



Final PhD Thesis Report

Codes4strains

Tracking bacterial pathogens through sources, geography and time using stable phylogenetically informative genome codes.

Responsible OHEJP Partner: Institut Pasteur



GENERAL INFORMATION

European Joint Programme full title	Promoting One Health in Europe through joint actions on foodborne zoonoses, antimicrobial resistance and emerging microbiological hazards.
European Joint Programme acronym	One Health EJP
Funding	This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.
Grant Agreement	Grant agreement n° 773830
Starting Date	01/01/2018
Duration	60 Months

DOCUMENT MANAGEMENT

PhD Report	Final PhD Thesis Report Y5 (2022)
PhD Reference	PhD16 - Codes4strains
PhD candidate	HENNART Mélanie
PhD Lead Supervisor	BRISSE Sylvain (Institut Pasteur)
PhD second supervisor	CRISCUOLO Alexis (Institut Pasteur)
Other supervisor (s)	
Due month of the deliverable	M66
Actual submission month	MXX (to be completed by the WP6 team)
Type <i>R: Document, report</i> <i>DEC: Websites, patent filings, videos, etc.; OTHER</i>	R
Dissemination level <i>PU: Public (default)</i> <i>CO: confidential, only for members of the consortium (including the Commission Services).</i>	PU <i>This is the default setting. This report will be entirely copied into One Health EJP Work Package 6 Deliverable D6.18 - Thesis Reports of up to 17 PhD studentships.</i>
Dissemination <i>Author's suggestion to inform the following possible interested parties.</i>	<div style="display: flex; flex-wrap: wrap;"> <div style="width: 33%;">OHEJP WP 1 <input type="checkbox"/></div> <div style="width: 33%;">OHEJP WP 2 <input type="checkbox"/></div> <div style="width: 33%;">OHEJP WP 3 <input type="checkbox"/></div> <div style="width: 33%;">OHEJP WP 4 <input type="checkbox"/></div> <div style="width: 33%;">OHEJP WP 5 <input type="checkbox"/></div> <div style="width: 33%;">OHEJP WP 6 <input type="checkbox"/></div> <div style="width: 33%;">OHEJP WP 7 <input type="checkbox"/></div> <div style="width: 33%;">Project Management Team <input checked="" type="checkbox"/></div> <div style="width: 33%;">Communication Team <input type="checkbox"/></div> <div style="width: 33%;">Scientific Steering Board <input type="checkbox"/></div> <div style="width: 33%;">National Stakeholders / Program Owners Committee <input type="checkbox"/></div> <div style="width: 33%;">EFSA <input type="checkbox"/> ECDC <input type="checkbox"/> EEA <input type="checkbox"/> EMA <input type="checkbox"/> FAO <input type="checkbox"/> WHO <input type="checkbox"/> OIE <input type="checkbox"/></div> </div> Other international stakeholder(s): Social Media: Other recipient(s):



1. PhD Project team composition

PhD student:

Mélanie HENNART – Sorbonne Université, Collège doctoral, F-75005 Paris, France

Lead PhD Supervisor:

Sylvain BRISSE – Institut Pasteur, Université Paris Cité, Biodiversity and Epidemiology of Bacterial Pathogens, Paris, F-75015, France

Second Supervisor:

Alexis CRISCUOLO – Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, F-75015 Paris, France

2. Abstract

Infectious diseases are a global public health concern, particularly due to antimicrobial-resistance in some pathogenic bacteria. *Klebsiella pneumoniae* is one of the most worrying multiresistant bacteria. *Corynebacterium diphtheriae*, which causes diphtheria, remains largely susceptible to first-line antibiotics, including penicillin, and can be controlled through vaccination, but re-emerges when vaccination coverage is insufficient. Among the effective infection control measures, the accurate detection and identification of these pathogens, as well as their epidemiological monitoring, play a key role. In the recent years, the implementation of whole-genome sequencing (WGS) has revolutionised bacterial genotyping, by providing discrimination at the strain level. Genomic sequencing also enables the detection of variants and their important characteristics, such as virulence or antimicrobial resistance.

The research work of this thesis is structured around two main axes.

The first axis provides bioinformatic analyses of the population structure of antimicrobial resistance in *C. diphtheriae*. Based on a large clinical isolates collection from metropolitan France and its overseas territories, a genome-wide association study (GWAS) was performed to determine the genetic basis behind resistance phenotypes, as well as their associations with diphtheria toxin production and other strain characteristics. A new penicillin resistance gene was discovered on a mobile element in *C. diphtheriae*. A genotyping bioinformatics tool, DIPHTOSCAN, was developed, to harmonize and facilitate the detection and genotyping of the main virulence factors and resistance genes of *C. diphtheriae*. We also developed strain nomenclatures from assembled genomes. DIPHTOSCAN further enables the prediction of biovars and of the toxicity of strains. The public availability of the tool (<https://gitlab.pasteur.fr/BEBP/diphtoscan>) and its ease of use will advance the genomic epidemiology of the agent of diphtheria, the clinical management of patients and knowledge on the links between animal and human diphtheria cases (particularly those caused by *C. ulcerans*). This thesis also advanced knowledge on the genotype-clinical phenotype links in *C. diphtheriae*.

