



Use of a whole genome sequencing-based approach for *Mycobacterium tuberculosis* surveillance in Europe in 2017–2019: an ECDC pilot study

Elisa Tagliani ¹, Richard Anthony^{2,11}, Thomas A. Kohl ^{3,4,11}, Albert de Neeling², Vlad Nikolayevskyy ^{5,6}, Csaba Ködmön⁷, Florian P. Maurer ^{8,9}, Stefan Niemann^{3,4}, Dick van Soolingen², Marieke J. van der Werf ⁷ and Daniela Maria Cirillo ¹ on behalf of the ECDC molecular surveillance project participants¹⁰

@ERSpublications

The implementation of a WGS-based surveillance system for monitoring the emergence of MDR-TB outbreaks in Europe is feasible and has the potential to provide supporting evidence to better elucidate cross-border transmission patterns <https://bit.ly/2ZTnPjk>

Cite this article as: Tagliani E, Anthony R, Kohl TA, *et al.* Use of a whole genome sequencing-based approach for *Mycobacterium tuberculosis* surveillance in Europe in 2017–2019: an ECDC pilot study. *Eur Respir J* 2021; 57: 2002272 [<https://doi.org/10.1183/13993003.02272-2020>].

ABSTRACT Whole genome sequencing (WGS) can be used for molecular typing and characterisation of *Mycobacterium tuberculosis* complex (MTBC) strains. We evaluated the systematic use of a WGS-based approach for MTBC surveillance involving all European Union/European Economic Area (EU/EEA) countries and highlight the challenges and lessons learnt to be considered for the future development of a WGS-based surveillance system.

WGS and epidemiological data of patients with rifampicin-resistant (RR) and multidrug-resistant (MDR) tuberculosis (TB) were collected from EU/EEA countries between January 2017 and December 2019. WGS-based genetic relatedness analysis was performed using a standardised approach including both core genome multilocus sequence typing (cgMLST) and single nucleotide polymorphism (SNP)-based calculation of distances on all WGS data that fulfilled minimum quality criteria to ensure data comparability.

A total of 2218 RR/MDR-MTBC isolates were collected from 25 countries. Among these, 56 cross-border clusters with increased likelihood of recent transmission (≤ 5 SNPs distance) comprising 316 RR/MDR-MTBC isolates were identified. The cross-border clusters included between two and 30 resistant isolates from two to six countries, demonstrating different RR/MDR-TB transmission patterns in Western and Eastern EU countries.

This pilot study shows that a WGS-based surveillance system is not only feasible but can efficiently elucidate the dynamics of in-country and cross-border RR/MDR-TB transmission across EU/EEA countries. Lessons learnt from this study highlight that the establishment of an EU/EEA centralised WGS-based surveillance system for TB will require strengthening of national integrated systems performing prospective WGS surveillance and the development of clear procedures to facilitate international collaboration for the investigation of cross-border clusters.

This article has supplementary material available from erj.ersjournals.com

Received: 11 June 2020 | Accepted after revision: 17 July 2020

Copyright ©ERS 2021. This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0.

Introduction

Multidrug-resistant (MDR) and extensively drug-resistant (XDR) tuberculosis (TB) represent a major threat to global TB control. With a total of 999 MDR-TB cases reported in 2018, there is a low overall prevalence of MDR-TB in the European Union (EU)/European Economic Area (EEA) (3.7% of all reported TB cases), but the proportion of MDR-TB remains high in countries like Lithuania (17.1%) and Estonia (24.6%) [1]. Romania alone, with 354 MDR-TB cases notified in 2018, reported more than one third (35.4%) of the total number of MDR-TB cases in the entire EU/EEA [1].

M/XDR-TB is difficult and expensive to treat; in 2018, the treatment success rate among MDR-TB cases in the EU/EEA was just 49.9% [1]. For the development of optimal TB control and infection prevention strategies, accurate tracing of in-country and cross-border MDR-TB transmission and rapid identification of emerging M/XDR-TB clones is of critical importance. The introduction of the Mycobacterial Interspersed Repetitive Units Variable Number Tandem Repeats (MIRU-VNTR) methodology for genotyping of *Mycobacterium tuberculosis* complex (MTBC) in EU/EEA Member States has enhanced our knowledge of the phylogeny and transmission patterns of clinical MTBC strains. However, the use of MIRU-VNTR for surveillance of MDR-TB in the EU/EEA has been hampered by suboptimal coverage (26.2% in 2015), delays in reporting and lack of epidemiological data to confirm chains of inter-human transmission [2, 3]. In addition, the intrinsic lack of discriminatory power of MIRU-VNTR makes this technology suboptimal for supporting contact tracing [4].

