

# **In the shadow of medicine: the glaring absence of occurrence records of human-hosted biodiversity**

Rémy Poncet, Olivier Gargominy

Submitted to: Online Journal of Public Health Informatics  
on: May 02, 2024

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

*Table of Contents*

**Original Manuscript..... 4**  
**Supplementary Files..... 13**  
    Figures ..... 14  
        Figure 1..... 15  
        Figure 2..... 16

# In the shadow of medicine: the glaring absence of occurrence records of human-hosted biodiversity

Rémy Poncet<sup>1\*</sup> MSci, MPhil; Olivier Gargominy<sup>1\*</sup> MSci

<sup>1</sup>PatriNat (OFB, MNHN) Paris FR

\*these authors contributed equally

## Corresponding Author:

Rémy Poncet MSci, MPhil  
PatriNat (OFB, MNHN)  
36 rue Geoffroy-Saint-Hilaire  
Paris  
FR

## Abstract

Microbial diversity is vast, with bacteria playing a crucial role in human health. However, data on the locations (geographic occurrence) of bacteria associated with humans remain scarce. This lack of information hinders our understanding of human-microbe relationships and disease prevention. Here, we show that existing solutions, such as France's Système d'Information sur le Patrimoine Naturel (SINP) framework, can be used to efficiently collect and manage occurrence data on human-associated bacteria. This user-friendly system allows medical personnel to easily share and access data on bacterial pathogens. By implementing similar national infrastructures and considering human-associated bacteria as biodiversity data, we can significantly improve public health research and our understanding of the One Health concept, which emphasizes the interconnectedness of human, animal, and environmental health.

(JMIR Preprints 02/05/2024:60140)

DOI: <https://doi.org/10.2196/preprints.60140>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.  
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.  
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [http](#)

## Original Manuscript

**Viewpoint [or “Original paper”, this can be discussed, depending on what the Ed. board think, the manuscript can be adapted]**

## **In the shadow of medicine: the glaring absence of occurrence records of human-hosted biodiversity**

### **Abstract**

Microbial diversity is vast, with bacteria playing a crucial role in human health. However, data on the locations (geographic occurrence) of bacteria associated with humans remain scarce. This lack of information hinders our understanding of human-microbe relationships and disease prevention. Here, we show that existing solutions, such as France's Système d'Information sur le Patrimoine Naturel (SINP) framework, can be used to efficiently collect and manage occurrence data on human-associated bacteria. This user-friendly system allows medical personnel to easily share and access data on bacterial pathogens. By implementing similar national infrastructures and considering human-associated bacteria as biodiversity data, we can significantly improve public health research and our understanding of the One Health concept, which emphasizes the interconnectedness of human, animal, and environmental health.

**Keywords:** Human microbiome; Bacterial occurrence data; Public health; One Health; Biodiversity data gap; Medical data integration

### **Introduction**

Recent estimates put global diversity between 563 and 2,206 million species, with the bacterial kingdom being the most diverse at between 364 and 1,950 million species [1]. There are no synthetic data on the types of environments or (micro)habitats preferred by bacteria on a global scale, but to obtain an overview of the main environments colonized by microbial species (including bacterial, archaeal, and eukaryotic microbial species), we can use the three-level environment classification system associated with the 23,828 samples collected during the Earth Microbiome Project as a proxy [2]. These data revealed 52.32% free-living species (11.01% in saline environments and 88.99% in nonsaline environments) and 47.68% host-dependent species (20.16% of plants and 79.84% of animals). Of the animal-derived samples, 49.88% came from the gut, 32.64% from the body surface, 13.86% from secretions, and 3.62% from other parts of the body. Available data from other sources, such as MGnify [3], focus in an unbalanced way on certain biomes regarding host-dependent species (figures as of March 9, 2024): human ( $n = 249,568$ ), other mammals ( $n = 54,432$ ), plants ( $n = 46,148$ ), birds ( $n = 8,524$ ), and insects ( $n = 2,921$ ); however, von Meijenfeldt et al. [4] used a balanced subsample of these data to show that biome annotations correlated relatively well with taxonomic profiles (at the order level) associated with sequence data, suggesting the possible existence of a biome-phylogeny relationship (see also Malard & Guisan [5] for evidence of microbial niche differentiation).