The second axis relates to infra-species genomic taxonomy. A new approach of genome-based classification and nomenclature of strains was developed using *K. pneumoniae* as a model. This work describes the design and implementation of a barcoding system that uses *Life Identification Number* (LIN) codes based on the *core-genome* MLST (cgMLST) typing approach. This innovative taxonomic strategy provides precise and stable strain nomenclatures and was applied to the ‘one health’ model *Klebsiella pneumoniae* species complex, a ubiquitous group of pathogens. The proposed unified taxonomy of *K. pneumoniae* strains will facilitate advances on the biology of its sublineages across niches, time and space, and will endow surveillance networks from different activity sectors, including food microbiology and veterinary and human medicine, with the capacity to efficiently monitor and



control the emergence of sublineages of high public health relevance.

Based on the contributions and concepts developed in the two above axes, several case studies were carried out: identification and characterisation of a new species (*C. rouxii*), which was previously misidentified as *C. diphtheriae*; genomic epidemiology studies of diphtheria in different world regions from animal and human clinical sources; showing that *C. rouxii* probably represents a novel zoonotic pathogen. These applications of genomic taxonomy in combination with antimicrobial resistance gene detection illustrate the potential of the methods and tools developed during this thesis to support genomic research and surveillance of pathogenic bacteria in a one health perspective.

The work of this thesis has shed light on important questions in the biology of *C. diphtheriae* and *K. pneumoniae* populations and has developed concepts and tools that meet the modern needs of medical microbiology and public health actors. The speed and reduced costs of sequencing offer the possibility for microbiology laboratories to have access to this new approach in routine. Their exploitation and interpretation with standardized and automated approaches, such as the ones developed in this thesis, will facilitate analysis of genomic sequence data, thus contributing, hopefully, to a better control of infectious diseases.

Keywords: pathogen tracking, strain nomenclature, genomic classification, antibiotic resistance, epidemiological surveillance

3. Introduction and Objectives

4. Materials and Methods

5. Scientific Results and Discussion

6. Conclusions and Future work and perspectives

} Thesis

7. PhD project self-evaluation

The research project has been largely and successfully completed, as we have addressed most of the objectives stated in the initial work plan. We developed an innovative genomic nomenclature and typing approach called cgLIN codes, which combines the advantages of the cgMLST approach (reproducibility, standardization) with the LIN (Life identification numbers) code approach (complete stability). LIN codes are attributed to each genome, and the code conveys a notion of phylogenetic proximity with other strains of the same species, based on the number of different alleles between the cgMLST profiles of the genomes.

A slight deviation from the initial plan is that we did not include the SNP address approach in the PhD work, even though it was originally planned. We did conduct a comparison between the cgMLST and SNP address approaches, and the results showed that the methods were not easily comparable. Besides, the SNP address approach has not been largely implemented (to our knowledge, only PHE/UKSHA uses it, on a restricted number of pathogens). Finally, our partner in charge of SNP address developments changed job and moved from PHE/UKSHA to the Netherlands, meaning he had no access anymore to the comparative databases. For these reasons, it was decided not to pursue the SNP address comparative evaluation.

To compensate, we used a similar approach to the SNP address, a multi-level grouping system based on cgMLST profiles (instead of SNPs in the SNP address approach), which allowed us to create a multilevel taxonomic barcode system. We also made sure that this new nomenclature approach would



be linked to existing nomenclature systems, such as the 7-gene MLST, to make it backwards-compatible with this widely used language, and to facilitate adoption of the genomic nomenclature.

It should be noted that, even if this was not planned initially within this PhD work, our LIN code method was implemented and made available on the BIGSdb platform (<https://bigsdb.pasteur.fr/>) as well as in PathogenWATCH (<https://pathogen.watch/>), two prominent genomic epidemiology and strain taxonomy global platforms (<https://bigsdb.pasteur.fr/klebsiella/cgmlst-lincodes/> ; <https://cgps.gitbook.io/pathogenwatch/technical-descriptions/typing-methods/klebsiella-lin-codes>), allowing the scientific community to access and use it for other bacteria. These two achievements are very significant additional deliverables.

Finally, although our initial work plan aimed to apply this novel approach to *Escherichia coli*, we were unable to do so as our PHE partner moved away from *E. coli* surveillance. However, we largely compensated this loss by studying and applying our approaches to another pathogen, *Corynebacterium diphtheriae*, on which significant work was done during this PhD (see abstract and below).

Overall, this PhD project resulted in 10 scientific publications and the development of two important bioinformatics tools (DIPHTOSCAN and LIN codes) that are already being used by the scientific community.



