G, Lavin A, Batlle A, Krizanac I, Skoric D, Ermacora P, Ulubas-Serce C, Caglayan K, Jarausch W, Foissac X (2011) Multilocus sequence analysis reveals the genetic diversity of European fruit tree phytoplasmas and supports the existence of inter-species recombination. Microbiology 157: 438-450. https://doi.org/10.1099/mic.0.043547-0
- Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40: 677-697. https://doi.org/10.1146/annurev.ecolsys.110308.120159
- Etropolska A, Jarausch W, Jarausch B, Trenchev G (2015) Detection of European fruit tree phytoplasmas and their insect vectors in important fruit-growing regions in Bulgaria. Bulgarian Journal of Agricultural Science 21: 1248-1253.
- EUR-Lex (2020) Commission Implementing Regulation (EU) 2019/2072 of 28 November 2019 establishing uniform conditions for the implementation of Regulation (EU) 2016/2031 of the European Parliament and the Council, as regards protective measures against pests of plants, and repealing Commission Regulation (EC) No 690/2008 and amending Commission Implementing Regulation (EU) 2018/2019. http://data.europa.eu/eli/reg_impl/2019/2072/oj. Accessed on: 2020-12-04.
- Gottwald TR (2010) Current epidemiological understanding of citrus huanglongbing. Annual Review of Phytopathology 48: 119-139. https://doi.org/10.1146/annurev-phyto-073009-114418
- Graham CH, Ferrier S, Huettman F, Moritz C, Peterson AT (2004) New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology & Evolution 19: 497-503. https://doi.org/10.1016/j.tree.2004.07.006
- Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AI, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C (2013) Predicting

species distributions for conservation decisions. Ecology Letters 16: 1424-142. https://doi.org/10.1111/ele.12189
- Hadidi A, Barba M, Candresse T, Jelkmann W (2011) Virus and virus-like diseases of pome and stone fruits. The American Phytopathological Society Press, St Paul, Minnesota. URL: https://apsjournals.apsnet.org/doi/pdf/10.1094/9780890545010.fm [ISBN 978-0-89054-396-2]
- Inoue H (2010) The generic affiliation of Japanese species of the subfamily Psyllinae (Hemiptera: Psyllidae) with a revised checklist. Journal of Natural History 44 (5-6): 333-360. https://doi.org/10.1080/00222930903437325
- Jarausch B, Tedeschi R, Sauvion N, Gross J, Jarausch W (2019) Chapter 3: Psyllid vectors. In: Bertaccini PW, Rao GP, Mori N (Eds) Transmission and management of phytoplasma associated diseases. II. Springer, Singapore. [ISBN 978-981-13-2831-2]. https://doi.org/10.1007/978-981-13-2832-9_11
- Jarausch W, Jarausch B, Fritz M, Runne M, Etropolska A, Pfeilstetter E (2019) Epidemiology of European stone fruit yellows in Germany: the role of wild *Prunus spinosa*. European Journal of Plant Pathology 154: 463-476. https://doi.org/10.1007/s10658-019-01669-3
- Khamis FM, Rwomushana I, Ombura LO, Cook G, Mohamed SA, Tanga CM, Nderitu PW, Borgemeister C, Stamou M, Grout TG (2017) DNA barcode reference library for the African citrus triozid, *Trioza erytreae* (Hemiptera: Triozidae): vector of African citrus greening. Journal of Economic Entomology 110: 2637-2646. https://doi.org/10.1093/jee/tox283
- Lister AM, Climate Change Research Group (2011) Natural history collections as sources of long-term datasets. Trends in Ecology & Evolution 26: 153-154. https://doi.org/10.1016/j.tree.2010.12.009
- Loginova MM (1974) The psyllids (Psylloidea, Homoptera) of the Mongolian People's Republic II. Nasekomye Mongoli 4: 51-66. [In Russia].
- MacLeod A, Anderson H, Follak S, Van Der Gaag DJ, Potting R, Smith J, Steffek R, Vloutoglou I, Holt J, Karadjova O, Kehlenbeck H, Labonne G, Reynaud P, Viaene N, Anthoine G, Holeva M, Hostachy B, Ilieva Z, Karssen G, Krumov V, Limon P, Meffert J, Niere B, Petrova E, Peyre J, Pfeilstetter E, Roelofs W, Rothlisberger F, Sauvion N, Schank N, Schrader G, Schroeder T, Steinmller S, Tjou-Tam-Sin L, Ventsislavov V, K V, Wesemael W (2012) Pest risk assessment for the European Community plant health: a comparative approach with case studies. EFSA Supporting Publications 9: 319E. https://doi.org/10.2903/sp.efsa.2012.EN-319
- Marie-Jeanne V, Bonnot F, Thébaud G, Peccoud J, Labonne G, Sauvion N (2020) Multi-scale spatial genetic structure of the vector-borne pathogen '*Candidatus* phytoplasma prunorum' in orchards and in wild habitats. Scientific Reports 10: 5002. https://doi.org/10.1038/s41598-020-61908-0
- Maw HE, Foottit RG, Hamilton KG, Scudder GG (2000) Checklist of the Hemiptera of Canada and Alaska. NRC Research Press, 220 pp. [ISBN 978-0-660-18165-3]
- Meynard C, Gay P, Lecoq M, Foucart A, Piou C, Chapuis M (2017) Climate-driven geographic distribution of the desert locust during recession periods: Subspecies' niche differentiation and relative risks under scenarios of climate change. Global Change Biology 23: 4739-4749. https://doi.org/10.1111/gcb.13739

