Geographic microevolution of *Mycobacterium ulcerans* sustains

Buruli ulcer extension, Australia

Jamal Saad^{1,2+}, Nassim Hammoudi^{1,2+}, Rita Zgheib^{1,3}, Hussein Anani^{1,3}, Michel Drancourt^{1*}.

1. IHU Méditerranée Infection, Marseille, France.

2. Aix-Marseille-Université, IRD, MEPHI, IHU Méditerranée Infection, Marseille, France.

3. Aix Marseille Univ, Institut de Recherche pour le Développement (IRD), Service de Santé des Armées, AP-HM, UMR Vecteurs Infections Tropicales et Méditerranéennes (VITROME), Institut Hospitalo-Universitaire Méditerranée Infection, Marseille, France.

+ These two authors equally contributed to this work.

*Corresponding author. Michel DRANCOURT, Institut Hospitalo-Universitaire Méditerranée-Infection, 19-21 Boulevard Jean Moulin, 13385, Marseille cedex 05, France ; tel +33413732401; fax : +33413732402; e-mail: michel.drancourt@univ-amu.fr

Word count Abstract = 235

Word count Text = 2,389

Figures number: 5

Tables number: 7

Keywords = *Mycobacterium*, *Mycobacterium ulcerans*, Buruli ulcer, Australia, whole genome sequencing, diversity.

Abstract. The reason why severe cases of Buruli ulcers caused by *Mycobacterium ulcerans* are emerging in some South Australia counties has not been determined. In this study, we measured the diversity of *M. ulcerans* complex whole genome sequences (WGS) and reported a marker of this diversity. Using this marker as a probe, we compared WGS diversity in Buruli ulcer-epidemic South Australia counties versus non-epidemic Australian counties and further refined comparisons at the level of counties where severe Buruli ulcer cases have been reported. Analyzing 218 WGS (35 complete and 183 reconstructed WGS, including 174 Australian WGS) yielded 15 *M. ulcerans* complex genotypes, including three genotypes specific to Australia and one genotype specific to South Australia. A 1,068-bp PPE family protein gene exhibiting genotype-specific sequence variations was employed to further probe 13 minority clones hidden in sequence reads. The repartition of these clones significantly differed between South Australia and the rest of Australia. In addition, a significantly higher prevalence of 3/13 clones was observed in South Australia counties of the Mornington Peninsula, Melbourne and Bellarine Peninsula than in other South Australia counties. The data presented in this report suggest that the microevolution of three *M. ulcerans* complex clones drove the emergence of severe Buruli ulcer cases in some South Australia counties. Sequencing one specific PPE gene served to efficiently probe *M. ulcerans* complex clones. Further functional studies may balance the environmental adaptation and virulence of these clones.

Introduction.

Buruli ulcer, a handicapping skin and subcutaneous infection, was initially described in Australia along with the description of the first isolate of the causative *Mycobacterium ulcerans* (*M. ulcerans*).¹ Thereafter, Buruli ulcers have been reported in several tropical areas of Asia, Africa and South America; therefore, 36 countries are currently notifying Buruli ulcer cases to the World Health Organization (WHO). In these endemic countries, the evolving incidence of Buruli ulcer is highly variable: the WHO is reporting a 64% decrease in notified Buruli ulcer cases worldwide over the past 9 years:¹ in fact, the incidence of Buruli ulcer is decreasing in most West African countries, which declared the highest numbers of cases during the 1980-2000 period, whereas it is increasing in its birthplace, South Australia.^{2,3} The infection is specifically re-emerging in Victoria, southeastern Australia, extending to previously Buruli ulcer-free countries where severe forms are now encountered.² For West Africa, we previously reported a significant correlation between dropping incidence in Buruli ulcers and climate change,⁴ while the reasons for the pejorative epidemiological evolution in Australia are unknown, although the hypothesis of currently undetected virulent clones of *M. ulcerans* has emerged.^{5,6}