8. Progress of the project: milestones and deliverables

Deliverables

PhD reference	PhD Project deliverable number	Deliverable name (Original name, if different from the actual one)	Delivery date from AWP (month)	Date delivered (month)	Comments	Integrative categories*
PhD16-FBZ2/AMR6.1-Codes4strains	D-PhD16-2.1	Pilot Genome set	M24	M24	https://doi.org/10.5281/zenodo.4471354	
	D-PhD16-3.1	cgMLST schemes for Kp and Ec	M36	M36	https://doi.org/10.5281/zenodo.4471354	
	D-PhD16-3.2	LINcodes algorithm defined and implemented on full dataset	M30	M30	https://doi.org/10.5281/zenodo.5284876	
	D-PhD16-3.3	SNapperDB Ec implemented for full dataset	M33	M33	https://doi.org/10.5281/zenodo.4471354	
	D-PhD16-3.4	SNapperDB Kp implemented for full dataset	M36		Not realized; We did not create SNapperDB for the pathogen Kp.	
	D-PhD16-4.1	Simulated dataset analysed	M42		Not realized; 7,000 public genomes were sufficient to constitute an adequate real genomic dataset.	
	D-PhD16-4.2	Publication on LINcode approach and comparison with cgMLST and SNP address approaches	M48	M48	https://doi.org/10.5281/zenodo.5284876	
	D-PhD16-4.3	A novel tool to define the population structure of species, called MSTclust		M40	https://doi.org/10.5281/zenodo.5284876	
	D-PhD16-4.4	Optimization of our inheritance algorithm to provide backwards compatibility of MLST groups with 7-gene MLST		M42	https://doi.org/10.5281/zenodo.5284876	
	D-PhD16-4.5	Comparison of the ANI metric with the cgMLST metric		M42	https://doi.org/10.5281/zenodo.5284876	
	D-PhD16-4.6	Designed a method to identify recombination in cgMLST		M40	https://doi.org/10.5281/zenodo.5284876	
	D-PhD16-4.7	Integration of MLST identifiers into the BIGSdb platform to make them publicly available		M42	https://doi.org/10.5281/zenodo.5284876	
	D-PhD16-5.1	Method implemented in partners' labs.	M54	M56	https://scholar.google.com/scholar?start=0&hl=fr&as_sdt=2005&sciodt=0,5&cites=16109542483182874956&scipsc=	
	D-PhD16-5.2	PhD viva	M57	M59	16 November 2022	

* Categories of Integrative activities : 1. Design and implementation of surveillance and control activities; 2. Harmonised protocols and applied best practice; 3. Databases of reference materials and data, incl. metadata; 4. Standardised data formats, aligned data analysis for interpretation of surveillance data; 5. Sharing and communication of surveillance data; 6. Sharing of best intervention activities ; 7. Prevention: aligned use of facilities and models; 8. Other (please specify);



Milestones

PhD reference	Milestone number	Milestone name	Delivery date from AWP	Actual delivery date	Achieved	Comments
PhD16-FBZZ/AMR6.1-Codes4strains	M-PhD16-2.1	Pilot genome dataset defined	M24	M24	Yes	
	M-PhD16-3.1	cgMLST schemes defined for Kp and Ec	M36	M36	Yes	
	M-PhD16-3.2	LINcodes algorithm defined	M30	M30	Yes	
	M-PhD16-3.3	SNapperDB databases set-up for both pathogens	M36	M48	No	Yes for Ec; not for Kp
	M-PhD16-4.1	Simulations completed	M42		No	Not realized; 7,000 public genomes were sufficient to constitute an adequate real genomic dataset.
	M-PhD16-4.2	Publication submitted	M48	M48	Yes	
	M-PhD16-4.3	A novel tool to define the population structure of species, called MSTclust		M40	Yes	
	M-PhD16-4.4	Optimization of our inheritance algorithm to provide backwards compatibility of MLST groups with 7-gene MLST		M42	Yes	
	M-PhD16-4.5	Comparison of the ANI metric with the cgMLST metric		M42	Yes	
	M-PhD16-4.6	Designed a method to identify recombination in cgMLST		M40	Yes	
	M-PhD16-4.7	Integration of MLST identifiers into the BIGSdb platform to make them publicly available		M42	Yes	
	M-PhD16-5.1	LINcodes method disseminated	M54	M56	Yes	
	M-PhD16-5.2	Viva presented	M57	M59	Yes	



9. Interactions with JRPs/JIPs or with external (EU or national) relevant projects or initiatives such as national action plans (AMR, Zoonoses etc.),

The novel LIN codes method developed during the course of the PhD project will naturally disseminate via the existing networks of collaboration in which the main investigators are involved: *K. pneumoniae* networks include MedVetKlebs (just finished JRP), KlebNET (JPIAMR support), SpARK (JPIAMR support), kleb-GAP, Nor-Kleb-Net (see MedVetKlebs final report), and KlebNET-GSP (funded by BMGF); and *Corynebacterium diphtheriae* surveillance networks at national and international levels include the French NRC @Pasteur, the ECDC, and international contacts.