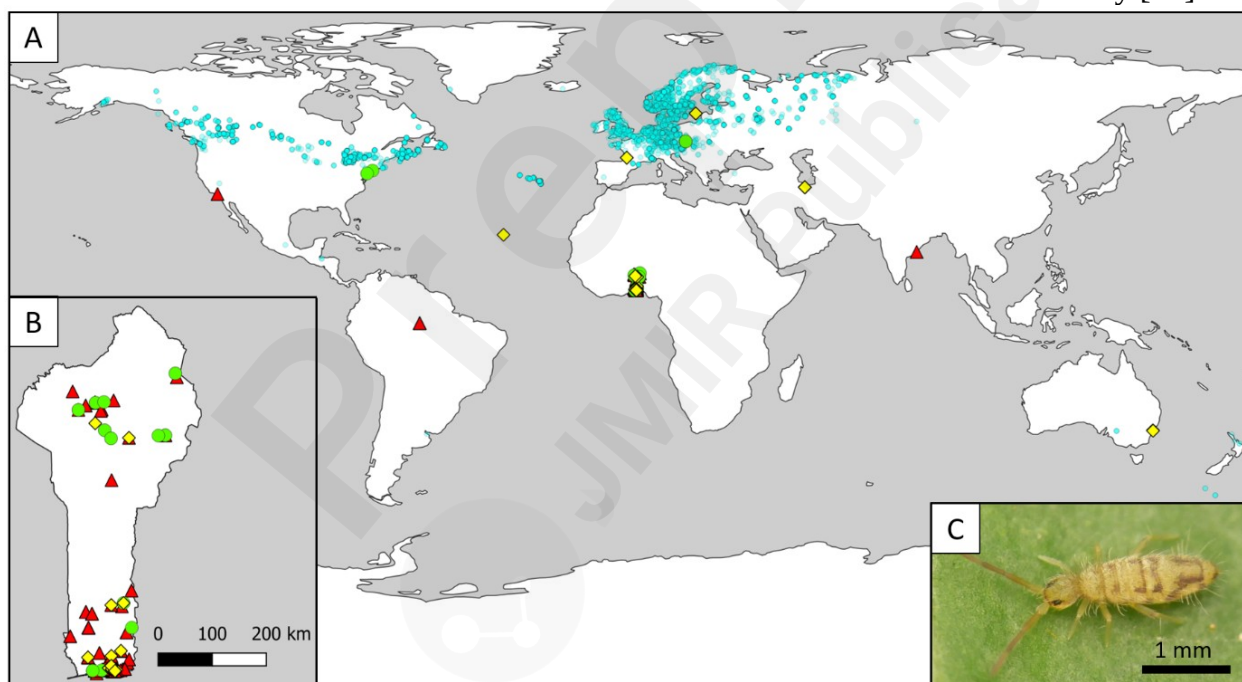
While the microbial diversity associated with the human species – *Homo sapiens* Linnaeus, 1758 – is the subject of large-scale studies (see, e.g., the Human Microbiome Project [6]), the data generated (mostly sequences) often lack taxonomic assignments or precise taxonomic assignments (e.g., to the species level) that would enable the relationships {microbe species} – {*Homo sapiens* Linnaeus, 1758} to be robustly documented. However, regarding human pathogenic bacteria, there are taxonomically assigned data at a specific level that allow these relationships to be established. For example, Bartlett et al. [7] listed 1,513 bacterial pathogens known to infect humans (based on the literature available before 2021). The list includes *Chlamydia trachomatis* (Busacca 1935) Rake

1957, *Neisseria gonorrhoeae* (Zopf 1885) Trevisan 1885, and *Treponema pallidum* (Schaudinn and Hoffmann 1905) Schaudinn 1905, three bacteria responsible for sexually transmitted infections, respectively: chlamydia, gonorrhea and syphilis. In 2020, an estimated 128.5 million people were infected with chlamydia [8], 82.4 million with gonorrhea [9], and 7.1 million with syphilis [10]. In a report dated March 7, 2024, the European Centre for Disease Prevention and Control (ECDC) reported a very significant increase in the number of cases of sexually transmitted infection in 2022 compared with that in the previous year, with the number of cases of gonorrhea increasing by 48% (70,881 confirmed cases) [11], the number of syphilis cases by 34% (35,391 confirmed cases) [12], and the number of chlamydia cases by 16% (216,508 confirmed cases) [13].

