Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal of Clinical Epidemiology 149 (2022) 36−44

Journal of Clinical Epidemiology

Toward an observatory of the evolution of clinical trials through phylomemy reconstruction: the COVID-19 vaccines example

Quentin Lobbé[a], David Chavalarias[a,*], Alexandre Delanoë[a], Gabriel Ferrand[b,c,d], Sarah Cohen-Boulakia[e], Philippe Ravaud[b,c,d], Isabelle Boutron[b,c,d]

[a]*CNRS, Complex Systems Institute of Paris Île-de-France, Paris, France*
[b]*Université de Paris, INSERM, INRAE, CNAM, CRESS, F-75004, Paris, France*
[c]*Centre d'Épidémiologie Clinique, AP-HP, Hôpital Hôtel-Dieu, F-75004, Paris, France*
[d]*Cochrane France, F-75004, Paris, France*
[e]*Laboratoire Interdisciplinaire des Sciences du Numérique, Université Paris-Saclay, CNRS, Paris, France*

Accepted 10 May 2022; Published online 27 May 2022

Abstract

Objectives: To visualize the evolution of all registered COVID-19 vaccine trials.

Study Design and Setting: As part of the living mapping of the COVID-NMA initiative, we identify biweekly all COVID-19 vaccine trials and automatically extract data from the EU clinical trials registry, ClinicalTrials.gov, IRCT and the World Health Organization International Clinical Trials Registry Platform. Data are curated and enriched by epidemiologists. We have used the phylomemy reconstruction process to visualize the temporal evolution of COVID-19 vaccines trials descriptions. We have analyzed the textual contents of 1,794 trials descriptions (last search in October 2021) and explored their collective structure along with their semantic dynamics.

Results: The structures highlighted by the phylomemy reconstruction processes synthesize the complexity of the knowledge produced by the research community. The reconstructed phylomemy clearly retrieves the five major COVID-19 vaccine platforms in the form of complete branches. The branches interactions reflect the exploration of a new approach to vaccine implementation moving from homologous prime vaccination to heterologous prime vaccination. Phylomemies also clearly identifies shifts in research questions, from vaccine efficacy to booster efficacy.

Conclusion: This new method provides important insights for the global coordination between research teams especially in crisis situations such as the COVID-19 pandemic.

Keywords: COVID-19; Vaccination; Phylomemy; Knowledge dynamics; clinical trials; co-word analysis

1. Introduction

Over the past 2 years, the ongoing COVID-19 pandemic has impacted a wide number of human domains: from economy to education and from public health to politics. Among others, Science swung early on into action to find both a cure and effective vaccine. This has resulted in an unprecedented volume of publications that have generated an information overload for the medical community. One of today's challenges is to synthesize the huge variety of the research avenues explored about COVID-19 research to improve coordination between the different research streams. Related work on this domain have focused on visualizing structured data or metadata to follow the mutation of the virus [1], the worldwide evolution of the pandemic [2] or the discovery of new treatments [3]. However, two points should be noticed. First, none of these works have

Guarantor: David Chavalarias, Complex Systems Institute of Paris Île-de-France, 113 Rue Nationale, 75013 Paris, France david.chavalarias@iscpif.fr

Data availability: The original COVID-NMA database can be downloaded at covid-nma.com. The preprocessing script can be downloaded at https://doi.org/10.7910/DVN/JTRI7A. The full list of root terms is available at https://doi.org/10.7910/DVN/JTRI7A. The reconstructed phylomemy is available for live explorations at http://maps.gargantext.org/phylo/vaccines_publications_10_2021/ and downloadable at https://doi.org/10.7910/DVN/JTRI7A.

* Corresponding author. Complex Systems Institute of Paris Île-de-France, 113 Rue Nationale, 75013 Paris, France. Tel./fax: +33-9-72-63-79-22.

E-mail address: david.chavalarias@iscpif.fr (D. Chavalarias).

https://doi.org/10.1016/j.jclinepi.2022.05.004

What is new?

Key findings

- The phylomemy reconstruction process applied to data automatically extracted from registries and annotated by epidemiologists allows identifying the evolution of main research questions.

What this adds to what is known?

- The phylomemy reconstruction process brings insights for the global coordination between research teams toward the creation of an observatory of the evolution of international clinical trials.

What is the implication, what should change now?

- We need to develop high-quality observatory of clinical trials based on data registered within clinical trials registries. New methods and tools such as the phylomemy reconstruction process are needed to explore these data and improve research planning coordination.

searched to analyze the temporal evolution of knowledge on COVID-19 (including the apparition of vaccines and their usage) using visualization techniques. Such analysis of past research may provide very important hints to understand current and future research investigations. Second, related work have focused on structured (meta) data and have rarely exploited the richness of the content of clinical trials. Nevertheless, the clinical trials available in the set of international primary and secondary trial registries [3,4] (i.e., all trials registered in the International Clinical Trials Registry Platform, Clinicaltrials.gov and the EU clinical trials registry) contain both large and precious information.

