BMJ Global Health

Clustering and visualisation of the GABRIEL network expertise in the field of infectious diseases

Cécile Chauvel,[1,2] Philippe Vanhems,[2,3] Marie-Charlotte Quemin,[4] Marianne Abifadel,[5] Shally Awasthi,[6] Sayera Banu,[7] Silvia Figueiredo Costa,[8] Sara Eyangoh,[9] Monzer Hamze,[10] Zakir Hossain,[11] Bourema Kouriba ,[12] Daniel Mukadi-Bamuleka,[13] Francine Ntoumi,[14] Abdoul-Salam Ouedraogo,[15] Phimpha Paboriboune,[16] Jean William Pape,[17] Chan Leakhena Phoeung ,[18] Firdausi Qadri ,[19] Ana Tereza Ribeiro Vasconcelos,[20] Graciela Russomando,[21] Luc Samison,[22] Marilda Agudo Mendonça Siqueira,[23] Nestani Tukvadze,[24] Jianwei Wang,[25] Florence Komurian Pradel [4]

To cite: Chauvel C, Vanhems P, Quemin M-C, *et al*. Clustering and visualisation of the GABRIEL network expertise in the field of infectious diseases. *BMJ Glob Health* 2025;**10**:e017595. doi:10.1136/bmjgh-2024-017595

Handling editor Fi Godlee

Received 19 September 2024
Accepted 2 May 2025

For numbered affiliations see end of article.

Correspondence to
Dr Florence Komurian Pradel; florence.pradel@fondation-merieux.org

ABSTRACT

Introduction The Global Approach to Biology Research, Infectious diseases and Epidemics in Low-income countries (GABRIEL) network is an international scientific network of 21 centres coordinated by the Merieux Foundation (Lyon, France). Mapping and characterising the similarities and differences in expertise and activities across four major infectious diseases (tuberculosis, antimicrobial-resistant infections, acute respiratory infections and emerging pathogens) among these centres would help to provide a better understanding of the network's capacity. It will also highlight how the applied methodology can enhance information sharing within research networks.

Methods Each centre responded to a questionnaire on their core activities and research themes. An advanced multivariate analysis was performed to relate all items together and highlight new synergies among members of the GABRIEL network. Similarities were found using a clustering algorithm and data were visualised using alluvial plots.

Results This strategy enabled to find new patterns in the GABRIEL network for the implementation of new projects on global health, regardless of geographical proximity or historical connections. Five clusters based on core activities, consisting of 6, 1, 3, 9 and 2 research units, respectively, have been identified, with clusters 1 and 4, including the majority of the units. Four clusters have been defined based on the four major infectious diseases, comprising 7, 3, 5 and 6 research units, respectively.

Conclusions The same methodology could also be applied to identify proximities on other networks of experts or between members of different networks for more efficient research or surveillance global programmes.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The capacities of the Global Approach to Biology Research, Infectious diseases and Epidemics in Low-income countries (GABRIEL) network are known individually for each member. However, there is no overall vision to highlight synergies and facilitate the development of collaborative research projects within the network.

WHAT THIS STUDY ADDS

⇒ This study provides a methodology for a multivariate and comprehensive analysis. It enables highlighting new similarities and diversities between centres of the GABRIEL network in terms of core activities or research themes, regardless of geographical proximity or historical connections and can identify potential weaknesses among GABRIEL members.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The similarities between centres will be used to develop future collaborative projects.

⇒ The differences between members in their areas of expertise and specialities in infectious diseases are an asset that reinforces the strength of the network.

⇒ The same study design and/or methodology could be applied to find synergies within other networks of experts or between several networks to implement more efficient global research or surveillance programmes.

INTRODUCTION

The Global Approach to Biology Research, Infectious diseases and Epidemics in Low-income countries (GABRIEL) network is an international scientific network coordinated by the Mérieux Foundation (Lyon, France).[1] It aims to strengthen the research and training capacities of laboratory members and improve the surveillance of infectious diseases with a major focus on public health

in countries of limited resources. The GABRIEL network also promotes a partnership with low and middle-income countries to support local responses in the field of infectious diseases and to address public health challenges through collaborative research, technology transfer, building capacity among 21 research institutions in Africa, Asia, Middle East, Caribbean, South America and Eastern Europe. Members of this network have diversified fields of expertise for activities such as research, surveillance and/or medical analysis.