10. Interactions with OHEJP stakeholders, national and international surveillance programmes.

The novel LIN codes nomenclature system will be used for wider communication and backwards-compatibility in the broad genomic epidemiology arena. Stakeholders such as EFSA, ECDC, national public health agencies including USA CDC, or WHO might end-up hearing from, evaluating and adopting this system, but this will also percolate naturally via the adoption of the approach by the epidemiological surveillance community for *Klebsiella* or other pathogens.

Future interactions with application developers and bacterial nomenclature platform administrators will be established to compare this approach with comparable approaches in the field, such as HierCC (Zhou *et al.*, 2021; Enterobase, Warwick, UK) or PopPunk (Lees *et al.*, 2019; <https://poppunk.net/>); or commercial software such as SeqSphere+ (Ridom & cgmlst.org).

Following our publication on the cgLIN codes concept and its implementation in *K. pneumoniae*, Martin M. C. Maiden and Keith A. Jolley were interested in applying the method on other pathogens of interest to the One Health perspective (e.g., *Campylobacter*) or Clinical/Antimicrobial resistance (e.g., *N. gonorrhoeae*). The tool is now available on PubMLST platform at Oxford and can be easily used by other partners.

11. Added value and benefits during PhD resulting from being part of the OHEJP doctoral programme and consortium

The annual conferences organized by OHEJP allowed me to get in touch with other researchers, including other PhD students. They also allowed me to share the results of the project with the One Health community, which gave an additional dimension to my thesis, *i.e.* to give a greater visibility to the project.

The OHEJP doctoral program allowed me to network with other One Health partners, such as those involved in the MedVetKlebs project.



12. Transferrable skills and Training

Name of Training Event	Topic	Dates (DD/MM/YY)	Organising Institute
Communicate effectively in writing	<p>Objectives</p> <ul style="list-style-type: none"> - Applying the principles of effective communication to concrete cases adapted to the needs of doctoral students - Learning how to write a document, in particular an article, adapted to its recipients in order to be read and understood <p>Practicing using the methodology specific to each document and professional writing style</p> <p>Contents</p> <ul style="list-style-type: none"> - Characteristics of writing in communication and marketing (objectives, targets, messages) - How to select relevant information and give it meaning - How to structure your writing (adapted plans and specific methodologies) - Writing to be read (principles of readability and popularisation of information) <p>Improving the presentation of your documents (rules of form)</p>	31/03/2022 and 01/04/2022	Sorbonne Université
Speed up the writing of your thesis	<p>Objectives</p> <ul style="list-style-type: none"> - Understanding the different approaches to the thesis - Managing the organisation of the thesis - Structuring and writing the document clearly <p>Contents</p> <ul style="list-style-type: none"> - Organisation of research work (project management approach) - General approach <p>Searching for the main theme</p>	11/04/2022	Sorbonne Université
Establishing the basis for a personalised professional project	<ul style="list-style-type: none"> • General aspects of the job market, organisations, functions and positions • Panorama of the results of surveys on the professional future of Sorbonne PhDs • Notion of Personalized Professional Project • Notion of competences and reflexive posture 	09/11/2020	Sorbonne Université
Open access: general aspects	<ul style="list-style-type: none"> • Development of arguments and editorial solutions for Open access • Presentation of institutional and disciplinary open archives • Learning to use the portal HAL Sorbonne University 	05/03/2020	Sorbonne Université
Conduct your interviews and meetings efficiently	<ul style="list-style-type: none"> • Preparing and structuring interviews and meetings • Anticipate and anticipate objections or questions • Learn how to manage the types of interlocutors • Improve verbal and gestural expression techniques • Build your argumentation and develop your strength of conviction and persuasion • Use the educational tools in the right situation • Practice dialogue, participation and facilitation skills <p>Learn how to regulate and manage voltages face-to-face or in a group setting</p>	09/03/2020 10/03/2020	Sorbonne Université
Open science, by and for the	<ul style="list-style-type: none"> • Definition of open science • Institutional context 	25/02/2020	Sorbonne Université



benefit of researchers	<ul style="list-style-type: none"> • The need for openness in science 		
Ethics of scientific research	<p>3h Workshop, based on the Dilemma Game of the Erasmus University of Rotterdam.</p> <ul style="list-style-type: none"> • A maximum of 5 groups of 5 people are trained at each workshop. Each group must answer a dozen questions on scientific integrity, chosen according to the origin of the participants. When the group does not agree on a response, it must discuss it to seek consensus. • Pooling of group results, using the same consensus-finding and discussion process where this is not possible. 	21/02/2020	Sorbonne Université
Discover the main principles of management	<ul style="list-style-type: none"> • History of management • Management styles and team types • Management and teamwork • The skills of the manager in situation 	12/02/2020	Sorbonne Université
Seminar BIBLIO@PHD	<ul style="list-style-type: none"> • The program is focused on points of view from different actors of scientific publishing and varies depending on the years. • The programs and slides of prior seminars are available online: https://paris-sorbonne.libguides.com/bibliodoctorat 	31/01/2020	Sorbonne Université
Discover the principles of effective written and oral communication.	<ul style="list-style-type: none"> • Decrypt the act of communication • Principles of effective written communication • Principles of effective oral communication 	22/01/2020	Sorbonne Université
Research ethics and scientific integrity	<ul style="list-style-type: none"> • Open to all students and the whole scientific community of Sorbonne University. • Each conference focuses on a specific theme and is given by an internationally recognized speaker. Speakers are selected from the recognized resources (US Office of Research integrity, French OFIS, EUA-CDE, LERU). The lectures are captured in video for live broadcast on the other campuses of Sorbonne 	19/12/2019	Sorbonne Université