Buruli ulcer is a non-transmissible disease contracted by direct or indirect contacts of breached skin with *M. ulcerans*-contaminated environments.^{1,7} Therefore, the evolution of the incidence of Buruli ulcers in any one endemic area as well as its geographical extension in any one country, partly reflects the evolving dynamics of *M. ulcerans* in the ecosystems where this opportunistic pathogen is thriving. *M. ulcerans* is a close derivative of *Mycobacterium marinum* (*M. marinum*) after their common ancestor evolved following a strong reduction of the genome and the acquisition of a 174-210-kb giant plasmid,⁸ encoding the nonribosomal synthesis of macrolide exotoxins named mycolactones.⁹ *M. ulcerans* is repeatedly detected in Buruli ulcer-endemic tropical countries in stagnant water ecosystems

to which rural populations that are developing Buruli ulcer are exposed;^{10,11} but only two environmental *M. ulcerans* isolates have been reported for which no genome sequence is available,^{12,13} meaning that the knowledge regarding the genomic diversity of the pathogen is limited to that of clinical isolates, which may not reflect the actual genomic diversity of *M. ulcerans* in the environment. Accordingly, the genomic microevolution of *M. ulcerans* clinical isolates has been reported in the Democratic Republic of Congo and Angola,¹⁴ but such observations were not placed in the background of evolving Buruli ulcers in the same geographic areas.

In this study, by analyzing the whole set of genomic data issued from studies of *M. ulcerans* in Australia,¹⁵ we reconstituted WGS data and observed the genomic microevolution of isolates specifically associated with the geographical expansion of *M. ulcerans* in coastal areas in South Australia¹⁵ and found one specific marker for that evolution.

Materials and methods.

Isolate culture and sequencing. Total DNA of nine isolates (including 01 *M. marinum* B and 08 *M. ulcerans*) maintained in the Reference Mycobacteriology Laboratory, IHU Méditerranée Infection, France, was extracted from several colonies of a 6-week-old subculture on Coletsos culture medium using the InstaGene matrix following the manufacturer's instructions (Bio-Rad, Marnes-la-Coquette, France). Then, the DNA (0.2 µg/µL) was sequenced using a MiSeq platform (Illumina, Inc., San Diego, CA, USA). DNA was fragmented and amplified by 12 cycles of PCR. After purification on AMPure XP beads (Beckman Coulter, Inc., Fullerton, CA, USA), the libraries were normalized and pooled for sequencing on a MiSeq instrument.¹⁶ Each sequencing run included one negative control (free of mycobacterial DNA) with ten mycobacterial samples to evaluate the quality

of output sequence reads. Furthermore, data reads from different laboratories were used in this study.

Genomic data analysis. In the first step, we retrieved WGS data for 44 *M. ulcerans* and *M. marinum* WGS (including the 09 WGS reported above), as well as inside reads deposited for 174 WGS made from *M. ulcerans* isolates in Australia (S1 Table).^{2,3,15} Then, the 182 *M. ulcerans* genome sequences and one *M. marinum* M were assembled and annotated using Spades version 3.14.0¹⁷ and Prokka version 1.14.6¹⁸, respectively. Roary 3.13.0¹⁹ was used with 95% identity to generate the pangenome, core genome and SNP-core genome of the 182 studied WGS (based on standard assembling software, such as SPAdes, A5-assembly and De novo). Moreover, PHYRE2 Protein Fold Recognition Server, CPHmodels-3.2, SCRATCH protein predictor (<http://scratch.proteomics.ics.uci.edu/index.html>) and Protter (<http://wlab.ethz.ch/protter/start/>) were used to predict the structure localization and function of genes of interest. CLC genomics 7 was used to map the output sequence reads against a gene and to detect the probability variant to estimate the zygosity (homozygous and heterozygous) of strains. Artemis software was used to visualize the heterogeneity profile of the studied strains (<https://www.sanger.ac.uk/science/tools/artemis>). WebLogo3 (<http://weblogo.threeplusone.com/create.cgi>) was also used to compare the gene sequences. We determined the properties of each genome (total length, G+C content, number of contigs and others) with Quast 5.0.2.²⁰ Average nucleotide identity (ANI) was calculated using OrthoANIu²¹ to validate the classification results of *M. ulcerans* isolates. Moreover, Seaview version 4²² was used to view the alignment profile of the core genome and to detect differences in substitution mutations, deletions, or insertions between isolates.

Results.