- Meynard C, Leroy B, Kaplan D (2019) Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? Ecography 42: 2021-2036. https://doi.org/10.1111/ecog.04385
- Munyaneza JE (2012) Zebra chip disease of potato: biology, epidemiology, and management. American Journal of Potato Research 89: 329-350. https://doi.org/10.1007/s12230-012-9262-3
- Muscatello A, Elith J, Kujala H (2021) How decisions about fitting species distribution models affect conservation outcomes. Conservation Biology https://doi.org/10.1111/cobi.13669
- Narouei-Khandan HA, Halbert SE, Worner SP, van Bruggen AH (2016) Global climate suitability of citrus huanglongbing and its vector, the Asian citrus psyllid, using two correlative species distribution modeling approaches, with emphasis on the USA. European Journal of Plant Pathology 144: 655-670. https://doi.org/10.1007/s10658-015-0804-7
- Ouvrard D (2021) Psyl'list - The world Psylloidea database. http://www.hemiptera-databases.org/psyllist. Accessed on: 2021-3-29.
- Peccoud J, Labonne G, Sauvion N (2013) Molecular test to assign individuals within the *Cacopsylla pruni* complex. PLOS One 8: e72454. https://doi.org/10.1371/journal.pone.0072454
- Peccoud J, Pleydell DR, Sauvion N (2018) A framework for estimating the effects of sequential reproductive barriers: implementation using Bayesian models with field data from cryptic species. Evolution 72: 2503-2512. https://doi.org/10.1111/evo.13595
- Rwomushana I, Khamis FM, Grout TG, Mohamed SA, Stamou M, Borgemeister C, Heya HM, Tanga CM, Nderitu PW, Seguni ZS (2017) Detection of *Diaphorina citri* Kuwayama (Hemiptera: Liviidae) in Kenya and potential implication for the spread of Huanglongbing disease in East Africa. Biological Invasions 19: 2777-2787. https://doi.org/10.1007/s10530-017-1502-5
- Sabaté J, Lavina A, Batlle A (2016) Incidence and distribution of '*Candidatus* Phytoplasma prunorum' and its vector *Cacopsylla pruni* in Spain: an approach to the epidemiology of the disease and the role of wild *Prunus*. Plant Pathology 65: 837-846. https://doi.org/10.1111/ppa.12464
- Sauvion N, Lachenaud O, Genson G, Rasplus J-Y, Labonne G (2007) Are there several biotypes of *Cacopsylla pruni*? Bulletin of Insectology 60: 185-186. URL: http://www.bulletinofinsectology.org/pdfarticles/vol60-2007-185-186sauvion.pdf
- Sauvion N, Lachenaud O, Mondor-Genson G, Rasplus J-Y, Labonne G (2009) Nine polymorphic microsatellite loci from the psyllid *Cacopsylla pruni* (Scopoli), the vector of European stone fruit yellows. Molecular Ecology Resources 9: 1196-1199. https://doi.org/10.1111/j.1755-0998.2009.02604.x
- Seljak G (2020) Jumping plant-lice of Slovenia (Insecta: Hemiptera: Psylloidea). Scopolia 98: 1-224. URL: https://www.zobodat.at/pdf/Scopolia_98_0001-0224.pdf
- Shimwela MM, Narouei-Khandan HA, Halbert SE, Keremane ML, Minsavage GV, Timilsina S, Massawe DP, JB J, van Bruggen AH (2016) First occurrence of *Diaphorina citri* in East Africa, characterization of the *Ca.* Liberibacter species causing huanglongbing (HLB) in Tanzania, and potential further spread of *D. citri* and HLB in Africa and Europe. European Journal of Plant Pathology 146: 349-368. https://doi.org/10.1007/s10658-016-0921-y