In the present work, we have thus designed a solution to visualize the content of not yet annotated textual fields (such as full trials descriptions) to reveal how knowledge evolves in pandemic times. More precisely, we have used the phylomemy reconstruction process [5] to reconstruct the temporal evolution of the semantic landscape of timestamped corpora of textual documents. We have applied our solution to the dataset on vaccines of the COVID-NMA database which results from an international initiative and provides a unique collection of highly curated, reviewed, and complete data on clinical trials based on mapping and reviewing trials registries.

This work has been intrinsically interdisciplinary by involving expertise in epidemiology, complex systems, visualization, and data science. The challenges to address are the following: How can we make the most of the crucial information stored in mutable and evergrowing databases of clinical trials? How can we create visualization to render the evolution of the database content and help epidemiologists interpret such visualizations?

2. Materials and methods

Our article aims at applying a new text mining method—phylomemy reconstruction—to reconstruct the temporal evolution of COVID-19 vaccines research. To that end, we choose to analyze a set of 1,794 clinical trials descriptions extracted from the COVID-NMA database. Our dataset has been collected and curated by the combined effort of epidemiologists, data integration, and complex systems researchers.

2.1. The COVID-NMA database

The COVID-NMA project is an international initiative aimed at providing a living mapping and a living systematic review of all trials assessing treatments and preventive interventions for COVID-19 [3,4]. The development of the COVID-NMA database relies on a full methodology designed to generate and make available a complete, comprehensive, integrated, nonredundant, and carefully annotated datasets on clinical trials. We automatically extract data from clinical registries on a weekly basis and provide assistance to epidemiologists on the curation and annotation process. Raw data are extracted from the EU clinical trials register, from the ClinicalTrials.gov managed by the United States National Library of Medicine, from the Iranian Registry of Clinical Trials registry and from the World Health Organization (WHO) International Clinical Trials Registry Platform—an international registry that assembles information on clinical trials registered in 17 primary registries to identify new trial assessing COVID-19 vaccines and update of previously registered trial records. Data are extracted from registries, annotated and enriched by epidemiologists, and then stored and made available through the COVID-NMA database.

2.2. Preprocessing the COVID-NMA database

To be used as input data, the COVID-NMA database needs to be preprocessed. But the unprecedented volume of trials related to COVID-19 has challenged our capacity to build a time-consistent and insightful visualization. We thus faced two main issues: dealing with the mutable nature of trials registries and collaboratively curating a core vocabulary from the trials descriptions.

We first note that international trials registries can be postupdated by research teams. Contrary to regular scientific publications, their textual contents are mutable by nature; the description of a given trial can be postupdated weeks after having been recorded. One can detail an experimental protocol or simply link some results to a trial. Usually, researchers can manually deal with such updates, but in our case, the tight temporality of the COVID-19