Whole genome analyses. We observed 41 different G+C% values among the 218 studied genomes, with G+C% values varying between a maximum of 65.8% for *M. marinum* strain DL240490 from Israel to a minimum of 63.73% for *M. ulcerans* SRR6346311 from Queensland, Australia and 63.37% for *M. ulcerans* CSURP9951 from Japan. Then, we observed that the genome size varied between a maximum length of 6.66 Mb for *M. marinum* strain M and a minimum length of 5.19 Mb for *M. ulcerans* SRR6346343 isolated from *Trichosurus vulpecula* in Gippsland, Australia (S1 Table). Studying ANI, core genome and SNPs based on core genome sequence (2,814-genes; 2,607,386-bp) data indicated six different *M. ulcerans* genotypes comprising 188 *M. ulcerans* isolates (A-F), one *Mycobacterium liflandii* genotype comprising two isolates (G), one *Mycobacterium pseudoshottii* genotype comprising four strains (H) and seven *M. marinum* genotypes comprising 24 isolates (I-O) (Figs 1a, 1b) (S2, S3 Tables). Focusing on the 174 *M. ulcerans* isolates made in Australia, we observed three different genotypes: genotype A included 163 clinical isolates, with SNPs varying between 1 and 950 (including isolates originating from six Australian counties, i.e., (Ballarat, Mornington Peninsula, Phillip Island, Melbourne, Gippsland and Bellarine Peninsula) and 06 animal isolates (01 *Canis lupus familiaris*, 03 *T. vulpecula*, 01 “koala”, 01 *Potorous longipes*) originated from four counties (Mornington Peninsula, Phillip Island, Gippsland and Bellarine Peninsula); genotype B included ten clinical isolates, with SNPs varying between 1 and 396 (including one isolate from Darwin, one isolate from Port Hedland and eight isolates from Queensland); and genotype C included one clinical isolate from Papua New Guinea grouped with *M. ulcerans* CSURQ0185 isolated from Côte d'Ivoire with 383 SNPs (S2 Table). We found that the values of SNPs between the genotypes were as follows: between genotypes A and B >1,000, between genotypes B and C >2,000 and between genotypes A and C >2,000.

Hotspot-specific target. Core genome sequence alignment of the 218 genomes examined in this study (Fig 1b) indicated that one 1,068-bp *PPE* gene was annotated as a PPE family protein in the *M. ulcerans* strain CSURP7741 reference genome (GenBank assembly accession number GCA_901411635.1). The PPE protein was predicted to be an alpha helical transmembrane protein with 0.934205 probability. This gene exhibited interesting heterogeneities, including that it was completely absent in *M. ulcerans* SRR6346343 isolated from *T. vulpecula* in Gippsland. First, we observed deletions in the 5' extremity of the gene resulting in a *PPE* gene size varying between 335 bp and 1,068 bp, correlating with the clusterization of the isolates in the core-genome tree (Fig 1b and 1c). These PPE deletions resulted in a non-transmembrane protein prediction with 0.90177 probability. Second, we observed a heterogeneous sequence hotspot sorting 216 isolates into three groups (this hotspot region was absent in *M. ulcerans* strain Harvey (JAOL01000001.1) and in *M. ulcerans* (SRR6346343). Group X1 was characterized by no deletion variant *PPE* gene at 376-377-378-379-380-381 (AATTTT) sequence position in three *M. ulcerans* isolates; group X2 was characterized by a TT deletion at positions 378-379 in 177 *M. ulcerans* isolates, in all four *M. pseudoshotii* isolates, in the two *M. liflandii* isolates and in all 24 *M. marinum* genotypes; and group X3 was characterized by a TTT deletion at positions 378-379-380 in six *M. ulcerans* isolates (Fig 2, S1 Table). Focusing on the 173 *M. ulcerans* isolates from Australia, we found that all these isolates had a 525-bp group X2 *PPE* gene (Fig 1c).

***M. ulcerans* diversity, Australia.** To question the clonal diversity of the Australian isolates of *M. ulcerans*, we mapped the output MiSeq reads of 182 genomes (issued from 174 Australian isolates plus eight isolates sequenced in IHU Méditerranée Infection, France) to the hotspot region. We searched for the specific target deletion in the depth of reads and found additional deletion forms. In total, 13 *PPE* variants (here designed PPE-1 , PPE-13)

