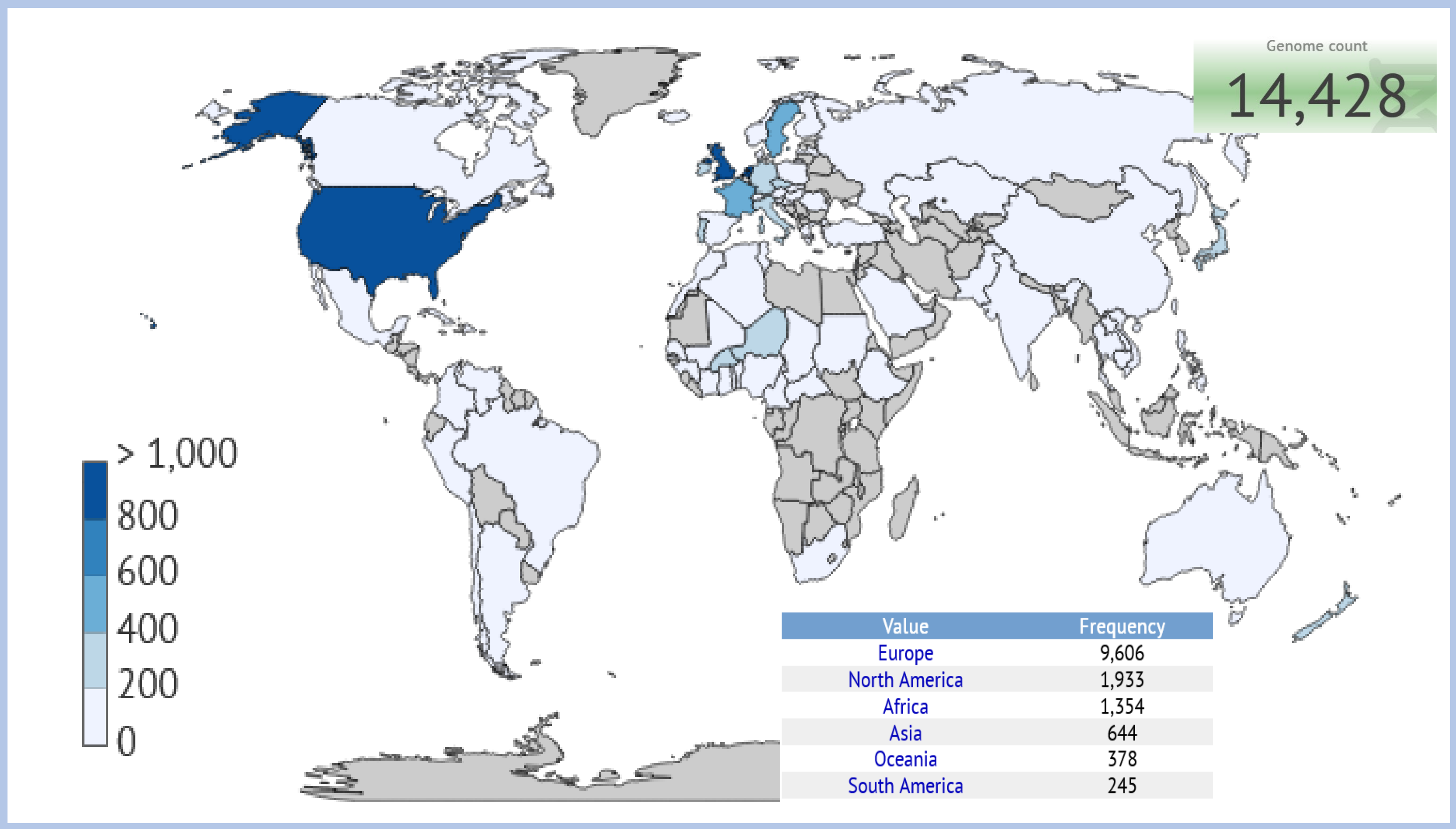
Global Meningitis Genome Library: The Power of Curation

Holly Bratcher[1], **Kasia Parfitt**[1*], Keith Jolley[1], Martin Maiden[1*]

[1]Department of Biology, University of Oxford, United Kingdom
*Corresponding authors: kasia.parfitt@biology.ox.ac.uk and martin.maiden@biology.ox.ac.uk

Meningitis remains a major public health challenge globally. The goal of the Global Meningitis Genome Library (GMGL)[1] within PubMLST[2] is to co-ordinate and host publicly available genome libraries for curated sets of meningitis-causing bacterial pathogens. It was built to address the WHO's roadmap to combatting meningitis. A curated library is defined as a coherent set of isolates which preferably include provenance records, annotated genome sequence data, and are linked to published literature.