## Methods & Results [to be removed if the category is Viewpoint?]

### Chlamydia, gonorrhea, and syphilis are textbook study cases

The high prevalence of these three sexually transmitted infections at both the global and European levels is a textbook case for showcasing the degree of centralization and availability of occurrence records of bacterial species associated with humans. We therefore searched for occurrence records of these three infectious species in the Global Biodiversity Information Facility (GBIF), and the results were surprisingly low: 28 georeferenced records for *Chlamydia trachomatis* [14], 205 for *Neisseria gonorrhoeae* [15], and 17 for *Treponema pallidum* [16]. Figure 1 shows the distribution of georeferenced records available in the GBIF for the three bacteria and, by way of comparison, for the springtail species *Entomobrya nivalis* (Linnaeus, 1758) [17], which belongs to a taxonomic group that is still little studied worldwide and benefits from a reduced scientific community [18].



**Figure 1.** Available georeferenced records of three bacteria causing frequent sexually transmitted infections. A, global distribution; B, focus on Benin (which is the only country where occurrence records have been at least partially systematically collected between 2007 and 2020); C, the habitus of the springtail *Entomobrya nivalis* (photo credit Alexis Orion) [19]. *Chlamydia trachomatis* is marked with yellow diamonds; *Neisseria gonorrhoeae* is marked with red triangles; *Treponema pallidum* is marked with green circles; and *Entomobrya nivalis* is marked with small transparent turquoise circles (note: the display of occurrence records for this springtail species is intended to highlight the lack of records concerning the three bacteria, which are nevertheless frequent).

## The ecosystem for data collection already exists

The transmission and centralization of occurrence records on species (or at any other taxonomic level) presupposes the existence of at least three elements. The first concerns nomenclature, i.e., the ability to properly write taxon names, to unambiguously designate the taxonomic concept under consideration at the time of taxonomic assignment of a specimen (regardless of the method used for specimen taxonomical assignation). For bacteria, nomenclature is governed by the International Code of Nomenclature of Prokaryotes [20] as well as by SeqCode [21], depending on the taxonomic description protocol employed. From a practical point of view, although taxon names can be found in the literature in which they were originally described, it is also possible to find their name (taxon and authority) in various databases such as the List of Prokaryotic names with Standing in Nomenclature [22], the Taxonomy Database of the National Center for Biotechnology Information [23], or the Catalogue of Life [24]. The second is the existence of data standards that guarantee formatting and enable interoperability between information systems. The Darwin Core standard [25] can be used to record and vehiculate data on the occurrence of human pathogenic bacteria. In particular, the standard allows the management of associated occurrences (one occurrence of pathogenic bacteria and one occurrence of host species) with the term *dwc:associatedOccurrences* and to qualify this association (e.g., parasite of, host of) with the term *dwc:relationshipOfResource* (see ENETWILD consortium et al. [26] and Astorga et al. [27] for more details). The third is access to software/applications for banking and managing species records (based on, or compatible with, GBIF-accepted data standards). This dimension is undoubtedly the least obvious since depending on the situation (country, taxonomic group concerned, players involved, etc.) there may or may not be an easily accessible service on offer. When such infrastructures do not exist, solutions such as Symbiota [28], PlutoF [29], or EarthCape [30] can compensate for this lack. In addition, other optional elements can be added to these three dimensions to improve data reuse, in line with the FAIR principles [31] – Findability, Accessibility, Interoperability, and Reusability. For example, data dictionaries can be designed using internationally shared ontologies, such as the Earth Microbiome Project Ontology [32] or the European Molecular Biology Laboratory – European Bioinformatics Institute Ontology [33].

## An example of the French turnkey ecosystem

Despite the very small number of occurrence records on human infectious bacteria in France, the ecosystem of solutions provided by the Office français de la Biodiversité and the Muséum National d'Histoire Naturelle makes it straightforward to manage all the stages in the data cycle through the Système d'Information sur le Patrimoine Naturel (SINP) framework, from data acquisition to distribution on the GBIF. Figure 2 shows the main components of the SINP involved:

- TAXREF, the national taxonomic repository [34] forms the backbone of the system, provided via the TAXREF-web application [35]: *i.* a taxonomic management application (taxon management, links with external Global Species Databases, links with Docs-Web [36] the bibliographic reference management system), *ii.* a taxon status management application (official presence status in the territories (e.g., taxon is present, absent, extinct, etc.), and conservation and regulatory status (International Union for Conservation of Nature red list evaluations, protected species, etc.)), *iii.* a taxon trait management application (e.g., free-living species, host-dependent species), *iv.* a taxon interaction management application (e.g., *Chlamydia trachomatis* is an endoparasite of *Homo sapiens*), and *v.* a taxon relation to the natural and seminatural habitats management application (based on the HABREF national repository [37]).
- CardObs application [38, 39] is designed to capture and manage taxon occurrence data. It is linked to the Metadata management application [40], which can be used to create datasets and declare parameters such as the data acquisition protocols. It allows the creation of stations



and the recording of one or more taxon occurrences for each of them. It is possible to establish relationships between taxa (e.g., *Chlamydia trachomatis* was collected from *Homo sapiens*) and to link a record to a voucher or a sequence accession ID in GenBank [41] or Barcode Of Life Data System [42]. Once entered, the data comply with the exchange standard for taxon observation and monitoring data [43] and undergo control procedures [44–46] before being entered into the Inventaire National du Patrimoine Naturel (INPN) [47] (data accessible on the OpenObs platform [48]) and then into the GBIF.

In short, for medical or laboratory staff who identify the presence of human parasitic bacteria in a sample, they must log on to CardObs and enter the following minimum information: station (e.g., patient's city of residence), date, data collector, taxon (e.g., bacterial species), related species (e.g., *Homo sapiens*), and type of link between the two species (e.g., bacteria is a pathogen of *Homo sapiens*). This entry, which takes just one or two minutes, would help to fill the gap in occurrence data on bacteria linked to humans in a very simple and efficient way.



**Figure 2.** Overview of the services available in France to bank and make medical data of interest available for understanding human-related biodiversity. TAXREF feeds CardObs with the official list of taxa and associated reference information. CardObs is connected to the Metadata application, making it easy to select the dataset in which a record is entered. The CardObs information system complies with the official data model for taxon observations (GBIF-compatible). Once entered, the data undergo conformity (integrity) and qualification tests (level of validity of the data – procedure variable according to the reference systems available), and if they are compliant, they are uploaded to the INPN and then to the GBIF.

## Discussion

The lack of banking and dissemination of records of human-hosted species (bacteria, parasites, etc.) suggests three facts: *i.* the lack of consideration of the human species as an ecosystem hosting biodiversity, *ii.* the failure to consider data generated in a medical context as data of interest regarding biodiversity, environment, and the One Health concept, and *iii.* the absence of a policy (or any other kind of incentive) for the systematic banking of these data (occurrence, species-species



relationship), enabling them to be made available to the scientific community and society. Awareness of the benefits of banking such data to make them accessible – made possible by the existence of an ecosystem of solutions addressing the entire data cycle – would support research and prevention in the field of public health and the human-environment relationship (at all scales, including the *human ecosystem*). Indeed, it is worth noting that in the ECDC reports on the three pathogenic bacteria used in the study case, France, Germany, Belgium, the Netherlands, Austria, and Italy were unable to systematically transfer summary data for the period 2018-2022. In addition, these three bacteria are known to interact with other species, e.g., House mouse (*Mus musculus* Linnaeus, 1758), Long-tailed Macaque (*Macaca fascicularis* (Raffles, 1821)), Wild Boar (*Sus scrofa* Linnaeus, 1758) for *Chlamydia trachomatis* [49], House mouse (*Mus musculus*) for *Neisseria gonorrhoeae* [50], and several monkey species for *Treponema pallidum* [51]. Host species of these pathogenic bacteria for which more than 500,000 occurrence data are available in the GBIF [52]. Thus, in line with the implementation of the One Health concept [53-55] and in compliance with the provisions of the Aarhus Convention [56] (in particular, articles 4. *Access to environmental information*, and 5. *Collection and dissemination of environmental information*), encouraging the banking and availability of biodiversity data generated in the medical sector would be a real step forward in the service of public health, biodiversity research, and society.

## Acknowledgements

The authors are thankful to Dr. Bastiaan von Meijenfildt for his helpful insights and review of the section dealing with the ecological preferences of bacteria.

## Conflicts of Interest

None declared.