13. Ethical Reviews

Comments of Ethics Advisors, January 2020	Comments PhD Project Supervisor, mid-2020	Comments of Ethics Advisors, October 2020
None	None	None



14. Scientific Publications

Publication date	Publication title	Authors	DOI reference	Zenodo reference	Is OHEJP acknowledged?	Is it a Green Open Access? *please specify embargo length	Is it a Gold Open Access?
28 March 2023	A Global <i>Corynebacterium diphtheriae</i> genomic framework sheds light on current diphtheria reemergence	Hennart M , Crestani C, Bridel S, Armatys N, Brémont S, Carmi-Leroy A, Landier A, Passet V, Fonteneau L, Vaux S, Toubiana J, Badell E, Brisse S.	10.1101/2023.02.20.529124	https://zenodo.org/record/7732371#.ZDU0hnZByF4	Yes	NA	NA
5 January 2023	<i>Corynebacterium</i> of the <i>diphtheriae</i> complex in companion animals: clinical and microbiological characterization of 64 cases from France	Museux K, Arcari G, Rodrigo G, Hennart M , Badell E, Toubiana J, Brisse S.	10.1101/2023.01.04.522820	https://zenodo.org/record/7817238#.ZDUypHZByF4	Yes	NA	NA
24 October 2022	Genomic epidemiology of <i>Corynebacterium diphtheriae</i> in New Caledonia	Tessier E, Hennart M , Badell E, Passet V, Toubiana J, Biron A, Gourinat AC, Merlet A, Colot J, Brisse S.	10.1101/2022.10.23.512725	https://zenodo.org/record/7874934#.ZEuJwXZByF4	Yes	NA	NA
27 January 2023	Multidrug-resistant toxigenic <i>Corynebacterium diphtheriae</i> sublineage 453 with two novel resistance genomic islands	Arcari G, Hennart M , Badell E, Brisse S.	10.1099/mgen.0.000923	https://zenodo.org/record/7781603#.ZDUumHZByF4	Yes		Yes
14 June 2022	A dual barcoding approach to bacterial strain nomenclature: Genomic taxonomy of <i>Klebsiella pneumoniae</i> strains.	Hennart M , Guglielmini J, Bridel S, Maiden MCJ, Jolley KA, Criscuolo A, Brisse S.	10.1093/molbev/msac135	https://zenodo.org/record/6855315#.ZCQr93ZByF4	Yes		Yes
18 November 2021	Genomic epidemiology and strain taxonomy of <i>Corynebacterium diphtheriae</i>	Guglielmini J, Hennart M , Badell E, Toubiana J, Criscuolo A, Brisse S.	10.1128/JCM.01581-21	https://zenodo.org/record/6617364#.ZCQ1V3ZByF4	Yes	Yes; 18 Nov. 2021	
26 May 2021	Ongoing diphtheria outbreak in Yemen: a cross-sectional and genomic epidemiology study	Badell E, Alharazi A, Criscuolo A, Almoayed KAA, Lefrancq N, Bouchez V, Guglielmini J, Hennart M , Carmi-Leroy A, Zidane N, Pascal-Perrigault M, Lebreton M, Martini H, Salje H, Toubiana J,	10.1016/S2666-5247(21)00094-X	https://zenodo.org/record/6393150#.YqbZ5XZBwuU	No		Yes



		Dureab F, Dhabaan G, Brisse S; NCPHL diphtheria outbreak working group.					
27 November 2020	Population genomics and antimicrobial resistance in <i>Corynebacterium diphtheriae</i>	Hennart M , Panunzi LG, Rodrigues C, Gaday Q, Baines SL, Barros-Pinkelrig M, Carmi-Leroy A, Dazas M, Wehenkel AM, Didelot X, Toubiana J, Badell E, Brisse S.	10.1186/s13073-020-00805-7	https://zenodo.org/record/4576982#.YqbYM3ZBwuU			Yes
13 May 2020	<i>Klebsiella pneumoniae</i> carriage in low-income countries: antimicrobial resistance, genomic diversity and risk factors	Huynh BT, Passet V, Rakotondrasoa A, Diallo T, Kerleguer A, Hennart M, Lauzanne A, Herindrainy P, Seck A, Bercion R, Borand L, Pardos de la Gandara M, Delarocque-Astagneau E, Guillemot D, Vray M, Garin B, Collard JM, Rodrigues C, Brisse S.	10.1080/19490976.2020.1748257	https://doi.org/10.5281/zenodo.3929396	No		Yes
28 February 2020	<i>Corynebacterium rouxii</i> sp. nov., a novel member of the diphtheriae species complex	Badell E, Hennart M , Rodrigues C, Passet V, Dazas M, Panunzi L, Bouchez V, Carmi-Leroy A, Toubiana J, Brisse S.	10.1016/j.resmic.2020.02.003	https://zenodo.org/record/6393067#.YqbiCHZBwuU	No	Yes; 12 months	