Implementing large epidemiological studies on infectious diseases requires strong collaboration between centres with often diverse skills and knowledge. Thus, a detailed description of each research centre within the GABRIEL network will be helpful to orient the best research strategy. The GABRIEL network research activities mostly focus on four main topics of public interest: acute respiratory infections (ARI),[2 3] tuberculosis,[4–7] antimicrobial resistance (AMR)[8–10] and emerging pathogens.[11 12]

Gaining a clear overview of expertise and activities within a network of numerous institutions is challenging and can hinder collaboration. Moreover, the complexity of a network, due to the number and diversity of its characteristics, can make it difficult to analyse and share this information effectively. The objectives of this study are twofold: first, to map the expertise of 21 research units belonging to the GABRIEL network based on their activities and expertise in four major infectious diseases (tuberculosis, AMR, ARI and emerging pathogens) to highlight their specificities and complementarities, and second, to discuss this methodology for optimising the information sharing within research networks. For this purpose, a mapping was conducted using multivariate analysis to capture the diversity of GABRIEL members, with similarities identified through a clustering algorithm and data visualised using alluvial plots.

METHODS

Study setting

The research units from academic, public and private institutions, located in 16 countries, participate in this study as part of the GABRIEL network (figure 1). The 21 enrolled research units were anonymised to ensure the confidentiality of the collected data.

Study design

The study is a cross-sectional survey consisting of a mapping of expertise and activities, with data collection conducted from May 2022 to January 2023. The data were provided by the heads of the research units.

Data collection

A questionnaire consisting over 90 questions was developed and filled in once to assess different aspects of expertise, including core activities (such as active or passive surveillance; basic or applied/clinical research; diagnostics; training), cross-disciplinary activities (clinical research, data collection and management, quality, shipment, etc), research themes (ARI including COVID-19, tuberculosis, AMR and emerging pathogens), collaborations, positioning and role at the national and international levels (eg, as a reference laboratory with specialised, nationally recognised technical expertise (ex for influenza), or as a focal point laboratory, coordinating networks and ensuring collaboration and data

Figure 1 Map of the GABRIEL members. GABRIEL, Global Approach to Biology Research, Infectious diseases and Epidemics in Low-income countries.

sharing at the international level. Geographical origin was not considered in this study to preserve anonymity. Before being sent to the study participants, the questionnaire was reviewed by the GABRIEL network steering committee, several experts from the Mérieux Foundation and pretested by a member of the GABRIEL network to ensure clarity and unambiguity of the questions. The collected data from the 21 research units were entered in a database and manually double-checked to ensure data quality. The questions selected for this study are related to the core activities, the research themes and the national or international positioning of the research institutions.

Clustering concept

Clustering methods were applied to the data collected retrospectively to find proximities without a priori regarding expertise and activities between the research units of the GABRIEL network. These statistical tools are dedicated to finding groups, called clusters, without pre-existing knowledge. Observations in the same cluster should be as homogeneous as possible, while observations in different clusters should be as discriminative as possible. Many variables can be analysed simultaneously, making interpretation easier when finding similarities and discrepancies in the data.

This study used binary variables, where a dot indicates the presence of a variable for the observation. However, when dealing with a large number of observations, variables or noisy data, the heatmap becomes challenging to interpret despite being intuitive.

Factorial analyses are dimension-reduction methods that represent data on two-dimensional scatterplots, where the axes are linear combinations of automatically computed variables. Each observation is represented as a dot, which can be colour-coded based on cluster membership. While clusters can be visually distinguished, interpreting the relationships between variables and clusters requires statistical expertise. As a result, these figures are not easily understandable without prior training and are not well suited for communication with a broad audience. Alluvial plots, in contrast, are newer tools for visualising changes over time or categorical variables. When combined with clustering, they are typically used to represent cluster evolution over time,[13] compare algorithms[14–16] or interpret clusters using a single variable.[17]

Data analysis

The clustering algorithm of partitioning around medoids (PAM) with Gower distance was applied separately for the list of core activities and the research themes.[18 19] The Gower distance is well suited for categorical variables, and PAM is a clustering technique that is robust to outliers and particularly efficient with small datasets. For both clusterings, the numbers of clusters were automatically determined between 2 and 5, the latter being the number of variables. For the sake of interpretation, it would have been difficult to decipher more than five groups of research institutions. The final numbers of clusters maximised the average silhouette widths of all observations.[20]

For each research institution, cluster membership and values of all variables were displayed in alluvial plots, which usually display distributions of observations in streams. We refer the reader to the literature for a comprehensive explanation on alluvial plots.[21]

All statistical analyses were performed using R Statistical Software (V.4.2.2[22]). R packages cluster and easyalluvial were used for clustering and plotting alluvial diagrams, respectively.

Table 1 Descriptive statistics of the replies to the questionnaire for all laboratories of the GABRIEL network

		N (%)
Core activities		
Surveillance	*Active*	16 (76)
	Passive	14 (67)
Research	*Basic*	6 (29)
	Clinical or applied	20 (95)
Training		19 (90)
Diagnostic services		17 (81)
Reference or focal point		14 (67)
Research themes		
Tuberculosis		11 (52)
AMR		10 (48)
ARI		18 (86)
Emerging pathogens		7 (33)
Other pathogens		7 (33)

AMR, antimicrobial resistance; ARI, acute respiratory infection, including COVID-19; GABRIEL, Global Approach to Biology Research, Infectious diseases and Epidemics in Low-income countries.