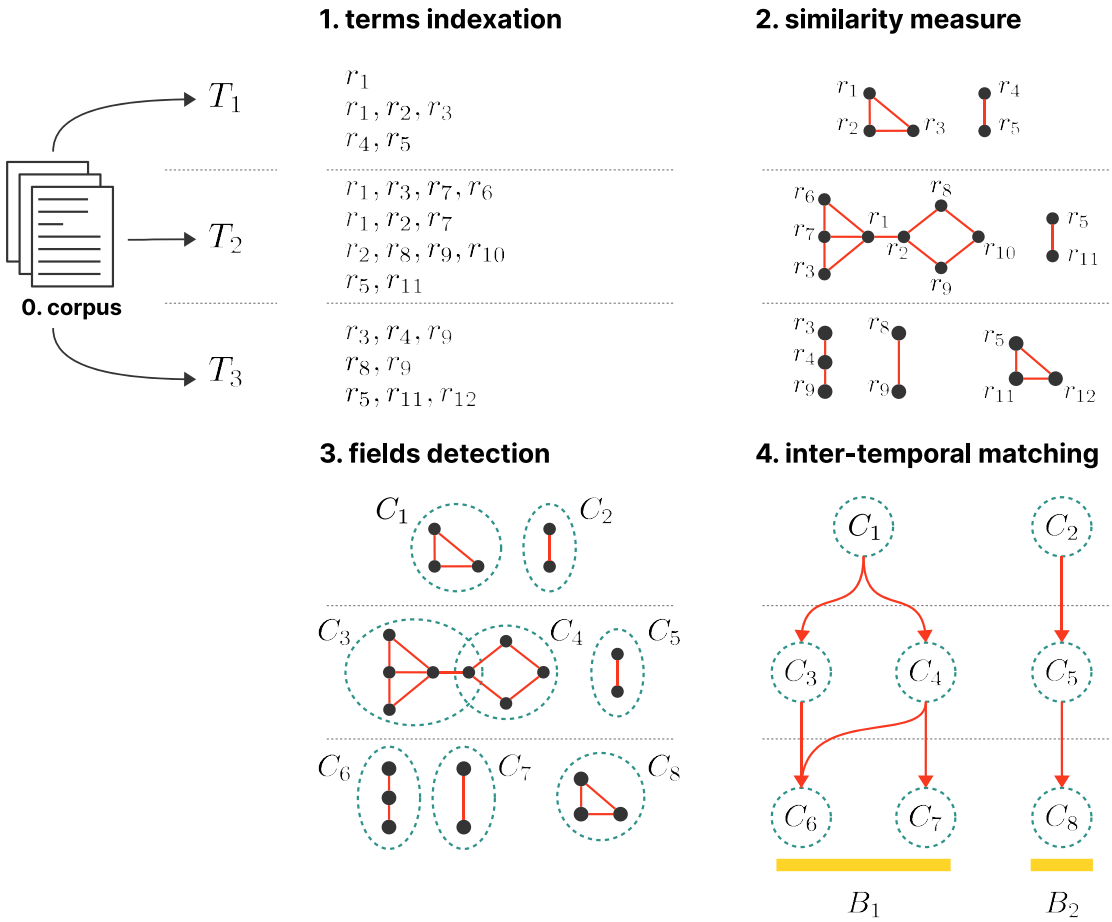

Fig. 1. The four operators of the phylomemy reconstruction process: (1) terms indexation, (2) similarity measures, (3) fields detection, and (4) intertemporal matching.

pandemic forced us to monthly reconstruct our visualization on top of hundreds of changeable trials. We have thus developed a time-consistent strategy; for each recorded trial, we have made the decision not to update the textual description of this trial. As subsequent trials referred to the first registered version of each vaccine's description, we chose to keep those first versions as references. By doing so, we do not break the temporal continuity of the phylomemy reconstruction process as we preserve the natural evolution of the descriptions. However, we have kept the metadata (i.e., trial phases, funding, associated publications, etc.) up-to-date with their most recent version.

The phylomemy reconstruction process requires the creation of a core vocabulary to visualize the evolution of the trials descriptions. To that end, we have first filtered from the COVID-NMA database a set of 1,794 records exclusively related to vaccination. Within each selected trial, we have merged the sections 'pharmacological treatment', 'treatment type', and 'treatment name' together to create a normalized description. The resulting corpus D_{vt} has later been collectively and collaboratively curated by epidemiologists; thanks to the free software Gargantext [6]. This software makes it possible to follow a human-driven approach where epidemiologists can validate and annotate each term they want to extract from the descriptions. We have thus created a core vocabulary as a list of 175 expressions, called root terms, that can have several variants.

2.3. The phylomemy reconstruction process

The phylomemy reconstruction process [5,7] combines advanced text-mining methods, scientometrics, and methods for the reconstruction of evolving complex networks to reconstruct the latent semantic structures of an unstructured—but timestamped—set of textual documents. Applied to a scientific corpus, it results in an inheritance network of research areas covered by all the collected publications. The phylomemy reconstruction process can be described as a combination of four subsequent operators summarized by Figure 1.

1. Terms indexation. By means of natural language processing algorithms and human validations—natural language processing algorithms and human validations are handled by the free software Gargantext [6]; we first extract from an original corpus

of documents (Fig. 1.0) a core vocabulary as a list $L=\{r_i|i\in I\}$ of sets r_i of equivalent expressions called roots (Fig. 1.1).

In our case study, the corpus is a set of 1,794 trials descriptions. The roots are all the technical and equivalent names (including characteristics variations and any misspelling) given for a same vaccine. For instance, the technical expressions ''rad5'' and ''rad26'' were aggregated into ''gam-COVID-vac''.

The corpus is then sliced into periods of interest $T^*=\{T_i\}_{1\leq i\leq k}, T_i\subset T$ for which roots' co-occurrences are computed.

In our case study, we consider 2 weeks periods starting every Monday from February 2020 to October 2021 and the output is a series of matrices of roots cooccurrences.

2. Similarity measure. Within each period of time and on the basis of its co-occurrences matrix, we estimate the semantic similarity between roots using the confidence measure [8]. The completion of this task results in a temporal series of graphs of similarity (Fig. 1.2).
3. Fields detection. For each period, a community detection algorithm—the frequent item set method [9]—is applied to detect subsets of densely connected roots within the graphs of similarity. Theses subsets C^T are called fields (Fig. 1.3) and their aggregated root expressions describe consistent research topics that were explored at a given period.

In our case study, the fields correspond to one or more descriptions of clinical trials sharing the same vaccine strategy. The output of this field detection step is a temporal series of clustering $C^*=\{C^T|T\in T^*\}$ with $C^T=\{C_j|j\in J^T\}$ and $C_j=\{r_i|r_i\in L, i\in I_j\subset I\}$ computed over all the periods. It describes all the research directions explored from February 2020 to October 2021.