Recently, whole genome sequencing (WGS) has become the gold standard for classification of MTBC strains and for tracing infection sources and transmission networks [5]. WGS can also be used for genotypic prediction of the antimicrobial susceptibility profile of a given isolate with sufficient accuracy to guide proper treatment [6–8]. The implementation of TB WGS in reference laboratories has shown that this technology can be a cost-effective and timesaving alternative to conventional MDR diagnostic and typing methods [9]. The role of WGS in detecting TB transmission and for investigating M/XDR-TB outbreaks in Europe has recently been demonstrated [10–13] and extensively reviewed [14–16]. Additional work on the effective translation of genomic data into meaningful information to guide informed decisions regarding public health interventions for tracing outbreaks and surveillance is still needed. More specifically, standardised approaches for data analysis to ensure compatibility of WGS data, quality assessment schemes, standardised nomenclature and validation of previously defined criteria for identifying recent transmission are needed [5].

Since 2015, the European Centre for Disease Prevention and Control (ECDC) has been supporting a number of pilot studies and projects for the application of WGS-based typing for public health protection [17]. In the framework of a pilot study on the use of WGS for molecular typing of MTBC strains in the EU/EEA, we evaluated the systematic use of a WGS-based approach for MDR-TB surveillance involving all EU/EEA countries [18]. In the project, we aimed to identify the major cross-border clusters and to provide genetic characterisation related to lineage and drug-resistance profiles of the rifampicin-resistant (RR)/MDR-MTBC strains circulating in the EU/EEA. We also identified critical areas to be addressed for the future development of a WGS-based surveillance system in Europe.

Methods

The pilot study was implemented according to pre-analytical, analytical and post-analytical parameters and methods agreed upon during a technical expert consultation convened at the start of the study (supplementary table S1).

Study participants

All EU/EEA countries were invited to participate in this study.

Affiliations: ¹Emerging Bacterial Pathogens Unit, Division of Immunology, Transplantation and Infectious Diseases, IRCCS San Raffaele Scientific Institute, Milan, Italy. ²Tuberculosis Reference Laboratory, Infectious Diseases Research, Diagnostics and Laboratory Surveillance (IDS), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands. ³Molecular and Experimental Mycobacteriology, National Reference Center for Mycobacteria, Research Center, Borstel, Germany. ⁴German Center for Infection Research, partner site Borstel-Hamburg-Lübeck-Riems, Borstel, Germany. ⁵Public Health England, London, UK. ⁶Imperial College London, London, UK. ⁷European Centre for Disease Prevention and Control, Stockholm, Sweden. ⁸Diagnostic Mycobacteriology, National Reference Center for Mycobacteria, Borstel, Germany. ⁹Institute of Medical Microbiology, Virology and Hygiene, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. ¹⁰See Acknowledgements for full details. ¹¹Both authors contributed equally.

Correspondence: Elisa Tagliani, Emerging Bacterial Pathogens Unit, Division of Immunology, Transplantation and Infectious Diseases, IRCCS San Raffaele Scientific Institute, Via Olgettina 58, 20132 Milan, Italy. E-mail: tagliani.elisa@dhsr.it

Data collection and validation

WGS and epidemiological data of patients with RR/MDR-TB were collected by the EUSeqMyTB consortium [18] from January 2017 to December 2019 (supplementary note S1). Local laboratory codes were pseudo-anonymised using a unique EU Sequencing Typing sample identifier to comply with EU Regulation 2016/679 on General Data Protection Regulation (GDPR). WGS data underwent quality checking and only those fulfilling pre-defined quality criteria were included in the study (supplementary note S1).