## Abbreviations

ECDC: European Centre for Disease Prevention and Control

GBIF: Global Biodiversity Information Facility

SINP: Système d'INformation sur le Patrimoine naturel (=Natural Heritage Information System)

## References

1. Li X, Wiens JJ. Estimating Global Biodiversity: The Role of Cryptic Insect Species. *Systematic Biology*. 2023;72(2):391-403. <https://doi.org/10.1093/sysbio/syac069>
2. Thompson L et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*. 2017;551:457-463. <https://doi.org/10.1038/nature24621>
3. Richardson L et al. MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Research*. 2022;51(D1):753-759. <https://doi.org/10.1093/nar/gkac1080>
4. von Meijenfildt FAB, Hogeweg P, Dutilh BE. A social niche breadth score reveals niche range strategies of generalists and specialists. *Nature Ecology & Evolution*. 2023;7:768-781. <https://doi.org/10.1038/s41559-023-02027-7>
5. Malard LA, Guisan A. Into the microbial niche. *Trends in Ecology & Evolution*. 2023;38(10):936-945. <https://doi.org/10.1016/j.tree.2023.04.015>
6. Gevers D et al. The Human Microbiome Project: a community resource for the healthy human microbiome. *PLoS biology*. 2012;10(8):e1001377. <https://doi.org/10.1371/journal.pbio.1001377>
7. Bartlett A, Padfield D, Lear L, Bendall R, Vos M. A comprehensive list of bacterial pathogens infecting humans. *Microbiology*. 2022;168(2):001269. <https://doi.org/10.1099/mic.0.001269>
8. WHO. Chlamydia. Updated on 17 July 2023. 2023. URL:

- <https://www.who.int/news-room/fact-sheets/detail/chlamydia> [accessed 2024-03-10]
9. WHO. Gonorrhoea (*Neisseria gonorrhoeae* infection). Updated on 18 July 2023. 2023. URL: [https://www.who.int/news-room/fact-sheets/detail/gonorrhoea-\(neisseria-gonorrhoeae-infection\)](https://www.who.int/news-room/fact-sheets/detail/gonorrhoea-(neisseria-gonorrhoeae-infection)) [accessed 2024-03-10]
  10. WHO. Syphilis. Updated on 31 May 2023. 2023. URL: <https://www.who.int/news-room/fact-sheets/detail/syphilis> accessed 10 March 2024 [accessed 2024-03-10]
  11. ECDC. Gonorrhoea. In: ECDC. Annual epidemiological report for 2022. Stockholm; 2024. URL: [https://www.ecdc.europa.eu/sites/default/files/documents/GONO\\_AER\\_2022\\_Report%20FINAL.pdf](https://www.ecdc.europa.eu/sites/default/files/documents/GONO_AER_2022_Report%20FINAL.pdf) [accessed 2024-03-10]
  12. ECDC. Syphilis. In: ECDC. Annual Epidemiological Report for 2022. Stockholm; 2024. URL: [https://www.ecdc.europa.eu/sites/default/files/documents/SYPH\\_AER\\_2022\\_Report\\_0.pdf](https://www.ecdc.europa.eu/sites/default/files/documents/SYPH_AER_2022_Report_0.pdf) [accessed 2024-03-10]
  13. ECDC. Chlamydia. In: ECDC. Annual Epidemiological Report for 2022. Stockholm;2024. URL: [https://www.ecdc.europa.eu/sites/default/files/documents/CHLAM\\_AER\\_2022\\_Report.pdf](https://www.ecdc.europa.eu/sites/default/files/documents/CHLAM_AER_2022_Report.pdf) [accessed 2024-03-10]
  14. GBIF. GBIF Occurrence Download (*Chlamydia trachomatis*), 10 March 2024. 2024. URL: <https://doi.org/10.15468/dl.judfc3> [accessed 2024-03-10]
  15. GBIF. GBIF Occurrence Download (*Neisseria gonorrhoeae*), 10 March 2024. 2024. URL: <https://doi.org/10.15468/dl.2bnevx> [accessed 2024-03-10]
  16. GBIF. GBIF Occurrence Download (*Treponema pallidum*), 10 March 2024. 2024. URL: <https://doi.org/10.15468/dl.49rn3g> [accessed 2024-03-10]
  17. GBIF. GBIF Occurrence Download (*Entomobrya nivalis*), 10 March 2024. 2024. URL: <https://doi.org/10.15468/dl.3cnygb> [accessed 2024-03-10]
  18. Potapov A. et al. Towards a global synthesis of Collembola knowledge – challenges and potential solutions. Soil Organisms. 2020;92(3):161-188. <https://doi.org/10.25674/so92iss3pp161>
  19. Orion A. Photo 42198057. 2019. URL: <https://uk.inaturalist.org/photos/42198057> [accessed 2024-03-10]
  20. Oren A. et al. International Code of Nomenclature of Prokaryotes. Prokaryotic Code (2022 Revision). International Journal of Systematic and Evolutionary Microbiology. 2023;73(5a):005585. <https://doi.org/10.1099/ijsem.0.005585>
  21. Hedlund BP et al. SeqCode: a nomenclatural code for prokaryotes described from sequence data. Nature Microbiology. 2022;7:1702-1708. <https://doi.org/10.1038/s41564-022-01214-9>
  22. LPSN. List of Prokaryotic names with Standing in Nomenclature. URL: <https://lpsn.dsmz.de> [accessed 2024-03-10]
  23. NCBI. Taxonomy. URL: <https://www.ncbi.nlm.nih.gov/taxonomy> [accessed 2024-03-10]
  24. Catalogue of Life. The most complete authoritative list of the world's species - maintained by hundreds of global taxonomists. URL: <https://www.catalogueoflife.org> [accessed 2024-03-10]
  25. Wiecek J et al. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS One. 2012;7(1):e29715. <https://doi.org/10.1371/journal.pone.0029715>
  26. ENETWILD consortium, Jaroszynska F, Body G, Pamerlon S, Archambeau A-S. Applying the Darwin Core data standard to wildlife disease – advancements toward a new data model. EFSA supporting publication; 2022. <https://doi.org/10.2903/sp.efsa.2022.EN-7667>
  27. Astorga F et al. Biodiversity data supports research on human infectious diseases: Global trends, challenges, and opportunities. One Health. 2023;16:100484.