15. Additional Outputs

The tools developed:

<https://gitlab.pasteur.fr/BEBP/diphtoscan>

<https://gitlab.pasteur.fr/BEBP/inheritance-algorithm>

<https://gitlab.pasteur.fr/BEBP/LINcoding>

The *K. pneumoniae* isolates database and the sequence database with LIN codes:

https://bigsdbs.pasteur.fr/cgi-bin/bigsdbs/bigsdbs.pl?db=pubmlst_klebsiella_isolates

[https://bigsdbs.pasteur.fr/cgi-](https://bigsdbs.pasteur.fr/cgi-bin/bigsdbs/bigsdbs.pl?db=pubmlst_klebsiella_seqdef&page=query&scheme_id=18&submit=1)

[bin/bigsdbs/bigsdbs.pl?db=pubmlst_klebsiella_seqdef&page=query&scheme_id=18&submit=1](https://bigsdbs.pasteur.fr/cgi-bin/bigsdbs/bigsdbs.pl?db=pubmlst_klebsiella_seqdef&page=query&scheme_id=18&submit=1)

16. Specific outcomes to highlight in dissemination and communications

Outcomes (deliverable, publication, folder, tool, etc.) of the project that might be suited for communication purposes to various audiences, for instance food safety or AMR scientists, national and international stakeholders, specific professionals, the general public, etc.

During the project, we first developed a taxonomy method for bacterial strains. The cgLIN tool is in open access and has been implemented in the BIGSdb (PubMLST and Pasteur) and pathogenwatch platforms, the two leading bacterial genomic epidemiology platforms.

This novel approach was praised by some at the specialized congress IMMEM in Bath, was the object of a communiqué by the Pasteur press office (<https://www.pasteur.fr/en/research-journal/news/improving-bacterial-strain-classification-more-effective-surveillance>) and is being implemented for other pathogens by international colleagues.

We also developed a bioinformatics pipeline, DIPHTOSCAN, which enables the harmonization, and facilitates, genomic epidemiology of the agents of diphtheria. This tool is available publicly and is described in a bioRxiv preprint (<https://www.biorxiv.org/content/10.1101/2023.02.20.529124v2>). It is being used by the ECDC Microbiology and Molecular Surveillance team to investigate the European diphtheria reemergence in 2022 (<https://www.ecdc.europa.eu/en/publications-data/increase-reported-diphtheria-cases-among-migrants-europe-due-corynebacterium>)

Are there any outcomes of this project that are already discussed or even implemented and in use at any institute of the project consortium, at stakeholders' organisations (ECDC, EFSA, EMA, EEA, FAO, OIE, WHO), or at the level of national authorities?

Our tools have been distributed publicly and some of them are in use. For example, a comment on a publication on multidrug resistant carbapenem-resistant *Klebsiella pneumoniae* ST23, related to an ECDC risk assessment (<https://www.ecdc.europa.eu/en/publications-data/risk-assessment-emergence-hypervirulent-klebsiella-pneumoniae-eu-eea>), pointed out the utility of our LIN code classification system, and underlined that its use would have avoided confusion raised by these publications, which was due to ST23 being made of two phylogenetically very distinct sublineages of *K. pneumoniae* (<https://doi.org/10.1093/jac/dkad028>).

The diphtOscan tool was presented to, and is being used by, the ECDC Microbiology and Molecular Surveillance team and other public health microbiology teams in national reference laboratories, to investigate the European diphtheria reemergence in 2022



(<https://www.ecdc.europa.eu/en/publications-data/increase-reported-diphtheria-cases-among-migrants-europe-due-corynebacterium>).

17. One Health impact

The project has defined, implemented and evaluated:

(i) a novel bioinformatics strategy to classify and name strains within pathogenic bacteria, from the level of deep subspecific lineages down to shallower levels of diversity that differentiate epidemiological related strains from non-related ones. The general applicability of the LIN codes approach means that in the future the classification and nomenclature of strains of other pathogens could benefit from the PhD project outcomes. We have discussed the approach with one member of the ECDC, Erik Alm (Principal Expert Applied Molecular Epidemiology at European Centre for Disease Prevention and Control), who expressed interest in a dissemination towards ECDC's molecular surveillance networks and EU national focal points. By facilitating the intercommunicability on bacterial strains across sectors and countries in the future, the project is highly relevant to multiple topics and objectives of the One Health perspective: antibiotic resistance clonal dissemination, emerging pathogens, host-jumps/species barrier crossing, cross-sector transmission, public health, and basic microbiology integration.

This project outcome has far-reaching impacts on possibilities to integrate efforts of agencies (e.g., at the international levels, ECDC, EFSA, PulseNet international) to detect, monitor, understand and control the spread of pathogens.