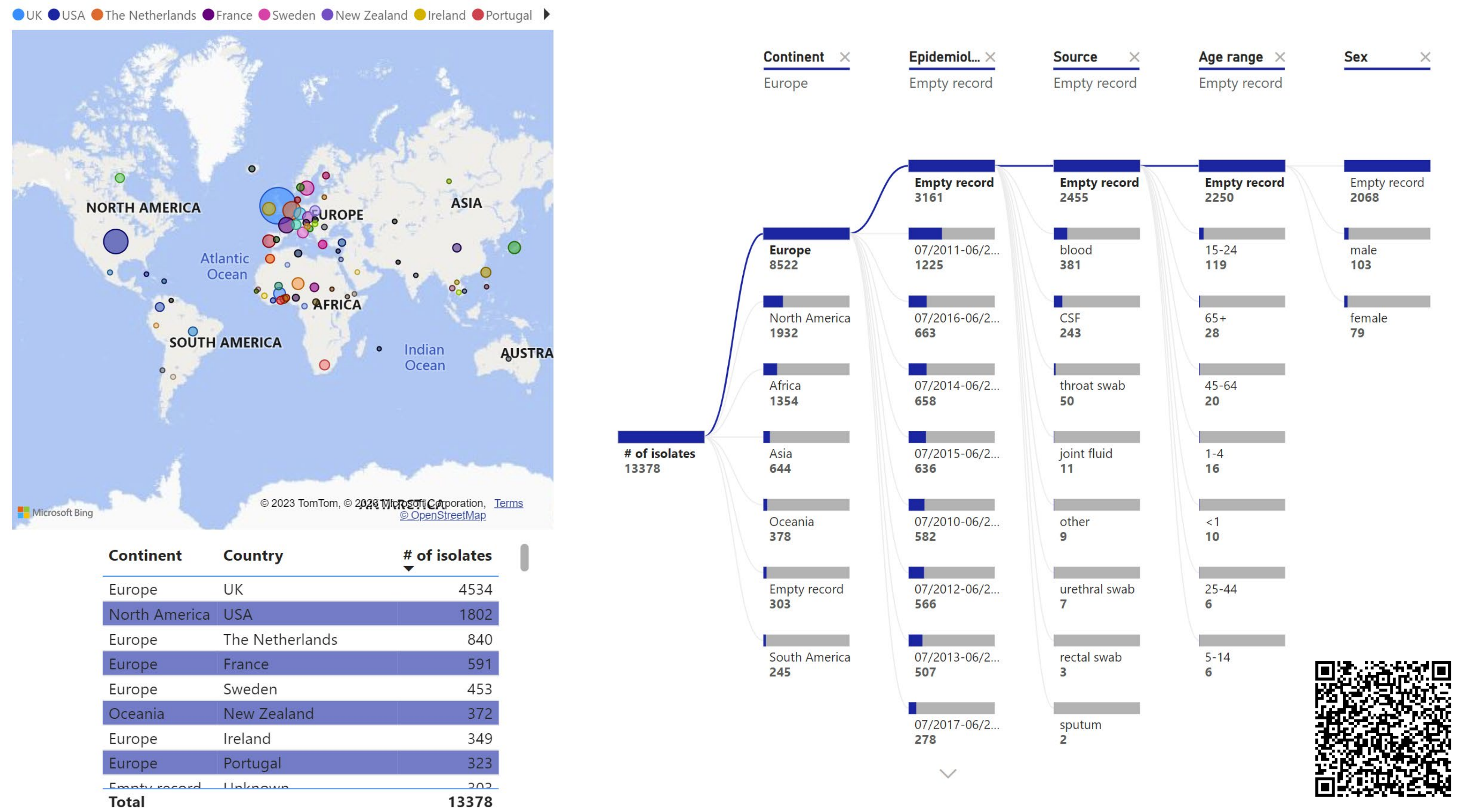
1. CURRENT DATABASE ISOLATES



Value	Frequency
Europe	9,606
North America	1,933
Africa	1,354
Asia	644
Oceania	378
South America	245

Figure 1 | World map illustrating the number of current isolates in the GMGL. The colour code on the left-hand side represents the total number of isolates present in each country. There are a total of 14,428 genomes in the database, spanning 6 continents. Europe has the highest number of isolates (*n* =9,606).

2. PROVENANCE & MISSING DATA



Continent	Country	# of isolates
Europe	UK	4534
North America	USA	1802
Europe	The Netherlands	840
Europe	France	591
Europe	Sweden	453
Oceania	New Zealand	372
Europe	Ireland	349
Europe	Portugal	323
Empty record	Unknown	303
Total		**13378**

Figure 2 | Screenshot from a dashboard (created by Margaret Varga, University of Oxford) which illustrates the missing provenance data within the GMGL database. The map shows the number of isolates in the database for each country (the bigger the circle, the greater the number of isolates). The right-hand side shows an example of a flow diagram which displays the amount of provenance data available for each country.