WGS-based relatedness analysis

The MTBC relatedness analysis was performed using two sequential analytical approaches: 1) core genome multilocus sequence typing (cgMLST) and 2) single nucleotide polymorphism (SNP)-based calculation of distances (supplementary note S2). The initial cgMLST-based analysis [19] allowed for “real-time” analysis of the rapidly expanding surveillance study database using a substantially reduced computational capacity compared to the SNP-based approach. The SNP-based analysis, using the MTBseq pipeline [20], was performed on the dataset of MTBC isolates clustering by cgMLST using a maximum distance threshold of ≤ 5 SNPs/alleles to identify cross-border clusters with increased likelihood of recent transmission [21].

WGS-based drug-resistance prediction and phylogenetic classification

Screening for drug-resistance variants and phylogenetic SNPs was performed using the MTBseq pipeline (supplementary note S2) [20]. Detected variants were annotated with drug resistance or susceptibility and phylogenetic SNPs according to the literature [22–27].

Cross-border cluster identification

An SNP-based cross-border cluster was defined as two or more RR/MDR-MTBC isolates having a difference of ≤ 5 SNPs, isolated in at least two different countries.

Data reporting to study participants

Study participants could assess the results of the SNP-based analysis through an access-controlled external webserver (www.euseqmytb.eu/) (supplementary note S3).

Coverage

The coverage by country was calculated using the number of MDR-TB cases notified to ECDC in 2018 as a reference when available [1]. For Latvia we used the number of MDR-TB cases notified to ECDC in 2017 [28], and for Italy the number of MDR-TB cases provided by the Italian National Institute of Health in 2018.

Results**Coverage of MDR-TB burden in the EU/EEA**

A total of 28 countries participated in the study by signing the material and data transfer agreement through their National TB Reference Laboratories, including Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden and the UK. The study included 2218 MTBC isolates from 2172 patients, isolated in 25 of the 28 participating EU/EEA Member States (89%) (table 1). The coverage was 76.7% of all reported MDR-TB cases in the region in 2018 (supplementary table S2). Only Lithuania and Italy had a coverage <70% (figure 1).

Genetic characterisation of the submitted isolates

A total of 2217 (99.95%) MTBC sequenced isolates fulfilled the minimum quality criteria. Among the RR/MDR-MTBC strains, the most represented lineages were lineage 4 (Euro-American; n=1404, 65.3%) and lineage 2.2.1 (Beijing; n=636, 29.7%), followed by lineage 3 (Delhi-CAS; n=63, 2.9%) and lineage 1 (East-African Indian (EAI); n=32, 1.5%) (figure 2 and supplementary note S4). Only two (0.1%) strains belonging to lineage 2.1 (East-Asian non-Beijing) and one (0.05%) lineage 5 (West-Africa 1) strain were detected.

WGS revealed mutations predicting resistance to rifampicin in 2151 (97.0%) isolates; among those, 1962 isolates (91.2%) had additional resistance to isoniazid (*i.e.* MDR-TB cases) (table 2). Among the RR/MDR-TB cases, resistance to fluoroquinolones was predicted in 581 (27%) isolates and additional resistance to any of the second-line injectable agents (*i.e.* XDR-TB) in 331 (15.4%) isolates [22, 23]. A total of 26 (1.2%) RR/MDR-MTBC isolates carried mutations predicting resistance to bedaquiline [26]. Additional

TABLE 1 *Mycobacterium tuberculosis* isolates submitted to the EUSeqMyTB consortium by year of isolation

Country	Isolates submitted by year of isolation n			Total n
	2017	2018	2019	
Austria	0	19	6	25
Belgium	5	8	6	19
Bulgaria	25	24	16	65
Croatia	0	2	1	3
Cyprus	0	0	0	0
Czech Republic	5	9	11	25
Denmark	0	4	4	8
Estonia	28	25	13	66
Finland	4	4	2	10
France	8	66	25	99
Germany	77	124	45	246
Hungary	5	13	4	22
Ireland	6	7	10	23
Italy	63	44	20	127
Latvia	18	33	34	85
Lithuania	77	60	42	179
Luxembourg	0	0	0	0
Malta	0	0	0	0
Netherlands	12	6	6	24
Norway	2	6	2	10
Poland	31	57	10	98
Portugal	12	19	6	37
Romania	335	336	195	866
Slovakia	2	4	2	8
Slovenia	0	1	0	1
Spain	17	37	12	66
Sweden	9	11	5	25
UK	27	30	24	81
Total per year	768	949	501	2218

details on the drug-resistance profile of the collected isolates and on the genomic regions considered to be associated with drug resistance are provided in supplementary note S5 and table S3, respectively.