(ii) Our tool to scan the genomes of diphtheria clinical isolates, including those from reservoir animals (pet cats and dogs mainly) for virulence, resistance and other characteristics will facilitate their molecular epidemiology and the understanding of transmission from animals to humans, or between humans in local, national or global scales. This tool is already in use by the ECDC Microbiology and Molecular Surveillance team (Daniel Palm and Andreas Hoefer, pers. Comm.).

(iii) We have also described a novel zoonotic pathogen of the *Corynebacterium diphtheriae* complex (*C. rouxii*). Isolates of this species are *tox*-negative and the distinction of this species from *C. diphtheriae* will improve risk assessment and diagnostic of diphtheria.

Through this project, the PI Sylvain Brisse was invited at One Health international conferences, including a keynote at the One Health EJP ASM in 2022, and a talk at the World One Health Congress in Singapore (November 2023). He was also invited to talk at the prestigious international KlebClub webinar series (May 10th, 2022) to talk about *Klebsiella* strain taxonomy. He thus interacted with other colleagues in the field.

We have also started collaborating on the diphtheria EU 2022 emergence with ECDC and other national reference laboratories and public health agencies on diphtheria; and S Brisse gave a talk on genomics of resistance in diphtheria at the Wellcome Trust conference Antimicrobial Resistance, Genomes, Big Data and Emerging Technologies (Nov. 2020).



18. List of dissemination and communication activities

Name of the activity:	"Journées Boris Ephrussi" 2022		
Date:	May 05 th -06 th 2022		
Place:	Paris		
	Yes / No		Yes / No
Organisation of a Conference		Participation to a Conference	Yes
Organisation of a Workshop		Participation to a Workshop	
Press release		Participation to an Event other than a Conference or a Workshop	
Non-scientific and non-peer-reviewed publication (popularised publication)		Video/Film	
Exhibition		Brokerage Event	
Flyer		Pitch Event	
Training		Trade Fair	
Social Media		Participation in activities organized jointly with other H2020 projects	
Website		Other	
Communication Campaign (e.g. Radio, TV)			
Specify the estimated number of persons reached, in the context of this dissemination and communication activity), in each of the following categories			
	Number		Number
Scientific Community (Higher Education, Research)	100+	Media	
Industry		Investors	
Civil Society		Customers	
General Public		Other	
Policy Makers			

Name of the activity:	OHEJP ASM
Date:	April 11-13 2022
Place:	Orvieto, Italy /online



Specify the Dissemination and Communication activities linked to the One Health EJP project for each of the following categories

	Yes / No		Yes / No
<i>Organisation of a Conference</i>		<i>Participation to a Conference</i>	Yes (poster)
<i>Organisation of a Workshop</i>		<i>Participation to a Workshop</i>	
<i>Press release</i>		<i>Participation to an Event other than a Conference or a Workshop</i>	
<i>Non-scientific and non-peer-reviewed publication (popularised publication)</i>		<i>Video/Film</i>	
<i>Exhibition</i>		<i>Brokerage Event</i>	
<i>Flyer</i>		<i>Pitch Event</i>	
<i>Training</i>		<i>Trade Fair</i>	
<i>Social Media</i>		<i>Participation in activities organized jointly with other H2020 projects</i>	
<i>Website</i>		<i>Other</i>	
<i>Communication Campaign (e.g. Radio, TV)</i>			

Specify the estimated number of persons reached, in the context of this dissemination and communication activity), in each of the following categories

	Number		Number
<i>Scientific Community (Higher Education, Research)</i>		<i>Media</i>	
<i>Industry</i>		<i>Investors</i>	
<i>Civil Society</i>		<i>Customers</i>	
<i>General Public</i>		<i>Other</i>	
<i>Policy Makers</i>			



Name of the activity:	OHEJP ASM		
Date:	April 11-13 2022		
Place:	Orvieto, Italy /online		
Specify the Dissemination and Communication activities linked to the One Health EJP project for each of the following categories			
	Yes / No		Yes / No
Organisation of a Conference		Participation to a Conference	Yes (Thesis in 3 minutes)
Organisation of a Workshop		Participation to a Workshop	
Press release		Participation to an Event other than a Conference or a Workshop	
Non-scientific and non-peer-reviewed publication (popularised publication)		Video/Film	
Exhibition		Brokerage Event	
Flyer		Pitch Event	
Training		Trade Fair	
Social Media		Participation in activities organized jointly with other H2020 projects	
Website		Other	
Communication Campaign (e.g. Radio, TV)			
Specify the estimated number of persons reached, in the context of this dissemination and communication activity), in each of the following categories			
	Number		Number
Scientific Community (Higher Education, Research)		Media	
Industry		Investors	
Civil Society		Customers	
General Public		Other	
Policy Makers			