4. Intertemporal matching. A temporal matching algorithm is then applied to identify meaningful kinship connections between fields from one period of time to another, that is, fields that belong to the same research stream. We finally highlight the different research streams B_k over time and called them branches of knowledge (Fig. 1.4).

The phylomemy reconstruction process makes it possible to draw the knowledge lineages at different resolutions through the tuning of a level of observation [5]. The complexity of the resulting semantic landscape can range from a wide 'continent' to an 'archipelago' of specialized branches of knowledge.

2.4. Visualizing phylomemies

The structures highlighted by a phylomemy reconstruction process synthesize the complexity of the knowledge produced by a research community. To make this newly reconstructed knowledge actionable and explorable, a phylomemy can be visualized as a temporal network [10]. Fields are represented by full circles and solid dark lines translate their kinship connections. Emerging terms (i.e., terms appearing for the first time in the phylomemy) are displayed over the whole structure as per the combined coordinates of their period and fields of appearance. Terms' size depends on their frequencies in the original corpus of trials. Branches are sorted from left to right so that closely related ones lie side by side. Interactive features can be used to reveal the entire field's content, follow the dissemination of a given term throughout the phylomemy, or simplify the scale of description of a selected branch.

3. Description of the resulting phylomemy

For our case study, we have used the corpus D_{vt} of 1,794 COVID-19 vaccines clinical trials (2.1) to reconstruct the weekly evolution of the research on COVID-19 vaccines between February 2020 and October 2021. Key expressions are extracted from the original descriptions of the tested vaccines, grouped into roots and then into fields. The reconstructed fields thus embody a set of trials at a given period of time. We here choose a level of observation $\lambda = 0.5$ to shape quite precise branches. The resulting phylomemy (Fig. 2) contains 175 roots and 550 fields distributed among 55 branches. The largest ones 'DNA-based vaccine', 'non-COVID vaccines', 'RNA-based vaccine', 'nonreplicating viral vector', 'inactivated virus' and 'protein subunit' are highlighted by yellow shades. Shades of blue indicate the proportion of randomized clinical trials among the total number of trials on which the corresponding field has been reconstructed. The visualization of a phylomemy can also offer its user to interactively highlight some key information, as, for example, the research paths addressing vaccine boost issues, highlighted in red at the bottom of this figure.

4. Following the worldwide tracks of COVID-19 vaccines

4.1. General observations

After having explored and analyzed Figure 2 alongside epidemiologists, we noticed that the reconstructed phylomemy clearly retrieves five major COVID-19 vaccine platforms in the form of complete branches. These platforms include the classical vaccine platforms, that is, 'nonreplicating viral vector', 'inactivated virus', and 'protein subunit' and the next-generation vaccine platform, that is, 'DNA-based vaccines' and 'RNA-based vaccines'. The visualization shows the continuous development of each branch and the way some of them started to interact and eventually blended while others stopped. Interestingly, trials of 'RNA-based vaccines' were registered very early in the course of the pandemic (February 2020) with trials