Cross-border cluster identification

The cgMLST analysis, based on a ≤ 5 alleles difference, grouped 1145 (51.6%) isolates into 307 clusters, ranging in size from two to 36 isolates, and including 244 national and 63 cross-border clusters. The 63 cross-border clusters comprised 449 isolates and ranged in size from two to 36 isolates, with a geographic distribution ranging from two to eight countries (supplementary table S4).

The SNP-based analysis, performed on the pool of MTBC isolates clustering by cgMLST using a ≤ 5 SNPs threshold, resulted in a reduction of the total number of clustered cases to 1017, the number of isolates in cross-border clusters to 316, and the total number of cross-border clusters to 56. This analysis broke a few cgMLST-based clusters (*i.e.* cgCL1, cgCL2 and cgCL4) into multiple SNP-based cross-border and national clusters, in line with the expected slightly higher sensitivity of the SNP-based approach. Cross-border clusters, identified by a unique SNP-cluster number (snpCL) (table 3), comprised between two and 30 isolates, with a geographic distribution ranging from two to six countries. Only three countries, Croatia, Portugal and Slovenia, were not involved in any cross-border cluster.

All subsequent analyses were performed on the pool of 316 RR/MDR-MTBC isolates included in cross-border clusters by SNP-based analysis.

Characteristics of RR/MDR-TB patients included in cross-border clusters

Sex did not differ between clustered and non-clustered RR/MDR-TB cases. Among the clustered RR/MDR-TB cases, 70.3% (95% CI 64.5%–75.0%) were male *versus* 71.6% (95% CI 69.5%–73.6%) of non-clustered ones. The mean age was 44 years (range 1–92 years), 41 years in clustered and 44 years in non-clustered cases.