3. DATA ENTRY & REUSE

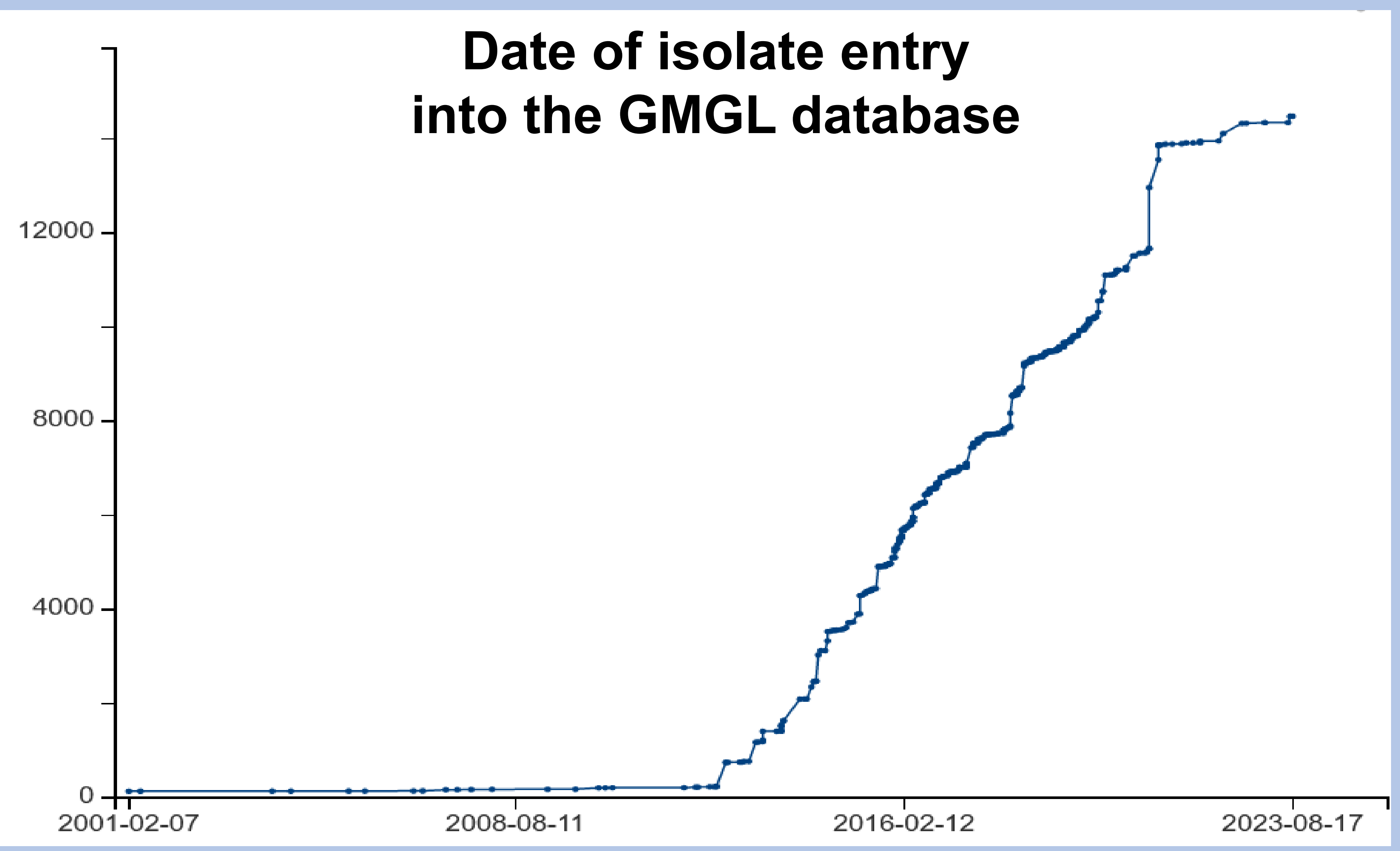


Figure 3 | Dates the *N. meningitidis* isolates were entered into the GMGL database (2001-2023). The graph is cumulative so illustrate the total number in the database at each time point.