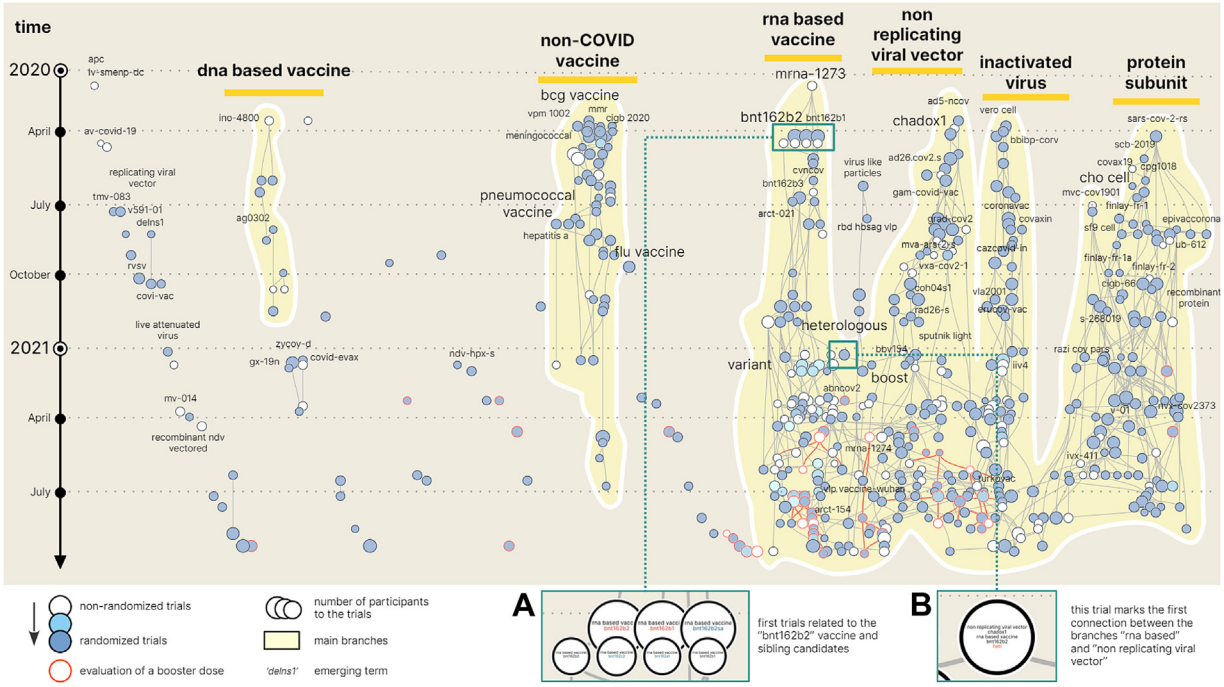

Fig. 2. Phylomemy of 1,794 COVID-19 vaccines trials recorded between February 2020 and October 2021 in the COVID-NMA database. Time goes by from top to bottom. The level of observation chosen for this reconstruction is $\lambda = 0.5$ (cf. [3]) to shape quite precise branches. Online and interactive version available at http://maps.gargantext.org/phylo/vaccines_publications_10_2021/.

evaluating the vaccine developed by Moderna TX (mRNA-1273) followed by the vaccine developed by Pfizer/BioNTech (BNT162b2) and sibling ones like BNT162b1 or BNT162b2sa that were not much longer tested (Fig. 2A). The number of trials increased rapidly and interactions with other widely explored techniques were observed shortly afterward, notably with the 'nonreplicating viral vector' family (ChAdOx1–AstraZeneca, Fig. 2B). The latest interaction involved the 'protein subunit' branch in July 2021. In contrast, 'DNA-based vaccines', with a first trial registered in April 2020, had a very limited number of trials planned and the whole branch stopped rapidly in 2020. Similarly, other platforms of 'replicating viral vector vaccine', 'virus-like particle vaccine', and 'live attenuated virus vaccine' showed a very limited development.

4.2. Repurposing non-COVID vaccines

As the development and approval of COVID-19 vaccines were expected to take time, researchers also explored repurposing non-COVID vaccines. Considering the lower severity of the disease in children and young adults, some researchers hypothesized the possible heterologous protective effect of these vaccines. Some evidence shows that live attenuated vaccines such as Bacille Calmette–Guerin and Measles, Mumps, Rubella can induce protective innate immunity, which could be central in controlling SARS-CoV-2 [11]. Although this hypothesis was appealing, it did not seem to expand into a wider research domain. The branch of 'non-COVID vaccines' appears and expands at the beginning of the pandemic but progressively decreases toward the end of 2020 as other more promising vaccines arose. Nevertheless, some researchers highlighted the need to adequately assess the use of non-COVID live attenuated vaccines as they could potentially boost response in high-risk populations, be used in addition to COVID-vaccines to increase effectiveness and durability of their effect, or be used to protect people exposed to COVID-19 patients [11].

4.3. Heterologous vaccination

The branches interactions reflect the exploration of a new approach to vaccine implementation moving from homologous prime vaccination (i.e., injections of two doses of the same vaccine) to heterologous prime vaccination (i.e., injection of the first dose of a given vaccine and the second dose of another vaccine). This is clearly shown in Figure 3 with the assessment of the heterologous prime vaccination of 'RNA-based vaccine' (BNT162b2-Pfizer/BioNTech) and 'nonreplicating viral vector' (ChAdOx1-AstraZeneka) in early 2021. This new approach was motivated by concerns about waning vaccine immunity but also by practical considerations. Following concerns about the safety of the AstraZeneca ChAdOx1 vaccine, the European Medicines Agency recommended giving a second dose Pfizer BNT162b2 vaccine to patients aged less than 55 years who received one dose of ChAdOx1-S-nCoV-19. Furthermore, decision makers needed flexibility to overcome the issue of vaccine availabilities during the vaccine rollout.