RESULTS

Descriptive statistics of the replies to the questionnaire selected for this study are displayed in table 1. More than 90% of the 21 research institutions performed clinical or applied research (N=20) and training (N=19), and around 80% performed active surveillance (N=16) or provided diagnostic services (N=17). Two-thirds were reference centres or focal points, and the same amount passively monitor pathogens (N=14). The less frequent core activity was the basic research (N=6, 29%).

Regarding research themes, 86% (N=18) of research institutions worked on ARI. Around half focused on tuberculosis (N=11, 52%) or AMR (N=10, 48%). Emerging pathogens and other pathogens were research themes for a third of the research institutions (N=7). Multivariate analyses considering all these parameters were needed to find relationships between laboratories regarding activities or research themes.

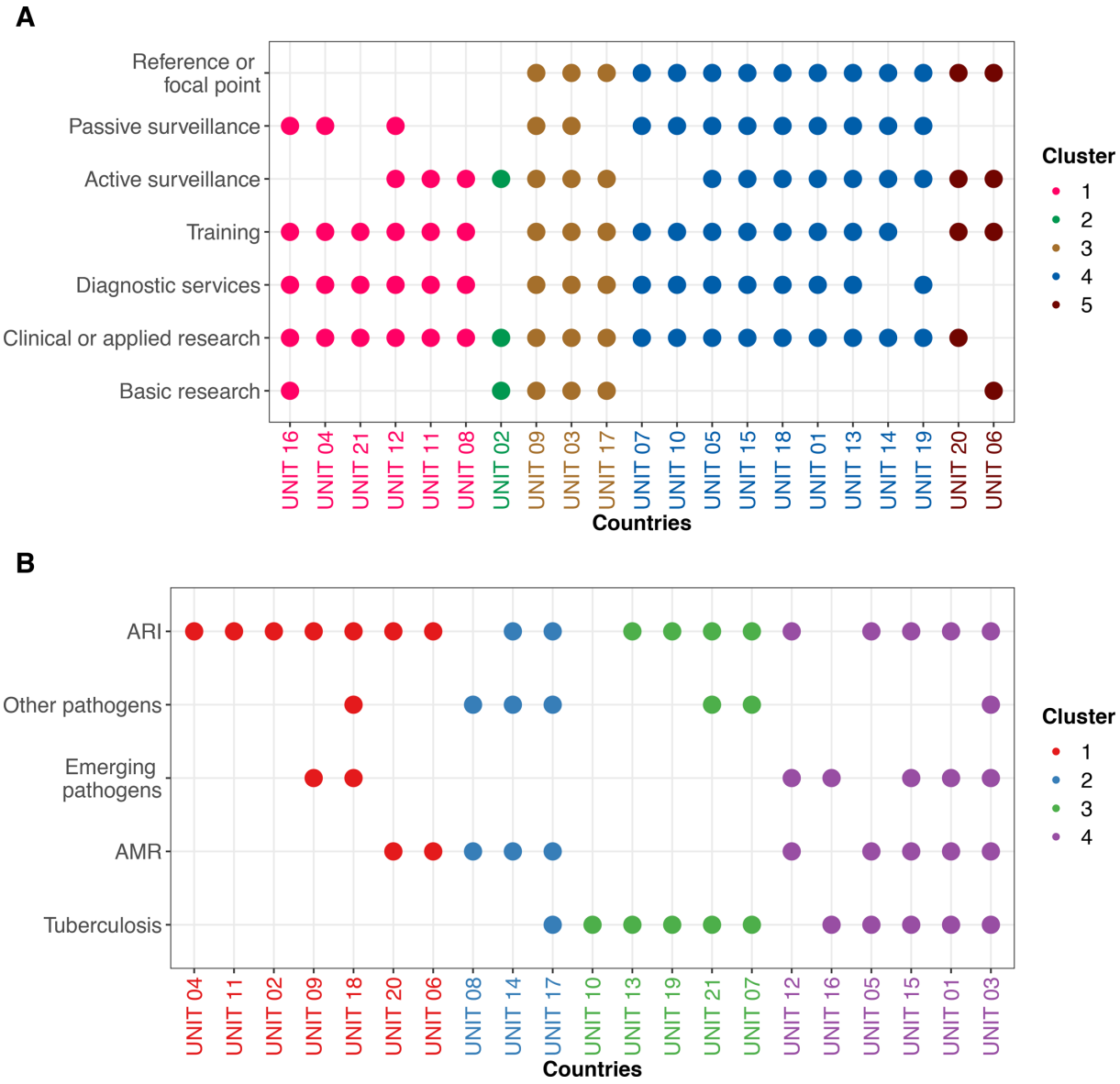

Figure 2 Clustering on core activities (A) and research themes (B). Plots of positive responses to each feature for each research institution were coloured according to cluster membership. A dot symbolises that the research institution in the column performs the type of core activity in row (A) or has activities related to the research theme in row (B). AMR, antimicrobial resistance; ARI, acute respiratory infection, including COVID-19.