- Steffek R, Swen Follak S, Sauvion N, Labonne G, MacLeod A (2012) Distribution of '*Candidatus* Phytoplasma prunorum' and its vector *Cacopsylla pruni* in European fruit growing areas: a literature survey. EPPO Bulletin 42: 191-202. https://doi.org/10.1111/epp.2567
- Suarez AV, Tsutsui ND (2004) The value of museum collections for research and society. BioScience 54: 66-74. https://doi.org/10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2
- Syfert MM, Serbina L, Burckhardt D, Knapp S, Percy DM (2017) Emerging new crop pests: ecological modelling and analysis of the south American potato psyllid *Russelliana solanicola* (Hemiptera: Psylloidea) and its wild relatives. PLOS One 12: e0167764. https://doi.org/10.1371/journal.pone.0167764
- Thébaud G, Sauvion N, Chadoeuf J, Dufils A, Labonne G (2006) Identifying risk factors for European stone fruit yellows from a survey. Phytopathology 96: 890-899. https://doi.org/10.1094/PHYTO-96-0890
- Thébaud G, Yvon M, Alary R, Sauvion N, Labonne G (2009) Efficient Transmission of '*Candidatus* Phytoplasma prunorum' is delayed by eight months due to a long latency in its host-alternating vector. Phytopathology 99: 265-273. https://doi.org/10.1094/PHYTO-99-3-0265
- Venette RC, Kriticos DJ, Magarey RD, Koch FH, Baker RH, Worner SP, Gomez Raboteaux NN, McKenney DW, Dobesberger EJ, Yemshanov D (2010) Pest risk maps for invasive alien species: a roadmap for improvement. BioScience 60: 349-362. https://doi.org/10.1525/bio.2010.60.5.5
- Warabieda W, Soika G, Ciesslinska M (2018) *Cacopsylla pruni* ir jo kaip '*Candidatus* Phytoplasma prunorum' pernejo svarba Lenkijoje [*Cacopsylla pruni* in Poland and its significance as a vector of 'C*andidatus* Phytoplasma prunorum']. Zemdirbyste-Agriculture 105: 177-182. URL: http://www.zemdirbyste-agriculture.lt/wp-content/uploads/2018/05/105_2_str23.pdf
- Wieczorek J, Bloom D, Guralnick R, Blum S, Dring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An evolving community-developed biodiversity data standard. PLOS One 7 (1): e29715. https://doi.org/10.1371/journal.pone.0029715
- Zhang Z- (2013) Phylum arthropoda. In: Zhang Z-Q (Ed.) Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness (Addenda 2013). Zootaxa 3703: 17-26. https://doi.org/10.11646/zootaxa.3703.1.6

Author-formatted, not peer-reviewed document posted on 20/05/2021. DOI: https://doi.org/10.3897/arphapreprints.e68864



М.М.Логинова

ПСИЛЛИДЫ (PSYLLOIDEA, HOMOPTERA) МОНГОЛЬСКОЙ НАРОДНОЙ РЕСПУБЛИКИ. II

M.M.Loginova. The psyllids (Psylloidea, Homoptera) of the Mongolian People's Republic.II

Psylla pruni (Scopoli, 1763).

Вост.: Хамар-Даба, 20 УІІ 1971 (М.Кз.); Дэрхийн-Цаган-Обо, 60 км ВСВ Баян-Бурда, 21 УІІ 1971 (К.). Всего 3 экз.

СССР: Молдавия, запад Украины, включая Закарпатье, Крым, Кавказ. Зап. Европа, кроме сев. Скандинавии. На различных косточковых рода Prunus: сливе, терне, альыче. Одно поколение в году, зимуют крылатые насекомые. В Европе часто вредит.

Figure 1.

Excerpt from a 1974 article from Loginova referring to *Cacopsylla pruni*, with translation and information about one of the localities cited, Hamar data. After Loginova (1974).

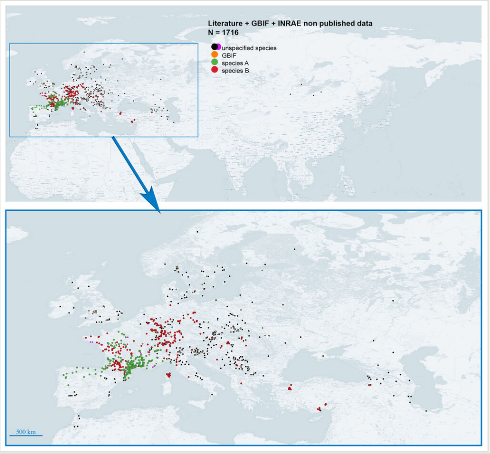


Figure 2.