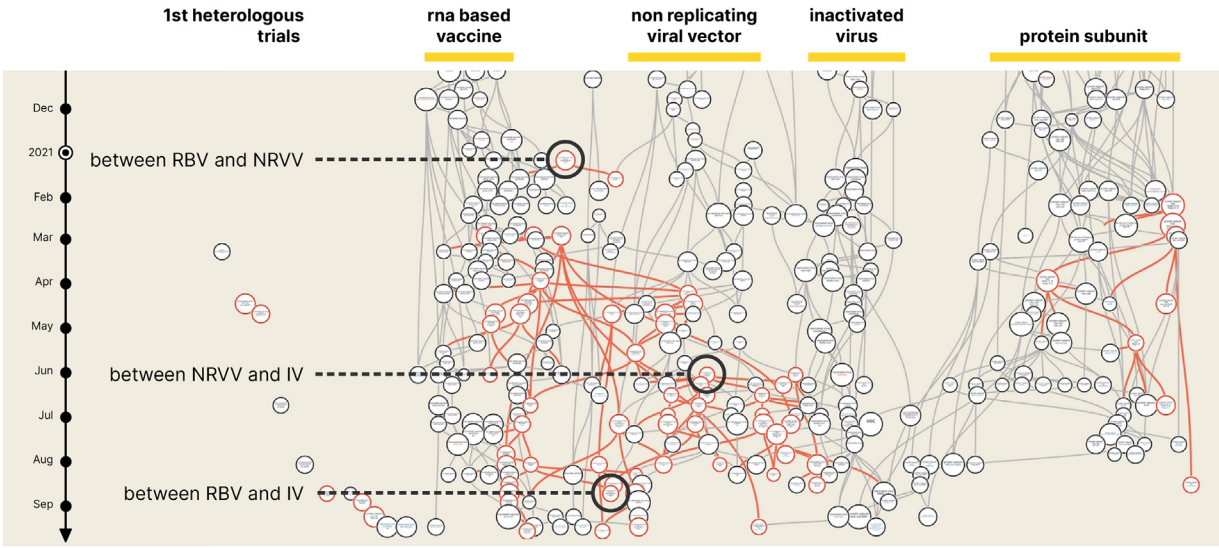

Fig. 3. A focus of Figure 2. In red are highlighted all the trials evaluating heterologous primary vaccination and heterologous booster. We circle the first heterologous trials involving different platforms.

This new approach proved to be relevant and other associations were evaluated: 'nonreplicating viral vector' and 'inactivated virus' in June 2021 and later 'RNA-based vaccine' and 'inactivated virus' in September 2021.

4.4. Boosters

Phylomemies are essential in identifying shifts in research questions. Although evidence of the beneficial effect of vaccines is mounting, research questions are moving toward exploring the effect of booster to overcome the waning of vaccine efficacy over time. Early in 2021, new trials assessing the impact of administrating a third dose (Fig. 2 shows red outline at the bottom) have been registered particularly for 'RNA-based vaccines' and 'nonreplicating viral vector' [12]. An important part of the research on boosters' effects is considering heterologous boosters.

4.5. Filters and specific research questions

By using metadata from the trials registries, we can filter the current phylomemy and thus push faceted observations to the fore or identify current main research questions. Phylomemies also provide important information on research planning and reporting. As shown in Figure 2, most trials registered are randomized controlled trials. Early in the pandemic, nonrandomized trials were primarily early-phase trials, whereas those registered in 2021 include both early-phase trials exploring new vaccines and phase-4 trials assessing vaccines safety.

We can explore the visualization to better understand how different countries participated in the overall research effort over time. For example, when filtering on the country (http://maps.gargantext.org/phylo/vaccines_countries_10_2021/), we see that trials conducted in the United States explored all vaccine platforms and that first registered trials frequently involved a center in the United States, confirming their leading role in clinical research (e.g., 'DNA-based vaccine', 'RNA-based vaccine', 'protein subunit'). Other important trials characteristics such as funding sources can also be highlighted (http://maps.gargantext.org/phylo/vaccines_fundings_10_2021/).

We also address the question of the publication of trial results (i.e., preprint or peer-reviewed articles). As shown in Figure 4, we currently have access to the results of a very limited number of planned trials. Although most of the COVID vaccine trials registered in early 2020 are published, most of the non-COVID vaccine trials are still unpublished. Thus this visualization can orient scientist toward a new question regarding the organization of the trials like the understanding whether these trials were actually conducted with unpublished results or were unable to recruit is an important issue. Phylomemies also offer flexibility and the possibilities to explore specific questions that may arise over time.

Using metadata and filters is a way to address specific questions through the phylomemy by highlighting and quickly finding subsets of trials within the registries. For example, there is currently some interest and concerns of decreased efficacy of vaccines in immunocompromised populations. It is possible to filter on this specific population and show that this specific population has been considered in clinical trials only recently (first time in 2021) and mainly for RNA-based vaccines. We can also make use of the 'type of patient' field to push trials related to specific public ('newborn', 'children', 'adolescent', 'pregnant women', etc.) to the fore of the visualization. A dedicated phylomemy can be explored at http://maps.gargantext.org/phylo/vaccines_publics_10_2021/.