The clustering algorithm based on activities enabled the construction of five clusters of research institutions (figure 2A). One cluster comprised UNIT 04, 08, 11, 12, 16 and 21 (in pink). All these research institutions offered training and diagnostic services, performed clinical or applied research and were not reference or focal points. In another cluster, UNIT 03, 09 and 17 (in light brown) provided all activities except passive surveillance of pathogens for UNIT 17. The nine research institutions in the blue cluster were all references or focal points that performed only clinical or applied research. In addition, all but one provided diagnostic services (UNIT 14) or training (UNIT 19), and two research institutions did not actively monitor pathogens (UNIT 07 and 10), whereas all provided passive surveillance. UNIT 06 and UNIT 20, both in a separate cluster (in dark brown), were references or focal points, training centres and actively monitored pathogens. Finally, UNIT 02 (in green) stood alone in one cluster with features different from all other research institutions in the GABRIEL network. Indeed, all types of research were conducted by UNIT 02, along with active surveillance programmes, while no other activities were implemented.

Regarding the research themes, the clustering algorithm produced four distinct clusters (figure 2B). The research theme 'tuberculosis' mainly drove the creation of clusters, and all the research institutions, except for three (UNITS 08, 10 and 16) studied ARI. The main features common to the seven research institutions in the red cluster were that they all worked on ARI and none on tuberculosis. The three research institutions in the blue cluster (UNITS 08, 14 and 17) worked on AMR and other

pathogens, while two worked on ARI. In the third green cluster, all research institutions conducted works on tuberculosis, but not on AMR or emerging pathogens. In the last purple cluster, the research institutions studied both AMR and emerging pathogens in combination with tuberculosis or ARI, except for UNITS 05 and 16, which worked either on AMR or emerging pathogens but not on both. The interpretation of this clustering on research themes was not as effortless as with the core activities, due to a high background noise observed within each single cluster, reflecting high variability in the data.

Alluvial plots were then used to display the outputs of both clustering (figure 3) to ease the interpretation. They allow all variables to be viewed simultaneously with

