Paris, July 7th, 2023

Dear Recommender,

Thank you for handling our manuscript and for the positive comments. We hereby submit the revised version, in which we have addressed the comments as follows.

Your editorial comments:

Dear authors, based on 2 independent reviews and in my own opinion your preprint is interesting and has merits, nevertheless, there are some issues you must fix in order to make it suitable for being recommended.
Best regards,

> **Our answer:** Thank you. We address the specific points you mention, in the answers to reviewers below.

Reviewers' comments:

Reviewer #1: *Reviewed by Ankur Mutreja, 05 May 2023 09:31*

Please go through my comments.

Summary: The authors of this study used DIPTHOSCAN to analyse the resurgence of diphtheria in France. The authors also added publically available data to develop a clearer picture. Using a combination of genome sequencing and phylogenetic analysis, they characterized the global population structure of the bacterium C. diphtheriae and identified factors that contribute to its ongoing spread and resurgence. The manuscript presents a valuable contribution to the field of infectious disease research, and the authors demonstrate a deep understanding of the complex genetic factors that underlie diphtheria's re-emergence.

> **Our answer:** Thank you for your appreciation.

Major comments:

1. Despite the author's assertion that DIPTHOSCAN is an easy-to-use tool, we observed certain issues in installing and using the publicly available DIPHTOSCAN tool that needs to be fixed. The fact that we found only the output file with information, such as species match, STs, 7 alleles, and tox types, led us to notice that some of the options listed were not executed.

> **Our answer:** Thank you for having tested diphtOscan. We have tested the tool on several computers and in our hands, it worked fine, including the options. We'd be happy to look into the specific issues you encountered, if you are willing to provide more details on the issue (computer used, system, dependencies missing?). Note that we have addressed a ticket that was open in github by an anonymous user; it dealt with a link in the AMRfinderPlus option that had to be modified; this was done on 28/04/2023 and should have solved the problem you encountered.

2. Another major concern is that the manuscript focuses almost exclusively on genomic and genetic factors and does not give adequate attention to the epidemiological, vaccine escape and clinical aspects of diphtheria.

Our answer: We believe our work touches on multiple epidemiological aspects, such as the strain diversity and links of isolates between geographical regions; as well as some clinical aspects, as clinical data were provided and discussed. We'd be happy to address more specific points if possible, given the retrospective nature of our work which limits the epidemiological and clinical data at hand. Regarding vaccine escape, we are afraid the design of our study does not enable us to explore this question; we believe it would require pre-vaccination era isolates datasets to see the shift in *tox* gene allele frequencies, for example. Nevertheless, our work places the vaccine strain PW8, mentioned at several parts of the manuscript, in the broad phylogenetic context of C. diphtheriae and its *tox* gene allelic diversity. As the translated amino-acid sequence of most frequent *tox* alleles in extant populations is 100% identical to the vaccine strain, there is no evidence for strong vaccine-escape evolution at first sight. We have added a sentence on *tox* alleles and toxin protein variation to link our work with a vaccine escape discussion for toxin variants in a previous work (Will et al., 2021).

3. The authors in the manuscript should have emphasised the reason for the disease re-emergence and the role of pathogen-associated genes role in colonisation and disease transmission.

Our answer: Our work has provided clues regarding the microbiological understanding of emergence; for example, it provides a detailed understanding of which sublineages are driving the emergence, which ones are newly observed and which ones were previously described; and their genomic and phenotypic features. We also show how human migrations have been driving the specific emergence observed in 2022. To address the comment of the reviewer, we have now also added a sentence on vaccination status of patients (end of first paragraph of Results). To highlight this factor as contributing to explain the emergence, we have added the sentence in the discussion: "The occurrence of cases of diphtheria among migrants, the vast majority of whom are not up to date with their vaccinations, raises concerns of the emergence of cluster cases in accommodation facilities for migrants, refugees or asylum seekers (Badenschier et al., 2022; Kofler et al., 2022). Professionals dealing with these populations need to be particularly vigilant in spotting clinical signs of diphtheria and ensuring that their vaccinations are up to date."

4. It would be valuable if authors could clarify the role of possible non-genetic factors in this resurgence of diphtheria, as well as evaluate the significance of their findings for public health policy and clinical practice.

Our answer: Thank you, please see above our answer to comment 3, where we added data on lack of vaccination and its association with this reemergence.

Minor comments:

1. It is better to include the number of tox-negatives (line: 103).

Our answer: We have added the number, thank you.

2. While stating the number of sequenced strains in the aim (line: 105) and methods (line: 122-124; line: 128-130), the numbers do not match.

Our answer: Thank you. We have modified and provided precisions.

3. Typo error (line: 288) NTCT.

Our answer: Thank you, corrected.

4. The authors may include more specific information on tox types and help readers comprehend how it differs from previous tox group classifications.

Our answer: To our knowledge, only one previous study analyzed comprehensively the diversity of tox (Will et al). We now provide in Figure S8, correspondence of *tox* alleles with this former study, based on the common isolates found between the two studies.