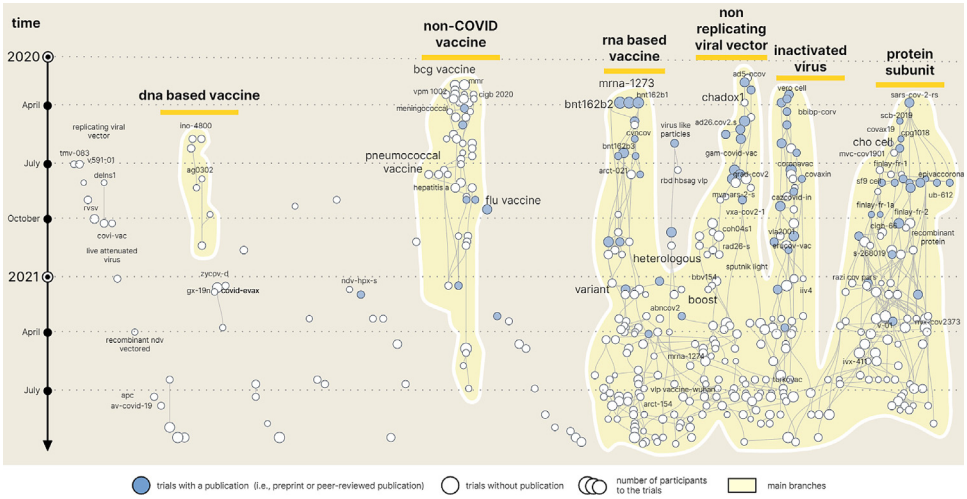

Fig. 4. Phylomemy of the randomized only COVID-19 vaccines trials. In blue, we highlight all the trials with an associated publication (i.e., preprint or peer-reviewed articles).

5. Discussion

By using the example of COVID-19 vaccine, our results show that phylomemies can improve our understanding and knowledge on research planning at a global level. At this stage, two main questions arise from our work: Is our method generalizable? Can we infer or predict future trends from the visualizations?

5.1. Why we need new approaches such as phylomemies?

The COVID-19 pandemic has been a clear illustration of the huge amount of avoidable research waste related to the lack of global coordination of clinical research [3]. To adequately plan clinical research and avoid research waste, we need new approaches. Researchers need tools to understand the current research landscape and prioritize research questions. However, data available are massive and one cannot synthesize easily the large amount of data generated (research planned and results produced). We need to develop infrastructures such as a global observatory of clinical research based on high quality data and tools to help stakeholders explore these data to improve our understanding of the ecosystem.

5.2. Generalization

Our case study is entirely focused on COVID-19 vaccines trials. But we defend that our approach is now generalizable to various contexts without extra effort. In the following, we develop this point by incrementally generalizing future domains of application as described hereafter:

- Integrating other COVID-NMA metadata. In consultation with epidemiologists, we have enriched our visualizations with the possibility to filter on a selected set of metadata such as the participants characteristics (age, pregnancy, etc.) (4.5). This list can be extended to all the structured fields of the COVID-NMA database by simply changing the preprocessing script (2.2). Furthermore, as phylomemies are designed to reveal the structure out of unstructured data, the COVID-NMA database could even be enriched with new fields such as subcategories of vaccine trials: ‘heterologous’, ‘boosters’, etc. In due time, our approach could influence the way scientists share trials information in registries by standardizing new metadata and unstructured textual content.
- Working on COVID-19 treatments instead of vaccines. Choosing between visualizing COVID-19 vaccines or treatments is also a matter of selecting the right field in the COVID-NMA database. Yet, we will still have to create a new core vocabulary (2.2). But thanks to Gargantext (the free text-mining software used to annotate the vaccines descriptions upstream from the phylomemies), it will only take a few days to collaboratively achieve this task and annotate hundreds of trials descriptions. A preliminary study can be found in [5].
- Visualizing trials related to another disease. Since the beginning of this study, the process designed to fill and integrate the COVID-NMA database has evolved making it possible to construct easily new databases gathering data on other kind of diseases. Indeed, the process to extract raw data from registries has a large generic step where we extract the description of each trial including information on its design, the inclusion/exclusion criteria for patients, the description of arms, and the set of outcomes. As an example, we have reconstructed the phylomemy of 1,798 trials related to Alzheimer disease and extracted from the WHO international registries. The resulting visualization can be explored at http://maps.gargantext.org/phylo/alzheimer/.

- Analyzing publication data. Phylomemies have been first designed to visualize the content of scientific publications. Thanks to the recent integration of phylomemies to the free software Gargantext, one can already reconstruct the temporal structure of various corpora of thousands of articles extracted from PubMed or from the Web of Science. In the context of a pandemic like COVID-19 pandemic, it would be relevant to explore in parallel both the phylomemy on trials and the phylomemy on articles or preprint to get a comprehensive view of the scientific landscape. A first exploration of COVID-19 literature can be found at http://maps.gargantext.org/maps/covid-19 in the form of a semantic graph.
- Monitoring various types of clinical trials on the fly. If all placed end-to-end, the continuous integration of WHO international registries within the COVID-NMA enrichment pipeline and through the phylomemy reconstruction process would enable the dynamic analysis of any kind of trials content as they arise. Such analytical workflow will require the creation of two teams: an integration team able to deal with the possible evolution of the registries and an annotation team dedicated to the creation and update of core vocabularies, but it would allow better coordination of scientific teams around the world in all medical fields and accelerate medical discoveries.