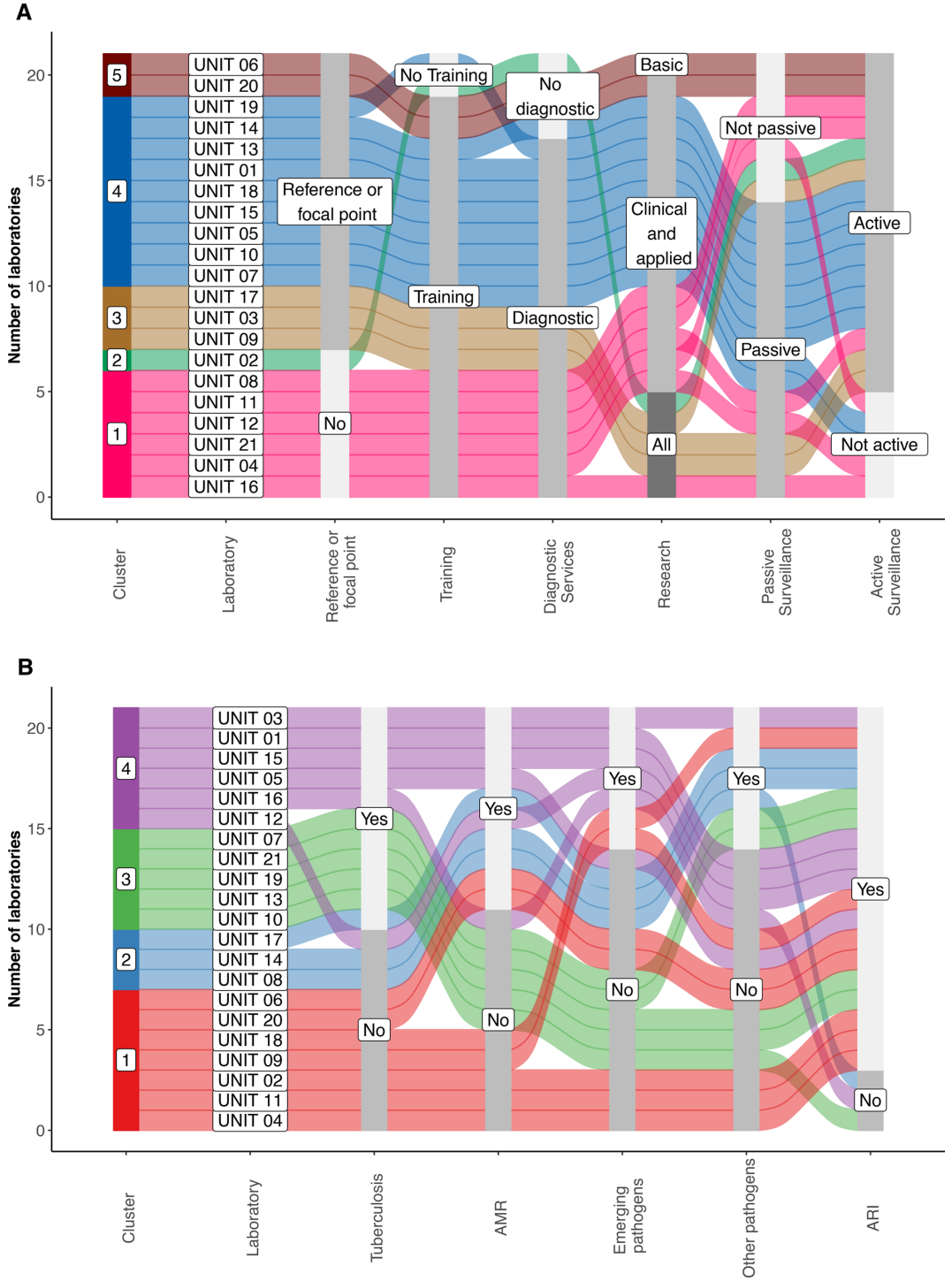

Figure 3 Alluvial plot illustrating the distribution of 21 research units, colour-coded by cluster membership across different: (A) core activities (reference or focal point; training; diagnostic services; research; passive and active surveillance) and (B) research themes (tuberculosis; AMR; emerging pathogens; other pathogens; ARI). AMR, antimicrobial resistance; ARI: acute respiratory infection, including COVID-19.

cluster trends shown by colour group. Group trends and similarities between clusters and the diversity of responses within each cluster can be analysed. Variables are displayed on the X-axis, while research institutions are piled up on the Y-axis, grouped and coloured according to their cluster membership. Each research institution's stream proceeds successively across the correct modality of each feature.

For the clustering of core activities (figure 3A), the reader can see why UNIT 02 was alone in one cluster (cluster 2 in green). The activities of this research institution differed from those of the others, and its stream was therefore atypical. Indeed, UNIT 02 was neither a reference laboratory nor a focal point, and it performed all research activities (basic and clinical/applied) like the three research institutions of cluster 3 (light brown). These were the only similarities of both clusters. UNIT 02 was one of the two research institutions that did not provide training, along with UNIT 19. For the other core activities, UNIT 02 was closer to UNIT 06 and UNIT 20, both research institutions of cluster 5 (dark brown), with no diagnostic nor passive surveillance activities but with active pathogen surveillance.

Clusters 3 and 4 were very similar, with the same absence or presence of core activities, except for basic research carried out solely by research institutions of cluster 3. Clusters 1 and 4 were similar, providing training and diagnostic services and performing exclusively clinical or applied research (except for UNIT 16). All research institutions in cluster 4 conducted passive pathogen surveillance, as did half of the research institutions in cluster 1.

The clusters will complement each other by leveraging their specific expertise to develop common research themes. For example, cluster 4, which has expertise in AMR, could collaborate with cluster 3, which specialises in tuberculosis but lacks an AMR component, to study resistance among Mycobacterium tuberculosis strains. Cross–cluster collaborations can effectively address complex research questions. The four clusters, comprising the majority of research units, have worked extensively on ARI—a field that has historically led to the discovery of emerging pathogens with significant public health impacts. In the event of an outbreak, they could establish dedicated research programmes and technological platforms to improve pathogen detection, benefiting from the expertise in emerging pathogens provided by the five UNITS in cluster 4. Geographical location plays a key role in enhancing cooperation and the flow of information. For example, due to their geographical proximity, which would allow rapid exchange of information and data, the UNITS present in Brazil can collaborate on epidemiological surveillance projects to track the spread of diseases such as dengue or Zika. On the other hand, when UNITS are located on different continents, they can contribute to epidemiological monitoring and provide alerts in case of emerging cases.

DISCUSSION

Clustering methods have a descriptive purpose and are frequently used in health studies for a broad range of data types, such as clinical data.[23–25] Graphical representations often simplify the interpretation of clustering analyses, with common types being classification trees, heatmaps and factorial analysis plots. Classification trees, derived from hierarchical clustering algorithms, can be unstable. Heatmaps, which display observations and variables in a matrix, are easy to read, especially for continuous values like biomarker or gene expression levels, where colour gradients indicate value ranges. This study used alluvial plots to interpret clusters based on the full set of variables used for clustering. By combining a clustering algorithm and alluvial plots, we visualise the characteristics of the GABRIEL network, providing a clear representation of its members' activities and expertise across four major infectious diseases. These graphs clearly illustrate the different connections or overlaps of activities and expertise within the network at a given time. The figures are intuitive and can be provided to communicate with a broad audience in a more understandable format than the other multivariate data representations.