5. What is the fitness gained by some of the SL to be predominant for causing an increased number of infections?

Our answer: Thank you for the interesting question. It is generally complex to pinpoint the particular determinants that confer a fitness advantage to any particular sublineage, as this is largely a multifactorial process and impossible to test in nature. Novel resistance genes, tox alleles or fimbriae or siderophore variations might all be involved in epidemiological reemergence. To acknowledge the important question raised by the reviewer, we have added 'to understand the microbiological determinants of (re)emerging sublineages,' in the discussion section.

6. While the authors present a wealth of data and analysis, for impact, the manuscript could be improved to make it a better read for a broader audience that doesn't understand the technicalities of phylogenomics.

Our answer: Thank you. We have slightly edited the manuscript in this direction. Part of the Results on diphtOscan developments were moved to Methods. We have described the phylogenetic structure of C. diphtheriae in what we believe are simple terms, but would be open to changes based on more specific suggestions.

Overall, this is an important and timely contribution to the field of infectious disease research, and the authors are to be commended for their comprehensive genomic analysis of diphtheria's re-emergence. With above comments addressed, I will support the acceptance of this work.

Our answer: Thank you very much for your very positive statement.

Reviewer #2: *Reviewed by anonymous reviewer, 31 May 2023 16:38*

Bioinformatics tool and global population framework was used to analyzed *C. diphtheriae genome. T*his work is very useful in examining the various genetic variants that frequently emerge in the *C. diphtheriae* infections. The sample size discussed was heterogenous and covered various sublineages. Overall, the study is an important contribution to the field.

Our answer: Thank you very much for your very positive evaluation.

Reviewer #3: *Reviewed by anonymous reviewer, 21 Jun 2023 14:17*

The preprint by Hennart et al, presents an extensive work for the development of a new bioinformatics platform aimed to analyze C. diphtheriae genomes focusing in relevant data such as the presence of a functional tox gene, other genes encoding virulence factors, genes encoding antibiotic resistance and genes that allow biotype discrimination.

The platform performance was evaluated using different sets of genomes including several from recently isolated bacteria in a context of Diphtheria reemergence.

The manuscript is well written and well explained; however, my main criticism is that it is very long and too technical and hence will be difficult to understand by non-bioinformatics readers, I recommend to shorten the method section (perhaps send some material to the supplementary methods).

Our answer: Thank you. We have shortened the Results section of the manuscript by moving some parts on diphtOscan to the Methods section. As it stands, the Methods section on bioinformatics methods does not seem very long to us, and a lot of details are indeed in the supplementary appendix already.

Also, although regarding the production of the Diphtheria toxin, the platform gave results that were mostly in agreement with the experimental determination of the gene presence by qPCR and its functionality (Elek test).

It predicted that only 50% of the non-toxigenic isolates were indeed non-toxigenic (L463), hence wrongly classifying 50% of the real non-toxigenic isolates as toxigenic.

And then in the discussion (L 659-661) you mention: “These cases may be attributable to (i) a lack of detection by the Elek test due to a low level of expression of the toxin gene in some strains, or (ii) yet unknown genetic mechanisms that abort tox gene expression entirely (unexplained true NTTB).”

Hence I would like to ask you what is the detection limit of Elek test? A very low expression of Tox would be equivalent to non-production?

Our answer: There is no detection limit of Elek’s test as this is not a quantitative assay; the Elek test is defined as positive when the human reader sees an immunoprecipitation arc, from a bacterial colony. Hence yes, if a strain produces very small amounts of toxin in the test, it might be wrongly classified as non-toxigenic.

Furthermore, perhaps these strains do not produce detectable toxin levels due defects in known tox gene regulation mechanisms, for example mutations in the regulatory region at the promoter level, or mutations in the main regulator gene encoding DtxR. Perhaps it mutates and becomes a supper repressor that represses the tox expression even in the absence of iron, etc.

Please add a section on tox gene regulation in the discussion and consider searching mutations in the regulatory regions of tox gene and in their regulators with your bioinformatics tools.

> **Our answer:** Thank you very much for this suggestion. We have further investigated the regulatory region upstream of the tox gene (300 nt upstream of tox start codon), and the DtxR sequence variation in face of toxigenicity or NTTB status. We found little variation and none that was associated with NTTB status. We have added a sentence on this in the Results, and described this analysis in Methods.

Minor:

1) L 162 “for 1 h DNA” add “,” after “h”.

> **Our answer:** Thank you for noticing; we have added a ‘.’, rather.

2) Gene and species names should be in italic in the references.

> **Our answer:** Thank you, we have re-formatted the bibliographic references according to PCI style.

Additional changes:

1. Following reviewer #1 comment, we have added Figure S8 with correspondence between our tox allele nomenclature and previous *tox* gene variants (Will et al). Fig S9 and S10 are previous Figures S8 and S9, respectively.
2. A small error of label was corrected in the ribotype phylogeny figure (ribotype Rossija was mixed-up in our analysis, wrongly suggesting a placement in the Mitis branch; it is now placed in the tree close to Sankt-Petersburg, consistent with previous literature (Grimont et al)
3. Table S1 with sublineages nomenclature was updated for 8 strains (fusion of groups)
4. The diphtOscan script was updated to incorporate a fix on an installation issue.