- <https://doi.org/10.1016/j.onehlt.2023.100484>
28. Symbiota. Open-Source Biodiversity Data Management Software. URL: <https://symbiota.org> [accessed 2024-03-10]
  29. PlutoF. Common Data Space for Biodiversity. URL: <https://plutof.ut.ee> [accessed 2024-03-10]
  30. EarthCape. Biodiversity Database Platform. URL: <https://earthcape.github.io> [accessed 2024-03-10]
  31. Wilkinson M et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016;3:160018. <https://doi.org/10.1038/sdata.2016.18>
  32. Earth Microbiome Project. EMP Ontology (EMPO). URL: <https://earthmicrobiome.org/protocols-and-standards/emp> [accessed 2024-03-10]
  33. European Bioinformatics Institute. EMBL-EBI Ontology Lookup Service. URL: <https://www.ebi.ac.uk/ols4> [accessed 2024-03-10]
  34. TAXREF Eds. TAXREF v17.0, référentiel taxonomique pour la France. PatriNat (OFB-CNRS-MNHN), Muséum national d'Histoire naturelle. Archive de téléchargement contenant 8 fichiers générés le 10 janvier 2024. Paris;2024. URL: <https://inpn.mnhn.fr/telechargement/referentielEspece/taxref/17.0/menu> [accessed 2024-03-10]
  35. TAXREF Eds. TAXREF-web – TAXREF, référentiel taxonomique national des organismes de France métropolitaine et outre-mer. URL: <https://taxref.mnhn.fr/taxref-web/accueil> [accessed 2024-03-10]
  36. INPN. Docs-Web, bibliothèque électronique de documents. URL: <https://inpn.mnhn.fr/docs-web/accueil> [accessed 2024-03-10]
  37. Gaudillat V, Andres S, Decherf B, La Rivière M, Vallez E. HabRef v7.0, référentiel des typologies d'habitats et de végétation pour la France. Guide méthodologique. PatriNat (OFB-MNHN-CNRS-IRD), Paris;2024. URL: [https://inpn.mnhn.fr/docs/ref\\_habitats/HABREF\\_7.0/Guide\\_methodologique\\_HABREF\\_V7.0.pdf](https://inpn.mnhn.fr/docs/ref_habitats/HABREF_7.0/Guide_methodologique_HABREF_V7.0.pdf) [accessed 2024-03-10]
  38. Gargominy O. Guide d'utilisation de CardObs - v3.0. Service du Patrimoine naturel, Paris;2013. URL: <https://inpn.mnhn.fr/docs-web/docs/download/231189> [accessed 2024-03-10]
  39. MNHN, PatriNat (OFB-MNHN-CNRS-IRD). CardObs, Saisie et gestion de données naturalistes. URL: <https://cardobs.mnhn.fr> [accessed 2024-03-10]
  40. INPN. SINP – Métadonnées. URL: <https://inpn.mnhn.fr/mtd/cadre> [accessed 2024-03-10]
  41. National Center for Biotechnology Information. GenBank. URL: <https://www.ncbi.nlm.nih.gov/genbank> [accessed 2024-03-10]
  42. BoldSystems. Barcode Of Life Data System V4. URL: <https://www.boldsystems.org> [accessed 2024-03-10]
  43. Vinet C et al. Standard d'échange pour les données d'observations et de suivis de taxons v3.0. PatriNat (OFB-MNHN-CNRS-IRD), Paris;2023. URL: <https://inpn.mnhn.fr/docs-web/docs/download/421824> [accessed 2024-03-10]
  44. Robert S et al. Guide méthodologique pour la conformité, la cohérence et la validation scientifique des données et des métadonnées du SINP – Volet 1 : occurrences de taxons, Version 1. Rapport pour le SINP, rapport MNHN-SPN 2016-77, Paris;2016. URL: <https://inpn.mnhn.fr/docs-web/docs/download/247411> [accessed 2024-03-10]
  45. Robert S et al. Procédure nationale de validation scientifique des données élémentaires d'échange du SINP pour les occurrences de taxons. Version 1. Service du patrimoine naturel, Muséum national d'histoire naturelle. Rapport SPN, Paris;2017. URL: <https://inpn.mnhn.fr/docs-web/docs/download/247445> [accessed 2024-03-10]