Indeed, this approach can help stakeholders identify expertise, needs and gaps to better prioritise health issues. Mapping expertise and activities with this method eased the identification of institutions specialised in specific diseases, such as tuberculosis or ARI, which have potential connections within a particular domain for collaborations. Targeting the centres of excellence within the network would help to quickly mobilise the proper expertise in the event of epidemics, for example, and to establish connections with stakeholders interested in collaborating on a specific infectious disease, including grant applications. On the other hand, alluvial plots also helped to visualise gaps and areas where human resources and capacity-building efforts need to be developed to set up multidisciplinary approaches to tackle infectious diseases.[26]

This methodology, when focused on activities, highlights generic expertise independent of research themes, which complement traditional networks that are purely theme-oriented.[27 28] Alluvial plots have been used to describe trajectories or transversal descriptive characteristics but not often in relation to healthcare resources or health institutions, considering their expertise. This pilot approach applied to the GABRIEL network provided an opportunity to perform a multivariate descriptive analysis, complemented by interesting data visualisation. As a result, communication is facilitated not only for experts but also for uninitiated individuals (administrative staff, policymakers, journalists, etc) regarding statistics. This approach helps facilitate collaboration between GABRIEL members with similar profiles, allowing them to identify activities that need improvement or development. It also promotes the sharing of information for grant applications and supports the implementation of similar analyses in the future to track changes and trends.

Another benefit is that the alluvial plots can be used to analyse dynamic data over time, enabling rapid identification of changes in activity and expertise. This may be linked to a better adaptation to the context of infectious diseases, particularly for research and surveillance activities. The data to be analysed can be more complex by including specific outcomes, such as the number of collaborative research projects or publications, which can provide information on the productivity and impact of the network.

However, complex datasets with many variables can overload the alluvial plots, making the data visualisation less transparent and difficult to interpret. It is also important to ensure that data supplied by multiple sources are consistent and complete with the same level of quality to avoid inaccurate or biased visualisation. Data processing, cleaning and optimisation are necessary to produce optimal clustering results.[20] In particular, a change in the number of clusters may influence the results and the interpretation. As for any statistical analysis, another limitation lies in the missing data, which might impact the cluster definitions. The criteria to be analysed should be chosen carefully to reduce these biases and avoid differential bias by centre.

Alluvial plots can be applied to a broader scope, such as clinical data, for example,[29] to illustrate patient clinical phenotypes, treatments and clinical outcomes, to provide a better understanding of the effectiveness of treatments and areas for improvement, or to illustrate multimorbidity acquisition sequences according to age, ethnicity and social deprivation.[30] It was recently applied to analysing bibliometric data on attention-deficit/hyperactivity disorder (ADHD) and understanding the network of characteristics of the 100 top-cited ADHD-related articles.[31]

These figures supported the visualisation of similarity trends within clusters and dissimilarities between them. However, the absence of some core activity or research theme for some research institutions within a cluster may be challenging to detect. In addition, retrieving either the presence or the absence of each feature with this matrix structure may be difficult for research institutions, making it hard to extract key messages. Another limitation of this transversal study is that it was not possible to monitor changes over time regarding expertise in research topics, which could come up, for example, through capacity building. The geographical location of the research units, along with their activities and domains of expertise, would have been interesting to analyse. However, it was not included in the list of variables in order to preserve the anonymity of the units.

CONCLUSION

Alluvial graphs are a powerful tool for visualising complex data in various domains to a broad audience. They are helpful in any context where understanding the pattern and distribution of activities and expertise is essential. This visualisation facilitates the identification of skills, gaps and needs among GABRIEL members and highlights their specificities which are a strength for the network. Similar approaches for various networks would enable the recognition of centres' expertise from a significant collaborative project perspective. This approach can facilitate the integration of databases across networks, thus allowing for the grouping of expertise or identifying complementarities. It is important that this study be repeated to observe the dynamic of the developments over time in activities and research themes.