were detected among the 182 isolate genomes, including PPE-1 [AATTTT], PPE-2 [AA--TT], PPE-3 [AA-AAT], PPE-4 [AA-ATT], PPE-5 [A--CTT], PPE-6 [AA-TTT], PPE-7 [AAATTT], PPE-8 [AA---T], PPE-9 [--TATT], PPE-10 [AA--CT], PPE-11 [AA--AT], PPE-12 [AA--GT] and PPE-13 [AA-CTT], in addition to the three variants reported above (Fig 3). As an example, we extracted corresponding reads of *M. ulcerans* strain CSURP9950, mapped them against hotspots using Bowtie software,²³ and then we performed a reassembly. As a result, we obtained two different sequences: a majority sequence (TT deletion) and a minority sequence characterized by one T deletion followed by one A substitution; both sequences exhibited only 92% sequence identity with one another. BLASTn results of these two sequences showed that the majority sequence matched with *M. ulcerans* group X2 with 100% identity, while the minority sequence matched with *M. ulcerans* group X1 with 99.2% identity. Furthermore, we observed that the *M. ulcerans* strain CSURP7741 lacked heterogeneity in the hotspot (S1 Fig). Overall, deep analysis yielded ten additional variants in the *M. ulcerans* PPE gene, updating to 13 variants in this gene (Including the 03 variants found by assembly genome analysis plus the 10 variants found by deep sequence analysis). Focusing on the 174 *M. ulcerans* isolates from Australia, we observed geographical clusterization within 11 clusters (S2 Fig), and in Victoria, we had 695 variants for 157 clinical isolates as follows: in Mornington Peninsula, we had 305 variants for 66 clinical isolates; in Melbourne, we had 177 variants for 37 clinical isolates; in Bellarine Peninsula, figures were 149/36; in Phillip Island, 14/4; in Gippsland, 41/13; and in Ballarat 9/1. Outside Victoria, we had 32 variants for 11 isolates as follows: Queensland, 23/8; Papua New Guinea, 2/1; Port Hedland, 4/1 and Darwin, 3/1 (S4, S5 Tables). Analyzing these output data by comparing the average of detected variants between isolates belonging to each counties using the Wilcoxon test indicated a significantly (p-value = 0.001834) higher polyclonal diversity of *M. ulcerans* isolates from Victoria, Southeast Australia than from North Australia (S6

Table). Moreover, inside Victoria, the analysis indicated that *M. ulcerans* isolates collected from the Bellarine Peninsula, Melbourne and Mornington Peninsula exhibited a significantly higher polyclonal diversity than *M. ulcerans* isolates collected from the other two Victoria counties (Gippsland and Phillip Island), with p-value = 0.02062, p-value = 0.0006314 and p-value = 0.0009121, respectively (S6 Table). Finally, we observed a significantly higher incidence of [AA-TTT], [AAATTT] and [AA--AT] variants in the three Buruli ulcer epidemic counties, Bellarine Peninsula, Melbourne and Mornington Peninsula^{3,15} compared to the non-epidemic counties using the chi-square (χ^2) test with p-value=0.001367, p-value=0.0002324 and p-value=0.04721, respectively (S7 Table).

For animal isolates in Australia, we found 04 variants in 03 *T. vulpecula* isolates made from the epidemic counties Mornington Peninsula and the non-epidemic counties Gippsland and Phillip Island. Moreover, we found 03 variants in one *C. lupus familiaris* isolate made in one epidemic counties of the Bellarine Peninsula, 02 variants from one *P. longipes* isolate made in the non-epidemic counties of Gippsland and 03 variants in one “koala” isolate obtained from the non-epidemic counties of Gippsland (S2 Fig, S4 Table). We observed that the number of detected *PPE* gene variants in animal isolates was lower than that of human isolates from the same counties. Specifically, 4/13 *PPE* variants were detected in common in these four animal species (with 03 of them being endemic to Australia), whereas 9/13 *PPE* variants were found exclusively in *M. ulcerans* isolates of human origin (S4 Table). In summary, [AA--TT], [AA-ATT], [AA--CT] and [AA--GT] variants were found in common in animals and humans in both Buruli ulcer epidemic and non-epidemic counties in Australia (S4 Table).

Discussion.