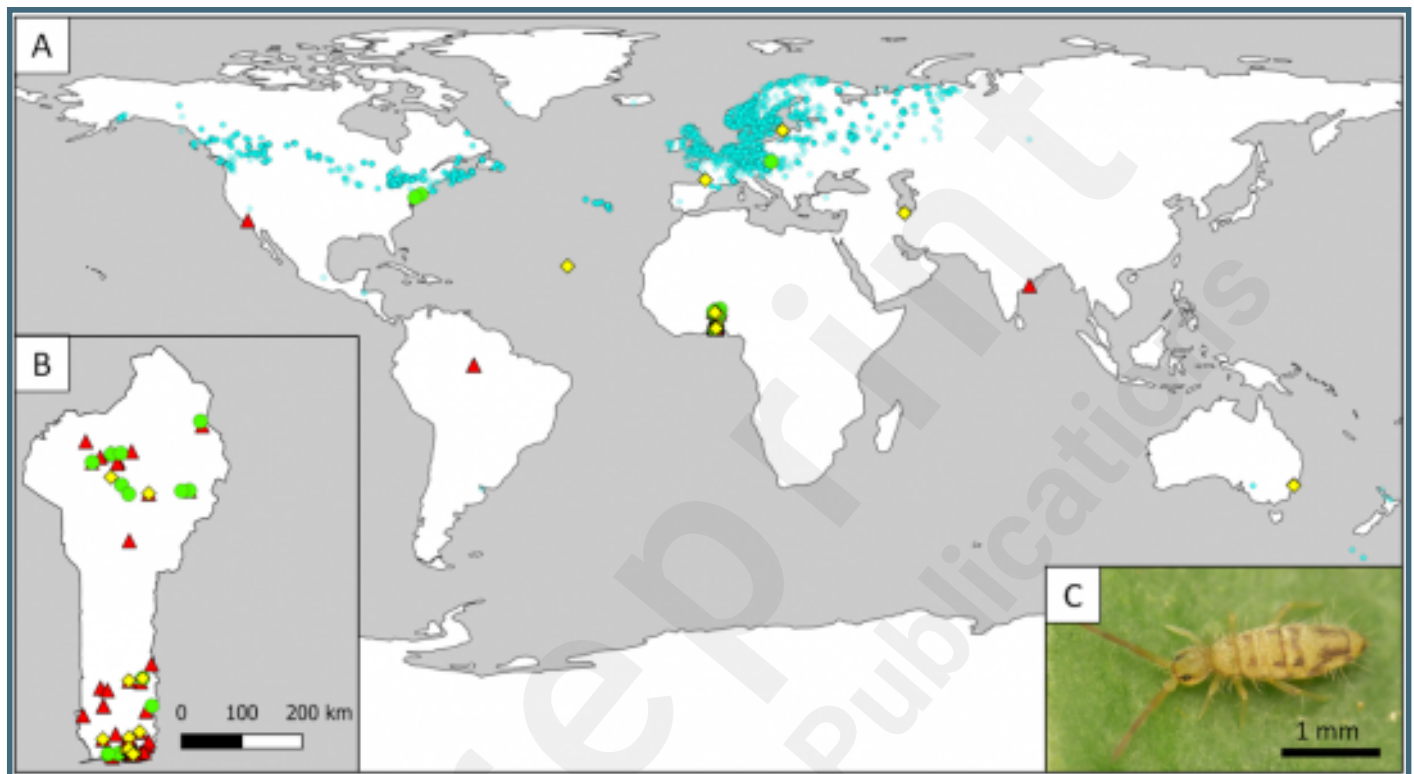
46. Robert S et al. Procédure de conformité et de cohérence des données et des métadonnées circulant entre les plateformes du SINP - Thématique : Occurrences de taxon, Version 1.0. Rapport pour le SINP. UMS 2006 Patrinat (AFB/CNRS/MNHN), Paris;2017. URL: <https://inpn.mnhn.fr/docs-web/docs/download/230990> [accessed 2024-03-10]
47. INPN. Inventaire National du Patrimoine Naturel. URL: <https://inpn.mnhn.fr> [accessed 2024-03-10]
48. INPN. OpenObs, Portail français d'accès aux données d'observation sur les espèces. URL: <https://openobs.mnhn.fr> [accessed 2024-03-10]
49. GloBi. What kind of organisms do *Chlamydia trachomatis* interacts with? 2024. URL: <https://www.globalbioticinteractions.org/?interactionType=interactsWith&sourceTaxon=Chlamydia%20trachomatis> [accessed 2024-03-11]
50. GloBi. What kind of organisms do *Neisseria gonorrhoeae* interacts with? 2024. URL: <https://www.globalbioticinteractions.org/?interactionType=interactsWith&sourceTaxon=Neisseria%20gonorrhoeae> [accessed 2024-03-11]
51. GloBi. What kind of organisms do *Treponema pallidum* interacts with? 2024. URL: <https://www.globalbioticinteractions.org/?interactionType=interactsWith&sourceTaxon=Treponema%20pallidum> [accessed 2024-03-11]
52. GBIF. GBIF Occurrence Download (*Macaca fascicularis*, *Mus musculus*, *Sus scrofa*), 11 March 2024. 2024. URL: <https://doi.org/10.15468/dl.qz93js> [accessed 2024-03-11]
53. Schwabe CW. Veterinary Medicine and Human Health, 3rd edition. Baltimore, Williams and Wilkins; 1984. ISBN:0683075942.
54. WSC. The Manhattan Principles. 2004. URL: <https://oneworldonehealth.wcs.org/About-Us/Mission/The-Manhattan-Principles.aspx> [accessed 2024-03-11]
55. Gruetzmacher K, et al. The Berlin principles on one health – Bridging global health and conservation. Science of The Total Environment. 2021;764:142919. <https://doi.org/10.1016/j.scitotenv.2020.142919>