Global map of the 1716 occurrence data available in the *C. pruni* dataset (map generated with QGIS 3.14). The map shows the distribution of cryptic species A (green dots) and B (red dots) according to available data. However, most of the data from the literature (black dots), GBIF (orange dots) or the Psylloidea catalogue of the "Faune de France" (currently being published) do not allow a distinction to be made between cryptic species.

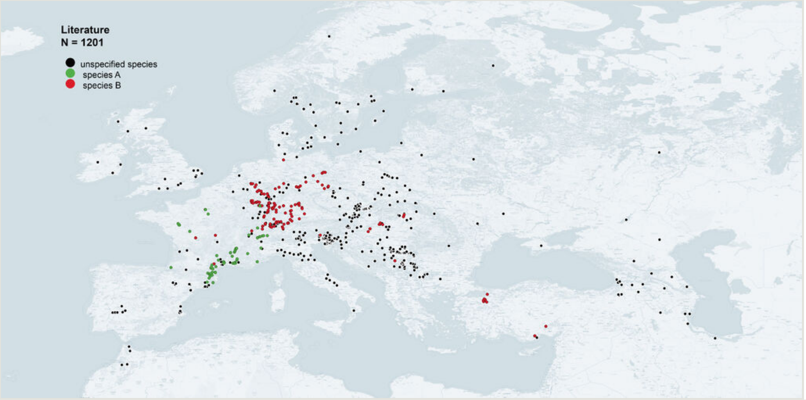


Figure 3.

Occurrence data of *Cacopsylla pruni* in the Western Palaearctic, obtained from our literature survey (map generated with QGIS 3.14).

Figure 4.

Examples of metadata accessible on the website of the Natural History Museum from links associated with GBIF references (e.g. https://www.gbif.org/occurrence/1265697015).

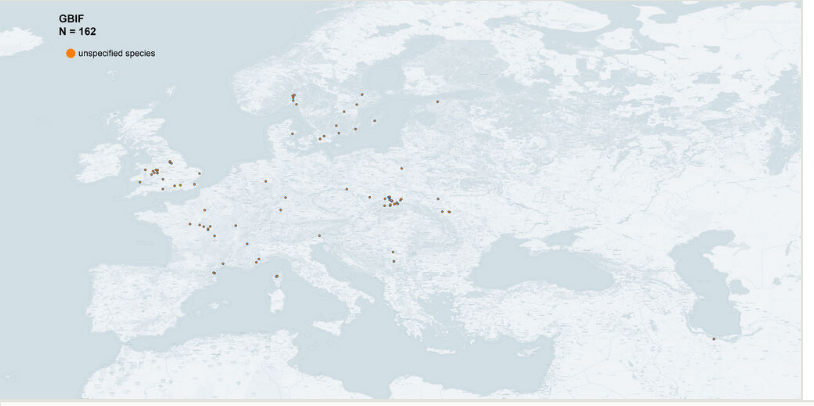


Figure 5.

Occurrence data of *Cacopsylla pruni* in Western Palaearctic from the GBIF database (map generated with QGIS 3.14).

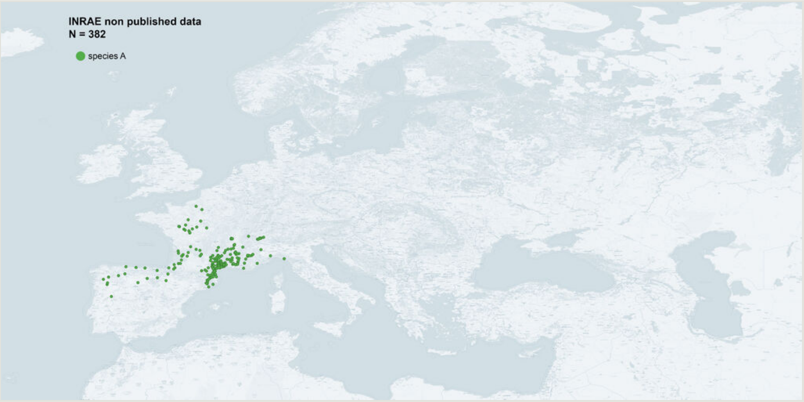


Figure 6.

Occurrence data of species of *Cacopsylla pruni* A in Western Palaearctic from sampling carried out by INRAE-Montpellier (map generated with QGIS 3.14).

Figure 7.

Occurrence data of species of *Cacopsylla pruni* B in Western Palaearctic from sampling carried out by INRAE-Montpellier (map generated with QGIS 3.14).