In this study, reviewing *M. ulcerans* WGS data indicated that we and colleagues routinely deposited in public genome sequence databases only one straight genome sequence per *M. ulcerans* isolate, although such WGS data were derived from cultured colonies (i.e., millions of cultured mycobacteria), as single-cell sequencing has never been used in that context to the best of our knowledge.²⁴ To overcome this limitation, we analyzed the entire set of available sequencing reads, instead of allowing any software to automatically analyze majority reads and ignore minority reads.^{24,25}

While the resulting sequence heterogeneity may have resulted from contamination and sequencing bias, it may also carry biologically relevant information, as illustrated in this report.^{26,27} Following this original pathway of WGS analysis, we indeed observed a significantly higher genomic diversity among *M. ulcerans* isolates collected in Buruli ulcer endemic counties of South Australia compared to non-endemic ones. We were even able to refine the analysis to the epidemiologically relevant scale of Australian counties of the Mornington Peninsula, Melbourne, and Bellarine Peninsula, where a pejorative evolution of the infection has been recently reported.^{2,3,15} At this scale, we noticed that *M. ulcerans* clones tagged by specific *PPE* gene variants were detected in clinical samples in common with animal samples, three of which were found only in Australia.

The genomic data reported in this study indicate a roadmap for further studies aimed at translating genomic characteristics into microbiological characteristics of *M. ulcerans* isolates in Australia to precisely support the increased incidence and virulence of Buruli ulcer in some counties of southeastern Australia.

We propose generalizing the genome analysis described in this report to highlight minority clones generally obscured by majority clones in routine analyses, thereby helping to

248 depict the biological diversity of medically relevant microorganisms, including opportunistic
249 pathogens of environmental origin, such as *M. ulcerans*.

250

251

252

253

254 **Acknowledgments.**

255 We would like to thank the Department of Microbiology and Immunology, Doherty
 256 Institute for Infection and Immunity, University of Melbourne, Melbourne, Victoria,
 257 Australia and Doherty
 258 Applied Microbial Genomics, Department of Microbiology and Immunology, Doherty
 259 Institute for Infection
 260 and Immunity, University of Melbourne, Melbourne, Victoria, Australia, for agreeing to
 261 publish the assembly
 262 genome of 174 clinical *M. ulcerans* isolates. J.S., N.H., R.Z. and H.A. are supported by a
 263 Ph.D. grants from the IHU Méditerranée Infection, Marseille, France.

264 **Authors' addresses:** Jamal Saad, Nassim Hammoudi, Rita Zgheib, Hussein Anani, Michel
 265 Drancourt. Aix-Marseille-Univ., IRD, MEPHI, IHU Méditerranée Infection, Marseille,
 266 France, E-mails : jsaad270@gmail.com, nassimveto15@live.fr, zgheibrita@gmail.com,
 267 husseinanani2289@hotmail.com, michel.drancourt@univ-amu.fr.

268

Fig 1: a: Pangenome tree with profile genes of 218 studied strains; b: core-genome tree of 218 studied strains; c: *PPE* gene tree of 217 strains with the gene length for each isolate. The three groups X1, X2 and X3 of the studied strains are indicated. Different colors were used to show the isolation counties, and a star was used to mark animal isolates from another.

Fig 2: Tree based on the common region of the *PPE* gene of 216 studied strains with the different groups X1, X2 and X3 depending on the *PPE* gene variant (at positions 378, 379 and 380).

Fig 3: Map showing the distribution of the different *PPE* variants in Australia. The table shows the 13 *PPE* variants (see Text), the Venn diagram shows the *PPE* variants found in the endemic area in Victoria county (red), the non-endemic area in Victoria county (orange) and the non-endemic area outside Victoria county (green).

Supporting information.

S1 Fig. Heterogeneity profile at the level of depth reads of several isolates based on 125 bp of the *PPE* gene, including sequence study variants.

S2 Fig. Heatmap and hierarchical clustering tree showing the profile of 174 Australian isolates based on the presence or absence of the 13 detected variants and the isolation region of isolates.

S1 Table. Genomic and clinical information's about isolates study.

S2 Table. Pairwise SNP distance matrix values between 218 studied strains.

S3 Table. Pairwise ANI values between 218 studied trains

S4 Table. Distribution of the different *PPE* variants in the different Australia counties.

S5 Table. Distribution of the different *PPE* variants in each isolates study of Australia.

294 **S6 Table.** Statistical chi-square (χ^2) test of detected variants between Victoria counties and
 295 outside Victoria in Australia.

296 **S7 Table.** Statistical chi-square (χ^2) test of variants between Victoria counties and outside
 297 Victoria in Australia.