5.3. Prediction

Phylomemies are not aimed at predicting future trends, although the analysis of their dynamics could give some hints. They help us to understand the present states of dynamical processes regarding their former evolution. Inferring upcoming developments can be pursued by combining the knowledge of field's experts (like epidemiologists) and the comprehension of hidden dynamics relieved by the phylomemies. In the case of this article, the nature of the input data already place us ahead of regular scientific time by visualizing registries data instead of publications data. Indeed, except for pandemic time, there usually is a gap of 5 years between the first trial and the publication of a new treatment. Phylomemies associated to trials registries might consequently be scientific tools for understanding possible future treatments.

5.4. Perspectives and insights for COVID-19 research

Global coordination between research teams is a key for accelerating innovation in Science, especially during crisis situations such as the COVID-19 pandemic. Reducing redundancies and providing heuristics to find new search paths as they arise can save time and lives [3]. We claim that phylomemy reconstruction could be instrumental to guide trialists, funders, and decision makers in biomedical research. In times of crises, it would enable them to better adapt to the evolution of the situation by following main research questions and identify less promising domains. It could also facilitate the identification of research gaps, research questions that may have been abandoned prematurely, and redundancy in research.

In a world where experts are increasingly specialized, our approach could draw attention to alternative solutions developed in other branches of science or to problems already encountered in research direction to be explored. It could also lead to new conceptual operations to be performed on a knowledge database, such as "give me all the branches of knowledge that are merging" or "suggest a promising combination of compounds to test". This could both accelerate research by making tangible the latent structure of innovation and promote collaborations between teams that would not otherwise be interested in each other's work. Phylomemy reconstructions may thus become collective and reflective tools to foster the worldwide collective coordination between researchers. This revolution in clinical trial processing is within reach. Nevertheless, it would imply having access to high-quality data on research planning and protocol.

Our case study focuses on a single disease, but this approach is fully generic and we call for a worldwide observatory for monitoring the dynamics of clinical trials. As it scales up, our approach could be implemented for any disease or research field.

Author contributions

All authors contributed equally to the paper. Gabriel Ferrand was responsible for the data collection process. Quentin Lobbé, David Chavalarias and Alexandre Delanoë were responsible for the text mining, data visualization processes, and perspectives. Sarah Cohen-Boulakia was responsible for the data integration process. Philippe Ravaud and Isabelle Boutron were responsible for the data interpretation process.

References

[1] Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of sars-cov-2 genomes. Proc Natl Acad Sci U S A 2020; 117:9241−3.
[2] the COVID-19 APHP-Universities-INRIA-INSERM Group, B. Early indicators of intensive care unit bed requirement during the covid-19 epidemic: a retrospective study in ile-de-France region, France. PLoS One 2020;15:1−12.
[3] Nguyen VT, Rivière P, Ripoll P, Barnier J, Vuillemot R, Ferrand G, et al. Research response to coronavirus disease 2019 needed better coordination and collaboration: a living mapping of registered trials. J Clin Epidemiol 2021;130:107−16.
[4] Boutron I, Chaimani A, Meerpohl JJ, Hróbjartsson A, Devane D, Rada G, et al. The COVID-NMA project: building an evidence ecosystem for the COVID-19 pandemic. Ann Intern Med 2020;173: 1015−7.

[5] Chavalarias D, Lobbé Q, Delanoë A. Draw me science − multi-level and multi-scale reconstruction of knowledge dynamics with phylomemies. Scientometrics 2021;22:1−31.
[6] Delanoë A., Chavalarias D. Mining the digital society - Gargantext, a macroscope for collaborative analysis and exploration of textual corpora. forthcoming.
[7] Chavalarias D, Cointet JP. Phylomemetic patterns in science evolution—the rise and fall of scientific fields. PloS one 2013;8:e54847.
[8] Dias G, Mukelov R, Cleuziou G. Mapping general-specific noun relationships to wordnet hypernym/hyponym relations. In: International conference on knowledge engineering and knowledge management. Berlin, Heidelberg: Springer; 2008:198−212.
[9] Uno T, Kiyomi M, Arimura H. Lcm ver. 2: Efficient mining algorithms for frequent/closed/maximal itemsets. In: Fimi.
[10] Lobbé Q, Delanoë A, Chavalarias D. Exploring, browsing and interacting with multi-level and multi-scale dynamics of knowledge. Inf Vis 2021;21(1):17−37.
[11] Chumakov K, Avidan MS, Benn CS, Bertozzi SM, Blatt L, Chang AY, et al. Old vaccines for new infections: exploiting innate immunity to control covid-19 and prevent future pandemics. Proc Natl Acad Sci U S A 2021;118.
[12] Krause PR, Fleming TR, Peto R, Longini IM, Figueroa JP, Sterne JAC, et al. Considerations in boosting COVID-19 vaccine immune responses. Lancet 2021;398:1377−80.