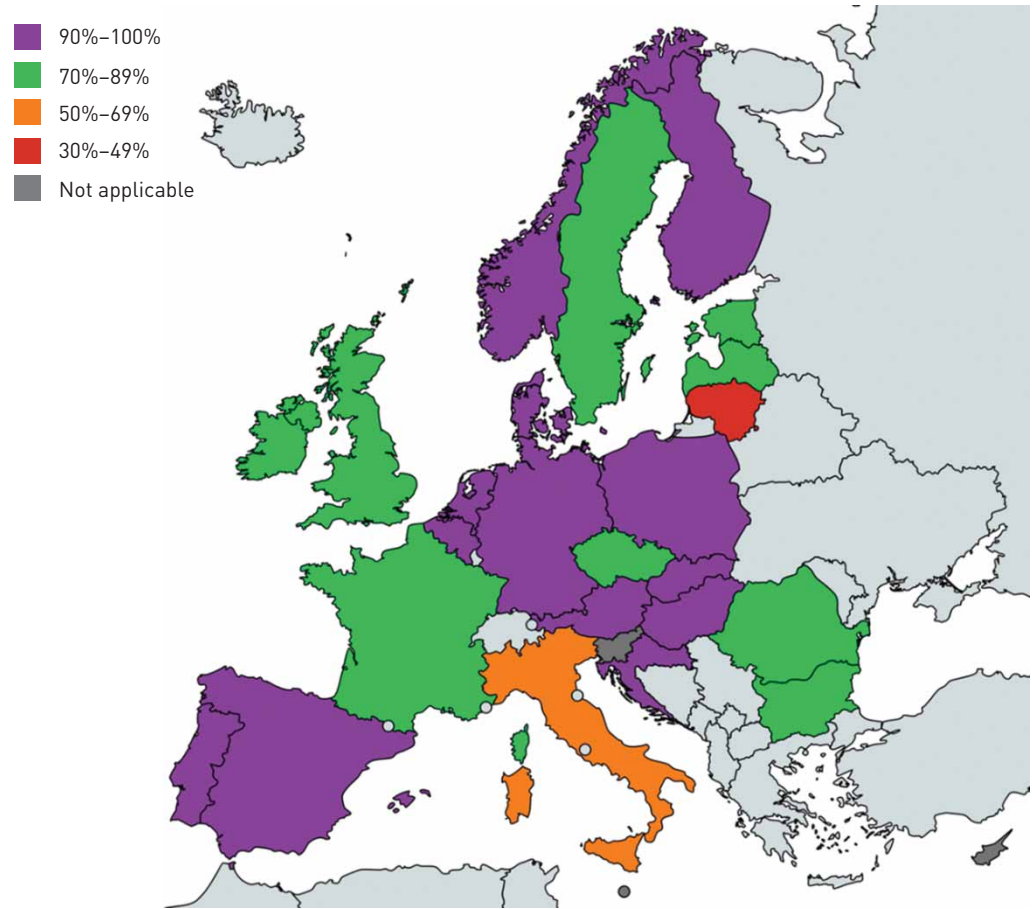


FIGURE 1 Percentage of multidrug-resistant (MDR)-tuberculosis (TB) coverage per European Union/European Economic Area country in 2018. The number of MDR-TB cases among all laboratory-confirmed TB cases notified to the European Centre for Disease Prevention and Control in 2018 was used as reference. The map was generated using the online tool available at www.mapchart.net/.

The majority ($n=242$, 76.6%) of the clustered RR/MDR-MTBC isolates belonged to lineage 4 (Euro-American), followed by lineage 2.2.1 (Beijing; $n=72$, 22.8%) and lineage 3 (Delhi-CAS; $n=2$, 0.6%) (supplementary figure S1). In lineage 4, the most frequently represented sub-lineages were 4.7–4.8 (mainly T; $n=91$, 28.8%), 4.2.1 (Ural; $n=37$, 11.7%), 4.3 (LAM; $n=30$, 9.5%) and 4.1.2.1 (Haarlem; $n=29$, 9.2%).

Four countries submitted ~70% of the clustered RR/MDR-MTBC isolates: Romania, Lithuania, Germany and Italy (table 4). Italy was also among the countries with the highest percentage of clustered RR/MDR-MTBC isolates among the total number of submitted isolates ($n=38$, 29.9%), together with Austria ($n=9$, 36.0%), Sweden ($n=8$, 32.0%) and Belgium ($n=5$, 26.3%).

Country of birth was reported for 198 of 316 (62.7%) clustered RR/MDR-TB cases. The percentage of clustered cases of foreign origin was between 80% and 100% in seven countries (Belgium, Finland, France, Ireland, Norway, Sweden and the UK), between 50% and 70% in four countries (Hungary, Italy, Slovakia and Spain), between 20% and 40% in two countries (Estonia and Poland) and 0% in three countries (Bulgaria, Latvia and Romania) (figure 3). Of the 56 cross-border clusters identified, 34 (60.7%) clusters were of “EU origin” and comprised 212 (67.1%) RR/MDR-TB patients originating from EU countries (origin missing for 60 cases). Among those, the majority (21 of 34) primarily included cases originating from Romania and Lithuania, the two EU countries with the highest MDR-TB burden, as well as a few cases from Italy and Poland. The remaining 13 clusters within this group included cases from Bulgaria, Estonia, Poland, Italy and Hungary. We also identified 15 (26.8%) cross-border clusters, including 81 (25.6%) RR/MDR-TB cases, almost exclusively of foreign (non-EU countries) origin (“non-EU origin” cross-border clusters). In this group, information on origin was missing for 40 (49.4%) of the cases. Only two of 15 clusters included individuals born in EU and non-EU countries. From the age of the two EU-born cases (one infant and one toddler), we suspect that transmission occurred within the family. Of the 15 “non-EU origin” cross-border clusters (61 RR/MDR-TB cases), 11 exclusively included TB patients

