McMASTER UNIVERSITY CONFIDENTIAL EXAMINER'S REPORT ON Ph.D. THESIS

Student No.	001336803	Department	Anthropology

Student Name	Katherine Eaton

Thesis Title	Big data, small microbes: Genomic analysis of the plague bacterium Yersinia pestis

Contribution to Knowledge: Originality, quality, quantity.	Excellent
Research (and Experimental Work, if applicable): Adequacy, thoroughness, technical skill (so far as can be judged on internal evidence)	Excellent
Understanding of the Subject: Adequacy of the review of previous work, choice of project, evaluation of results, validity of conclusions.	Excellent
Presentation: Organization of material, presentation of results (tables, figures), grammar, style, bibliography.	Excellent
Quality of Typescript: Makeup, neatness, freedom from errors.	Excellent
FINAL JUDGEMENT OF THE THESIS AS A WHOLE:	
The thesis can proceed to defence and is Excellent	

CONDITIONS OF APPROVAL, POINTS FOR DISCUSSION AT THE DEFENSE, OTHER SUGGESTIONS

The Ph.D. thesis manuscript prepared by Katherine Eaton is untitled 'Big data, small microbes: genomic analysis of the plague bacterium Yersinia pestis'. It presents within a total of 67 pages the work that she carried out under the supervision of Dr. Hendrik Poinar at the Department of Anthropology, Mc Master University.

The first section (pages 1-5) provides a brief introduction to the main research question, intertwined with personal anecdotes, and is written in such a style that it almost reads like a scientific book. The quality and clarity of the writing are not limited to this section only but are also apparent throughout the whole manuscript.

The second section (pages 6-11) is built around the first scientific article that Katherine Eaton published in 2020 as a single author. This article describes a computational package that she developed to automate the retrieval of ever-expanding pathogenic genome data from sequence repositories, together with metadata providing context on samples, sites, and underlying experimental work. That Katherine Eaton succeeded in developing and publishing such a computational tool on her own is impressive and speaks for her scientific maturity and excellence at designing and performing research, coding, and writing. These represent not less than the full range of qualities that are expected from professional scientists in her research area. This computational package has been tremendously useful for conducting the research work that Katherine Eaton presents in the following two sections. It will be equally useful for the community, as it provides a simple solution to important but recurrent technical challenges, such as accessing the most up-to-date data and traceability. The package distribution under open-source licensing shall warrant its wide adoption by the community.

The third section (pages 12-40) describes a study that is under review for publication at Nature Communications. It presents an analysis of an extensive reference panel of Yersinia pestis genomes, aimed at scaffolding and

dating the main phylogenetic nodes that underpinned the evolutionary trajectory of this pathogen. This work presents some of the most thorough investigations of the available genome sequence data, and does not only identify issues with some of the dating procedures previously (and most generally) applied, but also provides solutions. The quality of the figures presented in this work must be particularly emphasized, especially figure 5, which clearly conveys the main findings in the most compelling manner while summarizing sometimes complex analyses. While the implications of this study are manifold, perhaps the most valuable is to provide a careful evaluation of what ancient DNA data can or cannot do when it comes to reconstructing the short -term evolutionary trajectory during the second plague pandemics. Being able to honestly criticize and report on the main limitations of their own findings is also one of the most important, if not the most important, qualities of professional scientists.

The fourth section (pages 41-50) consists of a study submitted for publication in PNAS. it presents a longitudinal characterization of plague epidemics in Denmark between 1000 and 1800 CE, leveraging an extensive sample set of almost 300 archaeological bones. Collecting and analyzing data for such an amount of remains represents an incredible amount of work and, thus, an outstanding achievement. This resulted in the characterization of 9 ancient plague genomes, which provided new insights into the temporal and geographic dynamics of the second pandemics in the country, sometimes refining historical models based on textual evidence.

Finally, the last section (pages 51-54) provides a short conclusion to the Ph.D. work, and explores future directions.

To summarize, I have very much enjoyed reading Katherine Eaton's Ph.D. work and I am happy to recommend her work for the defense. I have no hesitations in saying that I consider this work amongst the top-10% Ph.D. work that I have assessed over the last 10 years, including the work from my own students. I am thus truly looking forward to the defense.

I would appreciate if it could provide the opportunity to discuss some of the following points:
-the future maintenance plans and actions for the computational package (especially for handling even larger data sets)
-the choice of the demographic models underlying the dating analyses presented in section 3
-the possible impact of dates modeling as point estimates and not full distributions
-the comparison of those timing estimates obtained and some of those most recently published, especially for the second pandemics
-the motivation for using some of the coverage thresholds, and Beastv1 instead of Beastv2
-the benefit for real-time PCR testing samples
-whether new wet-lab procedures could be developed to improve detection success
-whether radiocarbon dating data were obtained on the Danish samples
-the main limitations in current approaches to studying plague evolutionary history
-possible experimental design to study (epi)genetic changes in the host genome as a response to past pandemics
and more.

External Name	Ludovic Orlando	Date	2022-01-08 6:35:58AM