56. UNECE. Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters (Aarhus Convention). Aarhus, Denmark: United Nations Economic Commission for Europe; 1998. URL: <https://unece.org/environment-policy/public-participation/aarhus-convention/text> [accessed 2024-03-11]

## Supplementary Files

## Figures



Available georeferenced records of three bacteria causing frequent sexually transmitted infections. A, global distribution; B, focus on Benin (which is the only country where occurrence records have been at least partially systematically collected between 2007 and 2020); C, the habitus of the springtail *Entomobrya nivalis* (photo credit Alexis Orion) [19]. *Chlamydia trachomatis* is marked with yellow diamonds; *Neisseria gonorrhoeae* is marked with red triangles; *Treponema pallidum* is marked with green circles; and *Entomobrya nivalis* is marked with small transparent turquoise circles (note: the display of occurrence records for this springtail species is intended to highlight the lack of records concerning the three bacteria, which are nevertheless frequent).



Overview of the services available in France to bank and make medical data of interest available for understanding human-related biodiversity. TAXREF feeds CardObs with the official list of taxa and associated reference information. CardObs is connected to the Metadata application, making it easy to select the dataset in which a record is entered. The CardObs information system complies with the official data model for taxon observations (GBIF-compatible). Once entered, the data undergo conformity (integrity) and qualification tests (level of validity of the data – procedure variable according to the reference systems available), and if they are compliant, they are uploaded to the INPN and then to the GBIF.