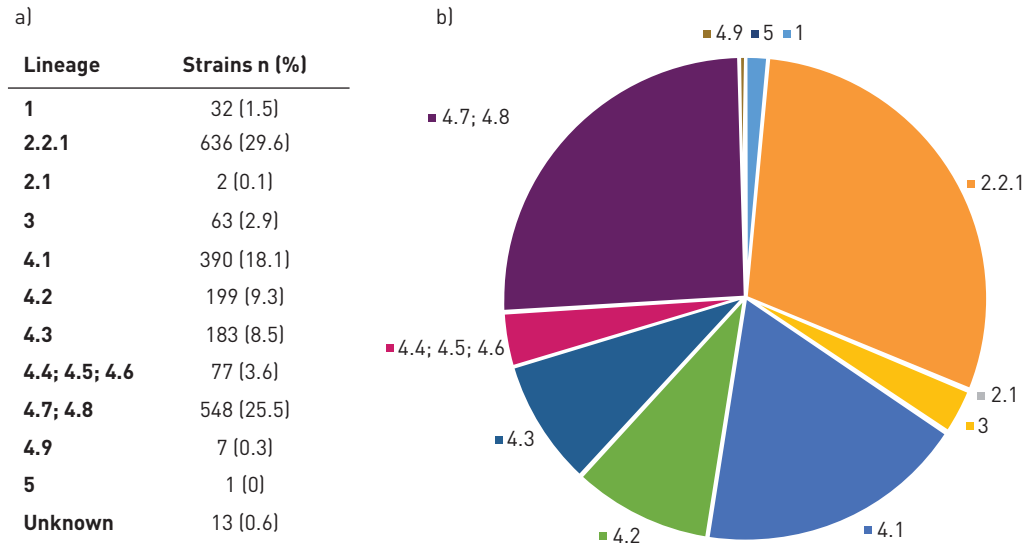


FIGURE 2 a, b) Lineage distribution of the 2151 rifampicin-resistant (RR)/multidrug-resistant (MDR) *Mycobacterium tuberculosis* complex (MTBC) isolates included in the EUSeqMyTB study. Lineages: 1: East-African Indian [EAI] (includes EAI and EAI Manila); 2: East-Asian (includes 2.2.1 Beijing and 2.1 East-Asian non-Beijing); 3: Delhi-CAS; 4.1: Euro-American (includes 4.1 and 4.1.2: Euro-American; 4.1.1: X-type; and 4.1.2.1: Haarlem); 4.2: Euro-American (includes 4.2 and 4.2.2: Euro-American; 4.2.1: Ural; and 4.2.2.1: TUR); 4.3: LAM; 4.4 (includes 4.4.1.1: S-type); 4.5: Euro-American; 4.6: Euro-American; 4.7: mainly T; 4.8: mainly T; 4.9: H37Rv-like; 5: West-Africa 1.

from Africa/sub-Saharan Africa, and four included patients from the Eastern European Region countries (Ukraine and Georgia). For the remaining seven (12.5%) cross-border clusters, comprising 23 (7.3%) RR/MDR-TB cases, information on origin was missing for 18 (78.3%) cases and therefore it was not possible to classify them into one specific group.

Examples of major cross-border clusters

Among the 56 cross-border clusters identified in this study, three (snpCL1, snpCL3 and snpCL8) stood out for size, number of countries involved and likelihood of recent transmission (*i.e.* low mean SNP difference between each isolate pair of the cluster, supplementary figure S2). The main characteristics related to cluster composition, drug-resistance profile of the isolates and epidemiological data of the MDR-TB cases involved are reported in supplementary table S5, while the corresponding minimum spanning trees are shown in figure 4. SnpCL3 and snpCL8 involve four Western EU countries and seem to be linked to migration from sub-Saharan Africa. SnpCL3 isolates are indistinguishable from those reported by WALKER *et al.* [10] for both genetic background and pattern of mutations, and indeed additional WGS-based relatedness analysis confirmed these MDR-TB cases as part of the “Horn of Africa” outbreak.

TABLE 2 Drug-resistance profile of the 2217 *Mycobacterium tuberculosis* isolates included in the database stratified by lineage

Lineage	R	H	MDR-TB	Z	SLIDs	FQs	XDR-TB
1 (EAI)	32 (1.5)	31 (1.6)	30 (1.5)	13 (1.0)	0 (0.0)	1 (0.2)	0 (0.0)
2 (Beijing/East-Asian non-Beijing)	638 (29.7)	632 (31.9)	630 (32.1)	443 (34.4)	337 (48.4)	257 (44.2)	179 (54.1)
3 (Delhi-CAS)	63 (2.9)	53 (2.7)	53 (2.7)	22 (1.7)	2 (0.3)	10 (1.7)	1 (0.3)
4.1 (Euro-American)	390 (18.1)	347 (17.5)	341 (17.4)	206 (16.0)	71 (10.2)	77 (13.3)	38 (11.5)
4.2 (Euro-American)	199 (9.3)	191 (9.6)	189 (9.6)	122 (9.5)	48 (6.9)	49 (8.4)	25 (7.6)
4.3 (LAM)	183 (8.5)	166 (8.4)	163 (8.3)	95 (7.4)	38 (5.5)	50 (8.6)	19 (5.7)
4.4, 4.5, 4.6 (Euro-American)	77 (3.6)	70 (3.5)	69 (3.5)	48 (3.7)	22 (3.2)	5 (0.9)	1 (0.3)
4.7, 4.8, 4.9 (Euro-American)	555 (25.8)	483 (24.3)	476 (24.3)	333 (25.9)	175 (25.1)	131 (22.5)	67 (20.2)
5 (West-Africa 1)	1 (0.0)	1 (0.1)	1 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Unknown	13 (0.6)	10 (0.5)	10 (0.5)	4 (0.3)	4 (0.6)	1 (0.2)	1 (0.3)
Total	2151 (97.0)	1984 (89.5)	1962 (88.5)	1286 (58.0)	697 (31.4)	581 (26.8)	331 (15.4)