4. STUDIES

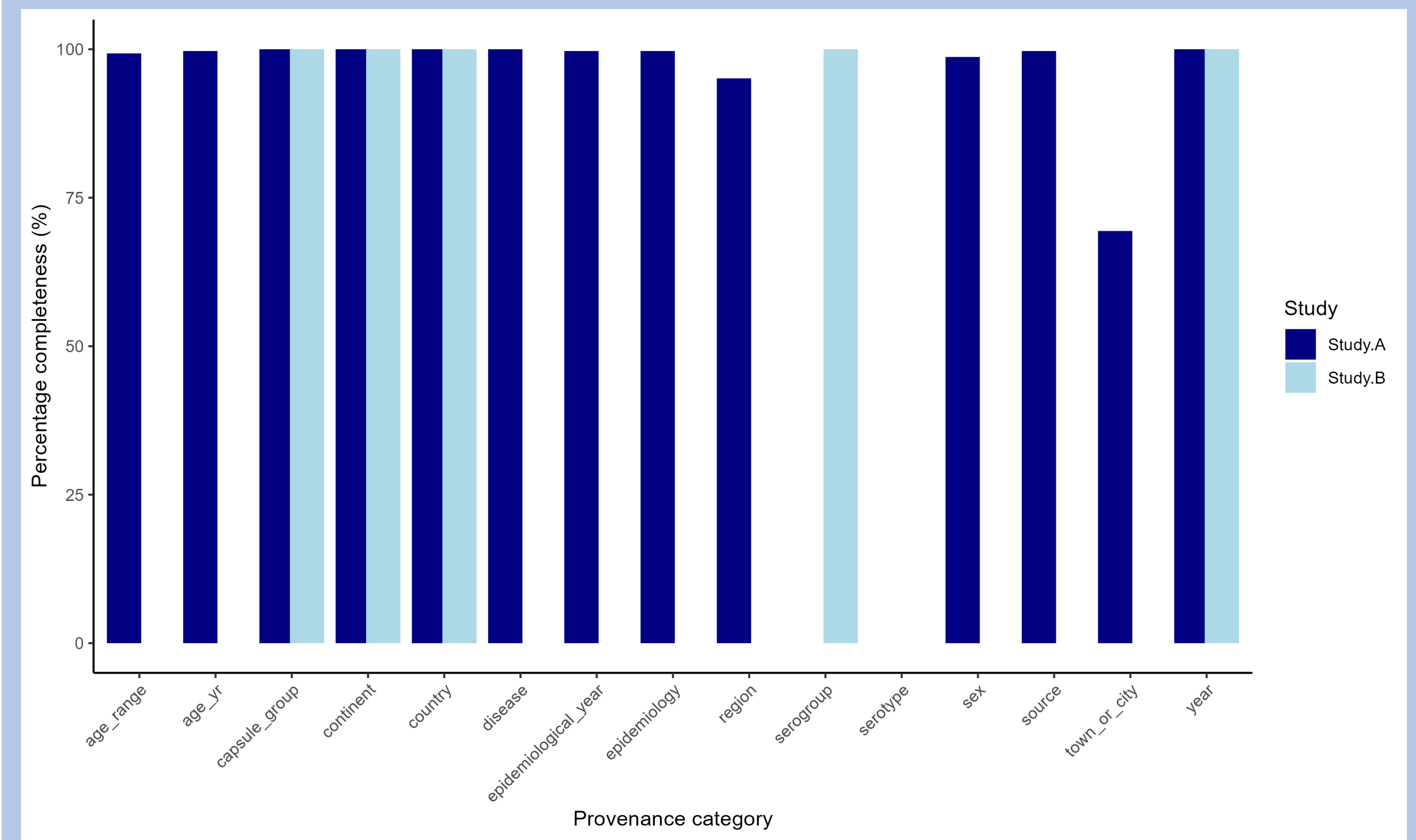


Figure 4 | Percentage completeness (%) of two anonymised datasets. Study A represents a more complete set of isolates with provenance data. Study B represents an isolate dataset which is less complete in terms of provenance data.

5. FUTURE DIRECTIONS

To improve the genome-associated metadata in the GMGL through i) gathering information from the published literature and ii) confirming/contacting publication authors.

6. REFERENCES

[1]Rodgers E, Bentley SD, Borrow R, *et al*. The global meningitis genome partnership. J Infect. 2020 Oct;81(4):510-520. doi: 10.1016/j.jinf.2020.06.064. Epub 2020 Jun 29. PMID: 32615197.

[2]Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018 Sep 24;3:124. doi: 10.12688/wellcomeopenres.14826.1. PMID: 30345391; PMCID: PMC6192448.

[3]Brehony C, Rodrigues CMC, Borrow R, *et al*. Distribution of Bexsero® Antigen Sequence Types (BASTs) in invasive meningococcal disease isolates: Implications for immunisation. Vaccine. 2016 Sep 7;34(39):4690-4697. doi: 10.1016/j.vaccine.2016.08.015. Epub 2016 Aug 9. PMID: 27521232; PMCID: PMC5012890.