Author affiliations
[1]Center of Excellence in Respiratory Pathogens, Hospices Civils de Lyon, Lyon, France
[2]Équipe Santé Publique, Épidémiologie et Écologie Évolutive des Maladies Infectieuses, Inserm U1111, CNRS UMR5308, ENS de Lyon, Université Claude Bernard Lyon 1, Centre International de Recherche en Infectiologie, Lyon, France
[3]Service Hygiène, Epidémiologie et Prévention. Centre Hospitalier Hôpital Edouard Herriot, Hospices Civils de Lyon, Lyon, France
[4]Fondation Mérieux, Lyon, France
[5]Rodolphe Mérieux Laboratory, Faculty of Pharmacy, Saint Joseph University of Beirut, Beirut, Lebanon
[6]Department of Pediatrics, King George's Medical University, Lucknow, India
[7]International Centre for Diarrheal Disease Research, Dhaka, Bangladesh
[8]Tropical Medicine Institute of Sao Paulo, University of Sao Paulo, Sao Paulo, Brazil
[9]Centre Pasteur du Cameroun, Yaounde, Cameroon
[10]Laboratoire Microbiologie Santé et Environnement, Faculty of public Health, Lebanese University, Tripoli, Lebanon
[11]Rodolphe Merieux Laboratory, Bangladesh Institute of Tropical and Infectious Disease, Chittagong, Bangladesh
[12]Charles Merieux Center of Infectiology, Bamako, Mali
[13]Rodolphe Merieux Laboratory, Institut National de Recherche Biomédicale, Goma, République démocratique du Congo
[14]Fondation Congolaise pour la Recherche Médicale, Brazzaville, République du Congo
[15]Laboratoire des Pathogènes Emergents et Ré-émergents, Université Nazi Boni, Bobo-Dioulasso, Burkina Faso
[16]Center of Infectiology Lao - Christophe Mérieux, Ventiane, Lao People's Democratic Republic
[17]Les Centres GHESKIO, Port-au-Prince, Haiti
[18]Rodolphe Merieux Laboratory, University of Health Sciences, Phnom Penh, Cambodia
[19]Institute for developing Science and Health Initiatives, Dhaka, Bangladesh
[20]Bioinformatics Laboratory, National Laboratory of Scientific Computing, Petropolis, Brazil
[21]Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción, San Lorenzo, Paraguay
[22]Charles Mérieux Center for Infectious Diseases, University of Antananarivo, Antananarivo, Madagascar
[23]Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil
[24]National Center For Tuberculosis And Lung Diseases, Tbilisi, Georgia
[25]Christophe Mérieux Laboratory, Chinese Academy of Medical Sciences & Peking Union Medical College Institute of Pathogen Biology, Beijing, China

X Firdausi Qadri @fqadri@icddrb.org

Acknowledgements The authors acknowledge the funders for their support and the staff of the participating research institutions for providing the data.

Contributors CC, PV, M-CQ, FKP conceptualised and designed the study and wrote the manuscript; PV, M-CQ, FKP created the questionnaire and collected the data; M-CQ and FKP performed the data management; CC analysed the data; MA, SA, SB, SFC, SE, MH, ZH, BK, DM-B, FN, A-SO, PP, JWP, CLP, FQ, ATRV, GR, LS, MAMS, NT, JW provided data and commented the manuscript; all authors approved the final version of the manuscript. FKP acted as guarantor.

Funding Fondation Mérieux and the Center of Excellence in Respiratory Pathogens provided financial support for this study. ATRV was supported by grants from CNPq (#307145/2021-2) and FAPERJ (E-26/201.046/2022).

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer-reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Bourema Kouriba http://orcid.org/0000-0002-3419-5972
Chan Leakhena Phoeung http://orcid.org/0000-0002-6538-3597
Firdausi Qadri http://orcid.org/0000-0002-4158-5995
Florence Komurian Pradel http://orcid.org/0000-0002-1023-0796