Data are presented as n (%). R: rifampicin; H: isoniazid; MDR: multidrug resistant; TB: tuberculosis; Z: pyrazinamide; SLID: second-line injectable drug; FQ: fluoroquinolone; XDR: extensively drug resistant.

TABLE 3 Cross-border clusters identified using an SNP-based approach

SNP-based cross-border clusters (≤ 5 SNPs threshold)			
Cross-border cluster name	Isolates n	Countries involved n	Lineage classification
snpCL 1	30	3	4.8 (mainly T)
snpCL 2	20	2	4.2.1 (Ural)
snpCL 3	16	4	4.6.2 (Euro-American)
snpCL 4	14	2	4.8 (mainly T)
snpCL 5	13	2	4.2.1 (Ural)
snpCL 6	13	3	4.8 (mainly T)
snpCL 7	12	3	2.2.1 (Beijing)
snpCL 8	12	4	4.2.2 (Euro-American)
snpCL 9	12	2	4.8 (mainly T)
snpCL 10	12	2	4.1.2.1 (Haarlem)
snpCL 11	10	2	4.3.3 (LAM)
snpCL 12	10	5	4.2.2 (Euro-American)
snpCL 13	9	6	2.2.1 (Beijing)
snpCL 14	7	4	2.2.1 (Beijing)
snpCL 15	7	5	2.2.1 (Beijing)
snpCL 16	7	2	2.2.1 (Beijing)
snpCL 17	5	4	4.3.3 (LAM)
snpCL 18	5	2	2.2.1 (Beijing)
snpCL 19	5	2	4.8 (mainly T)
snpCL 20	5	2	4.3.3 (LAM)
snpCL 21	5	2	4.2.2.1 (TUR)
snpCL 22	5	2	4.1.2.1 (Haarlem)
snpCL 23	4	3	4.3.3 (LAM)
snpCL 24	4	3	2.2.1 (Beijing)
snpCL 25–34 [#]	3	2 to 3	
snpCL 35–56 [¶]	2	2	

A sequential number from 1 to 56 was assigned to each cross-border cluster identified by SNP-based analysis, based on the size of the cluster and starting from the largest one. SNP: single nucleotide polymorphism; snpCL: SNP-based cluster. [#]: 10 clusters comprising three isolates from three (n=1) and two (n=9) countries, with the lineage distribution comprising 2.2.1 Beijing (n=3), 4.8 mainly T (n=2), 4.7 mainly T (n=1), 4.1.2.1 Haarlem (n=2), 4.2.2 Euro-American (n=1) and 4.1.1.1 X-type (n=1); [¶]: 22 clusters comprising two isolates from two countries, with the lineage distribution comprising 2.2.1 Beijing (n=7), 4.8 mainly T (n=3), 4.1.2.1 Haarlem (n=3), 4.2.1 Ural (n=2), 4.3.3 LAM (n=2), 4.3.1 LAM (n=1), 4.1.2 Euro-American (n=1), 4.2.2 Euro-American (n=1), 4.9, H37Rv-like (n=1) and 3 Delhi-CAS (n=1).

Conversely, the pattern of mutations and lineage of the isolates belonging to snpCL8 had not been previously reported in the literature, but, similarly to snpCL3, the majority (seven of 12) of MDR-TB cases belonging to this cluster were migrants from Somalia.