298

299

References.

1. MacCALLUM P, Tolhurst JC. A new mycobacterial infection in man. *J Pathol Bacteriol.* 1948;60(1):93-122.
2. Tai AYC, Athan E, Friedman ND, Hughes A, Walton A, O'Brien DP. Increased Severity and Spread of Mycobacterium ulcerans, Southeastern Australia. *Emerg Infect Dis.* 2018;24(1). doi:[10.3201/eid2401.171070](https://doi.org/10.3201/eid2401.171070)
3. Loftus MJ, Tay EL, Globan M, et al. Epidemiology of Buruli Ulcer Infections, Victoria, Australia, 2011-2016. *Emerg Infect Dis.* 2018;24(11):1988-1997. doi:[10.3201/eid2411.171593](https://doi.org/10.3201/eid2411.171593)
4. Drancourt M, Zingue D. Variations in temperatures and the incidence of Buruli ulcer in Africa. *Travel Med Infect Dis.* 2020;36:101472. doi:[10.1016/j.tmaid.2019.101472](https://doi.org/10.1016/j.tmaid.2019.101472)
5. O'Brien DP, Jeanne I, Blasdel K, Avumegah M, Athan E. The changing epidemiology worldwide of Mycobacterium ulcerans. *Epidemiol Infect.* Published online October 8, 2018:1-8. doi:[10.1017/S0950268818002662](https://doi.org/10.1017/S0950268818002662)
6. Lauer T. SNIDSRegionalComparisons. Published online 2020:11.
7. Hammoudi N, Cassagne C, Million M, et al. Investigation of skin microbiota reveals Mycobacterium ulcerans-Aspergillus sp. trans-kingdom communication. *bioRxiv.* Published online December 10, 2019:869636. doi:[10.1101/869636](https://doi.org/10.1101/869636)
8. Stinear TP, Mve-Obiang A, Small PLC, et al. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of Mycobacterium ulcerans. *Proc Natl Acad Sci U S A.* 2004;101(5):1345-1349. doi:[10.1073/pnas.0305877101](https://doi.org/10.1073/pnas.0305877101)
9. Pidot SJ, Hong H, Seemann T, et al. Deciphering the genetic basis for polyketide variation among mycobacteria producing mycolactones. *BMC Genomics.* 2008;9:462. doi:[10.1186/1471-2164-9-462](https://doi.org/10.1186/1471-2164-9-462)

- 324 10. Ross BC, Johnson PD, Oppedisano F, et al. Detection of *Mycobacterium ulcerans* in
325 environmental samples during an outbreak of ulcerative disease. *Appl Environ Microbiol.*
326 1997;63(10):4135-4138. doi:[10.1128/AEM.63.10.4135-4138.1997](https://doi.org/10.1128/AEM.63.10.4135-4138.1997)
- 327 11. Bratschi MW, Ruf M-T, Andreoli A, et al. *Mycobacterium ulcerans* persistence at a
328 village water source of Buruli ulcer patients. *PLoS Negl Trop Dis.* 2014;8(3):e2756.
329 doi:[10.1371/journal.pntd.0002756](https://doi.org/10.1371/journal.pntd.0002756)
- 330 12. Portaels F, Meyers WM, Ablordey A, et al. First cultivation and characterization of
331 *Mycobacterium ulcerans* from the environment. *PLoS Negl Trop Dis.* 2008;2(3):e178.
332 doi:[10.1371/journal.pntd.0000178](https://doi.org/10.1371/journal.pntd.0000178)
- 333 13. Zingue D, Panda A, Drancourt M. A protocol for culturing environmental strains of the
334 Buruli ulcer agent, *Mycobacterium ulcerans*. *Sci Rep.* 2018;8(1):6778. doi:[10.1038/s41598-](https://doi.org/10.1038/s41598-018-25278-y)
335 [018-25278-y](https://doi.org/10.1038/s41598-018-25278-y)
- 336 14. Vandellannoote K, Phanzu DM, Kibadi K, et al. *Mycobacterium ulcerans* Population
337 Genomics To Inform on the Spread of Buruli Ulcer across Central Africa. *mSphere.*
338 2019;4(1). doi:[10.1128/mSphere.00472-18](https://doi.org/10.1128/mSphere.00472-18)
- 339 15. Buultjens AH, Vandellannoote K, Meehan CJ, et al. Comparative Genomics Shows That
340 *Mycobacterium ulcerans* Migration and Expansion Preceded the Rise of Buruli Ulcer in
341 Southeastern Australia. *Appl Environ Microbiol.* 2018;84(8). doi:[10.1128/AEM.02612-17](https://doi.org/10.1128/AEM.02612-17)
- 342 16. Saad J, Combe M, Hammoudi N, et al. Whole-Genome Sequence of *Mycobacterium*
343 *ulcerans* CSURP7741, a French Guianan Clinical Isolate. *Microbiol Resour Announc.*
344 2019;8(29). doi:[10.1128/MRA.00215-19](https://doi.org/10.1128/MRA.00215-19)
- 345 17. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and
346 its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455-477.
347 doi:[10.1089/cmb.2012.0021](https://doi.org/10.1089/cmb.2012.0021)