REFERENCES

1 Komurian-Pradel F, Grundmann N, Siqueira MM, *et al*. Enhancing research capacities in infectious diseases: The GABRIEL network, a joint approach to major local health issues in developing countries. *Clin Epidemiol Glob Health* 2013;1:40–3.
2 Dananché C, Paranhos-Baccalà G, Messaoudi M, *et al*. Nasopharyngeal Viral and Bacterial Co-Detection among Children from Low- and Middle-Income Countries with and without Pneumonia. *Am J Trop Med Hyg* 2022;106:1086–93.
3 Bénet T, Picot VS, Awasthi S, *et al*. Severity of Pneumonia in Under 5-Year-Old Children from Developing Countries: A Multicenter, Prospective, Observational Study. *Am J Trop Med Hyg* 2017;97:68–76.
4 Nasrin R, Uddin MKM, Kabir SN, *et al*. Xpert MTB/RIF Ultra for the rapid diagnosis of extrapulmonary tuberculosis in a clinical setting of high tuberculosis prevalence country and interpretation of 'trace' results. *Tuberculosis (Edinb)* 2024;145:102478.
5 Chedid C, Andrieu T, Kokhreidze E, *et al*. In-Depth Immunophenotyping With Mass Cytometry During TB Treatment Reveals New T-Cell Subsets Associated With Culture Conversion. *Front Immunol* 2022;13:853572.
6 Chedid C, Kokhreidze E, Tukvadze N, *et al*. Relevance of QuantiFERON-TB Gold Plus and Heparin-Binding Hemagglutinin Interferon-γ Release Assays for Monitoring of Pulmonary Tuberculosis Clearance: A Multicentered Study. *Front Immunol* 2020;11:616450.
7 Bayaa R, Ndiaye MDB, Chedid C, *et al*. Multi-country evaluation of RISK6, a 6-gene blood transcriptomic signature, for tuberculosis diagnosis and treatment monitoring. *Sci Rep* 2021;11:13646.
8 Megueya AL, Makuetche K, Scaccia N, *et al*. First report of New Delhi metallo-beta-lactamase-5 (NDM-5)-producing Escherichia coli isolates from water environment in Cameroon. *J Glob Antimicrob Resist* 2024;36:489–91.
9 Scaccia N, da Silva Fonseca JV, Megueya AL, *et al*. Analysis of chlorhexidine, antibiotics and bacterial community composition in water environments from Brazil, Cameroon and Madagascar during the COVID-19 pandemic. *Sci Total Environ* 2024;932:173016.
10 Milenkov M, Proux C, Rasolofoarison TL, *et al*. Implementation of the WHO Tricycle protocol for surveillance of extended-spectrum β-lactamase producing Escherichia coli in humans, chickens, and the environment in Madagascar: a prospective genomic epidemiology study. *Lancet Microbe* 2024;5:100850.
11 Saadatian-Elahi M, Picot V, Hénaff L, *et al*. Protocol for a prospective, observational, hospital-based multicentre study of nosocomial SARS-CoV-2 transmission: NOSO-COR Project. *BMJ Open* 2020;10:e039088.
12 Vanhems P, Endtz H, Dananché C, *et al*. Comparison of the Clinical Features of SARS-CoV-2, Other Coronavirus and Influenza Infections in Infants Less Than 1-Year-Old. *Pediatr Infect Dis J* 2020;39:e157–8.
13 Acuña MA, Kasanetz F, De Luna P, *et al*. Principles of nociceptive coding in the anterior cingulate cortex. *Proc Natl Acad Sci U S A* 2023;120:e2212394120.
14 Ooka T, Raita Y, Fujiogi M, *et al*. Proteomics endotyping of infants with severe bronchiolitis and risk of childhood asthma. *Allergy* 2022;77:3350–61.
15 Constantine-Cooke N, Monterrubio-Gómez K, Plevris N, *et al*. Longitudinal Fecal Calprotectin Profiles Characterize Disease Course Heterogeneity in Crohn's Disease. *Clin Gastroenterol Hepatol* 2023;21:2918–27.
16 Crowson CS, Gunderson TM, Davis JM III, *et al*. Using Unsupervised Machine Learning Methods to Cluster Comorbidities in a Population-Based Cohort of Patients With Rheumatoid Arthritis. *Arthritis Care & Research* 2023;75:210–9.
17 Chotalia M, Ali M, Alderman JE, *et al*. Cardiovascular subphenotypes in patients with COVID-19 pneumonitis whose lungs are mechanically ventilated: a single-centre retrospective observational study. *Anaesthesia* 2022;77:763–71.
18 Kaufmann L, Rousseeuw P. Clustering by Means of Medoids. Data Anal Based L1-Norm Relat Methods. 1987;405–16.
19 Gower JC. A General Coefficient of Similarity and Some of Its Properties. *Biometrics* 1971;27:857.
20 Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 1987;20:53–65.
21 Rosvall M, Bergstrom CT. Mapping change in large networks. *PLoS One* 2010;5:e8694.
22 R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. n.d. Available: https://www.R-project.org/
23 Loftus TJ, Shickel B, Balch JA, *et al*. Phenotype clustering in health care: A narrative review for clinicians. *Front Artif Intell* 2022;5:842306.
24 Zhang F, Jonsson AH, Nathan A, *et al*. Deconstruction of rheumatoid arthritis synovium defines inflammatory subtypes. *Nature New Biol* 2023;623:616–24.
25 Pires J, Kraemer JG, Kuenzli E, *et al*. Gut microbiota dynamics in travelers returning from India colonized with extended-spectrum cephalosporin-resistant Enterobacteriaceae: A longitudinal study. *Travel Med Infect Dis* 2019;27:72–80.
26 Rwego IB, Babalobi OO, Musotsi P, *et al*. One Health capacity building in sub-Saharan Africa. *Infect Ecol Epidemiol* 2016;6:34032.
27 Igboh LS, McMorrow M, Tempia S, *et al*. Influenza surveillance capacity improvements in Africa during 2011-2017. *Influenza Other Respir Viruses* 2021;15:495–505.
28 Herstein JJ, Lowe JJ, Wolf T, *et al*. Leveraging a Preexisting Global Infectious Disease Network for Local Decision Making During a Pandemic. *Clin Infect Dis* 2022;74:ciab660:729–33:.
29 Chen H, Yu Q, Xie J, *et al*. Longitudinal phenotypes in patients with acute respiratory distress syndrome: a multi-database study. *Crit Care* 2022;26:340.
30 Ashworth M, Durbaba S, Whitney D, *et al*. Journey to multimorbidity: longitudinal analysis exploring cardiovascular risk factors and sociodemographic determinants in an urban setting. *BMJ Open* 2019;9:e031649.
31 Tsai YC, Chien TW, Wu JW, *et al*. Sing the alluvial plot to visualize the network characteristics of 100 top-cited articles on attention-deficit/hyperactivity disorder (adhd) since 2011: bibliometric analysis.