Name of the activity:	OHEJP ASM		
Date:	April 11-13 2022		
Place:	Orvieto, Italy /online		
Specify the Dissemination and Communication activities linked to the One Health EJP project for each of the following categories			
	Yes / No		Yes / No
Organisation of a Conference		Participation to a Conference	Yes (Roundtable discussion talk)
Organisation of a Workshop		Participation to a Workshop	
Press release		Participation to an Event other than a Conference or a Workshop	
Non-scientific and non-peer-reviewed publication (popularised publication)		Video/Film	
Exhibition		Brokerage Event	
Flyer		Pitch Event	
Training		Trade Fair	
Social Media		Participation in activities organized jointly with other H2020 projects	
Website		Other	
Communication Campaign (e.g. Radio, TV)			
Specify the estimated number of persons reached, in the context of this dissemination and communication activity), in each of the following categories			
	Number		Number
Scientific Community (Higher Education, Research)		Media	
Industry		Investors	
Civil Society		Customers	
General Public		Other	
Policy Makers			



Name of the activity:	"Journées Boris Ephrussi" 2021		
Date:	Poster : A new approach for naming bacterial strains, combining cgMLST and LIN codes May 27 th -28 th 2021		
Place:	Digital Conference		
Specify the Dissemination and Communication activities linked to the One Health EJP project for each of the following categories			
	Yes / No		Yes / No
Organisation of a Conference		Participation to a Conference	Yes
Organisation of a Workshop		Participation to a Workshop	
Press release		Participation to an Event other than a Conference or a Workshop	
Non-scientific and non-peer-reviewed publication (popularised publication)		Video/Film	
Exhibition		Brokerage Event	
Flyer		Pitch Event	Yes
Training		Trade Fair	
Social Media		Participation in activities organized jointly with other H2020 projects	
Website		Other	
Communication Campaign (e.g. Radio, TV)			
Specify the estimated number of persons reached, in the context of this dissemination and communication activity), in each of the following categories			
	Number		Number
Scientific Community (Higher Education, Research)		Media	
Industry		Investors	
Civil Society		Customers	
General Public		Other	
Policy Makers			



Name of the activity:	OHEJP Annual Scientific Meeting 2021 – poster presentation		
Date:	09-11 June 2021		
Place:	Online		
Specify the Dissemination and Communication activities linked to the One Health EJP project for each of the following categories			
	Yes / No		Yes / No
Organisation of a Conference		Participation to a Conference	Yes
Organisation of a Workshop		Participation to a Workshop	
Press release		Participation to an Event other than a Conference or a Workshop	
Non-scientific and non-peer-reviewed publication (popularised publication)		Video/Film	
Exhibition		Brokerage Event	
Flyer		Pitch Event	
Training		Trade Fair	
Social Media		Participation in activities organized jointly with other H2020 projects	
Website		Other	
Communication Campaign (e.g. Radio, TV)			
Specify the estimated number of persons reached, in the context of this dissemination and communication activity), in each of the following categories			
	Number		Number
Scientific Community (Higher Education, Research)	550+	Media	0
Industry	0	Investors	0
Civil Society	5	Customers	0
General Public	0	Other	0
Policy Makers	0		



Name of the activity:	One Health EJP Annual Scientific Meeting 2020		
Date:	27 th _ 29 th May 2020		
Place:	Digital Conference		
Specify the Dissemination and Communication activities linked to the One Health EJP project for each of the following categories			
	Yes / No		Yes / No
Organisation of a Conference		Participation to a Conference	Yes
Organisation of a Workshop		Participation to a Workshop	
Press release		Participation to an Event other than a Conference or a Workshop	
Non-scientific and non-peer-reviewed publication (popularised publication)		Video/Film	
Exhibition		Brokerage Event	
Flyer		Pitch Event	
Training		Trade Fair	
Social Media		Participation in activities organized jointly with other H2020 projects	
Website		Other	
Communication Campaign (e.g. Radio, TV)			
Specify the estimated number of persons reached, in the context of this dissemination and communication activity), in each of the following categories			
	Number		Number
Scientific Community (Higher Education, Research)		Media	
Industry		Investors	
Civil Society		Customers	
General Public		Other	
Policy Makers			



Name of the activity:	One Health EJP Annual Scientific Meeting 2020		
Date:	Oral : Three Minute Thesis (3MT) competition : Codes4strains		
Place:	27 th _ 29 th May 2020		
	Digital Conference		
Specify the Dissemination and Communication activities linked to the One Health EJP project for each of the following categories			
	Yes / No		Yes / No
Organisation of a Conference		Participation to a Conference	Yes
Organisation of a Workshop		Participation to a Workshop	
Press release		Participation to an Event other than a Conference or a Workshop	
Non-scientific and non-peer-reviewed publication (popularised publication)		Video/Film	
Exhibition		Brokerage Event	
Flyer		Pitch Event	Yes
Training		Trade Fair	
Social Media		Participation in activities organized jointly with other H2020 projects	
Website		Other	
Communication Campaign (e.g. Radio, TV)			
Specify the estimated number of persons reached, in the context of this dissemination and communication activity), in each of the following categories			
	Number		Number
Scientific Community (Higher Education, Research)		Media	
Industry		Investors	
Civil Society		Customers	
General Public		Other	
Policy Makers			