- 348 18. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*.
349 2014;30(14):2068-2069. doi:[10.1093/bioinformatics/btu153](https://doi.org/10.1093/bioinformatics/btu153)
- 350 19. Page AJ, Cummins CA, Hunt M, et al. Roary: rapid large-scale prokaryote pan genome
351 analysis. *Bioinformatics*. 2015;31(22):3691-3693. doi:[10.1093/bioinformatics/btv421](https://doi.org/10.1093/bioinformatics/btv421)
- 352 20. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for
353 genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075.
354 doi:[10.1093/bioinformatics/btt086](https://doi.org/10.1093/bioinformatics/btt086)
- 355 21. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to
356 calculate average nucleotide identity. *Antonie Van Leeuwenhoek*. 2017;110(10):1281-1286.
357 doi:[10.1007/s10482-017-0844-4](https://doi.org/10.1007/s10482-017-0844-4)
- 358 22. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user
359 interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*.
360 2010;27(2):221-224. doi:[10.1093/molbev/msp259](https://doi.org/10.1093/molbev/msp259)
- 361 23. Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics*.
362 2010;Chapter 11:Unit 11.7. doi:[10.1002/0471250953.bi1107s32](https://doi.org/10.1002/0471250953.bi1107s32)
- 363 24. Black PA, de Vos M, Louw GE, et al. Whole genome sequencing reveals genomic
364 heterogeneity and antibiotic purification in *Mycobacterium tuberculosis* isolates. *BMC*
365 *Genomics*. 2015;16:857. doi:[10.1186/s12864-015-2067-2](https://doi.org/10.1186/s12864-015-2067-2)
- 366 25. Lerch A, Koepfli C, Hofmann NE, et al. Development of amplicon deep sequencing
367 markers and data analysis pipeline for genotyping multi-clonal malaria infections. *BMC*
368 *Genomics*. 2017;18(1):864. doi:[10.1186/s12864-017-4260-y](https://doi.org/10.1186/s12864-017-4260-y)
- 369 26. Advani J, Verma R, Chatterjee O, et al. Whole Genome Sequencing of *Mycobacterium*
370 *tuberculosis* Clinical Isolates From India Reveals Genetic Heterogeneity and Region-Specific
371 Variations That Might Affect Drug Susceptibility. *Front Microbiol*. 2019;10:309.
372 doi:[10.3389/fmicb.2019.00309](https://doi.org/10.3389/fmicb.2019.00309)

- 373 27. Sobkowiak B, Glynn JR, Houben RMGJ, et al. Identifying mixed Mycobacterium
 374 tuberculosis infections from whole genome sequence data. *BMC Genomics*. 2018;19(1):613.
 375 doi:[10.1186/s12864-018-4988-z](https://doi.org/10.1186/s12864-018-4988-z)

a

b

c

		Democratic Republic of Congo
		Europe
		France
nsula		French Guiana
		Ghana
		Israel
		Japan
eninsula		Papua New Guinea
		Thailand
		Uganda
		UK
		USA
		USA: Philadelphia
		Unknown

Nº	PPE variants
1	[AATTTT]
2	[AA--TT]
3	[AA-AAAT]
4	[AA-ATT]
5	[A--CTT]
6	[AA-TTT]
7	[AAATTT]
8	[AA---T]
9	[--TATT]
10	[AA--CT]
11	[AA--AT]
12	[AA--GT]
13	[AA-CTT]