Differently from snpCL3 and snpCL8, the MDR-TB cases involved in snpCL1 originated from Europe, with >70% reporting Romania as country of birth. Notably, >90% (11 of 12) MDR-TB cases belonging to snpCL1 notified by Italy were resistant to fluoroquinolones and, among those, 50% were also resistant to bedaquiline. These include TB patients born in Romania and Italy.

Discussion

We have shown that implementation of WGS-based surveillance for monitoring the emergence of MDR-TB outbreaks nationally and internationally is not only feasible, but has the potential to provide supporting evidence to better elucidate cross-border transmission patterns in Europe.

Overall, the cross-border clustering rate resulting from this study is significantly lower than the one reported in the previous EU molecular surveillance project of MDR/XDR-TB based on VNTR typing and covering a timeframe of 9 years [3]. The reduced clustering rate observed in this study could be the result of the difference in sampling time between the two studies, and the lower coverage and selected sampling of the previous study, in addition to the higher resolution power of WGS compared to MIRU-VNTR [3].

TABLE 4 Country contribution to cross-border clusters[#]

Country	RR/MDR-TB cases in cluster n	Percentage among clustered isolates	Percentage of clustered RR/MDR-TB cases among submitted ones [¶]
Austria	9	2.8	36.0
Belgium	5	1.6	26.3
Bulgaria	6	1.9	9.2
Croatia	0	–	–
Czech Republic	6	1.9	24.0
Denmark	0	–	–
Estonia	5	1.6	7.6
Finland	2	0.6	22.2
France	11	3.5	11.1
Germany	40	12.7	16.3
Hungary	2	0.6	9.1
Ireland	3	0.9	13.0
Italy	38	12.0	29.9
Latvia	3	0.9	3.5
Lithuania	43	13.6	24.0
Netherlands	5	1.6	20.8
Norway	1	0.3	10
Poland	7	2.2	7.1
Portugal	0	–	–
Romania	101	32.0	11.7
Slovakia	2	0.6	25.0
Slovenia	0	–	–
Spain	9	2.8	13.6
Sweden	8	2.5	32.0
UK	10	3.2	12.3
Total	316	100	14.3

RR: rifampicin resistant; MDR: multidrug resistant; TB: tuberculosis. [#]: a cross-border cluster is defined as two or more RR/MDR *Mycobacterium tuberculosis* complex isolates having a single nucleotide polymorphism difference ≤ 5 , and isolated in at least two different countries; [¶]: n=2217.

Approximately 70% of the RR/MDR-TB patients involved in cross-border clusters were born in the country of notification. However, this percentage varied significantly among EU countries, being higher among Eastern EU countries and Baltic States and not exceeding 10% in several Western EU countries, suggesting different RR/MDR-TB transmission patterns in Western and Eastern EU countries. This is shown by the examples of three major cross-border clusters circulating in the EU. Two clusters, snpCL3 and snpCL8, showed similar dynamics. These clusters included patients born in Somalia and Sudan and notified by Austria, France, Germany, Italy, Netherlands and Sweden, suggesting that transmission occurred exclusively among the migrant population, possibly along the route of migration prior to entering the EU/EEA, or in the country of origin. Although data on country of birth were not systematically available, we did not identify evidence of MDR-TB transmission between migrants and people residing in Europe, owing probably to the efforts to identify cases shortly after entering the EU/EEA, to the better living conditions and accessibility to healthcare in general for migrants in Europe compared to the conditions during migration [12], and to their limited social integration in the majority of EU/EEA countries. The third cluster, snpCL1, represents a different example of transmission dynamics, in which 70% of the TB patients were born or were resident in the country of notification. Conventional epidemiological investigations conducted independently from this study confirmed that recent transmission had indeed occurred for at least one third of these patients. Most worryingly, this cluster included MDR-TB isolates with additional resistance to fluoroquinolones and bedaquiline. Although these pre-XDR-TB cases have so far only been diagnosed in Italy, these patients have different nationalities and have regular contacts with compatriots, thus increasing the chances of cross-border transmission of resistance. This cluster may represent a public health threat for the EU population if further transmission is not adequately prevented, illustrating the utility of prospective WGS in MTBC surveillance.

This is the first study reporting on the systematic use of an EU-wide WGS-based approach for MTBC surveillance purposes. To ensure data comparability we applied a common and standardised WGS

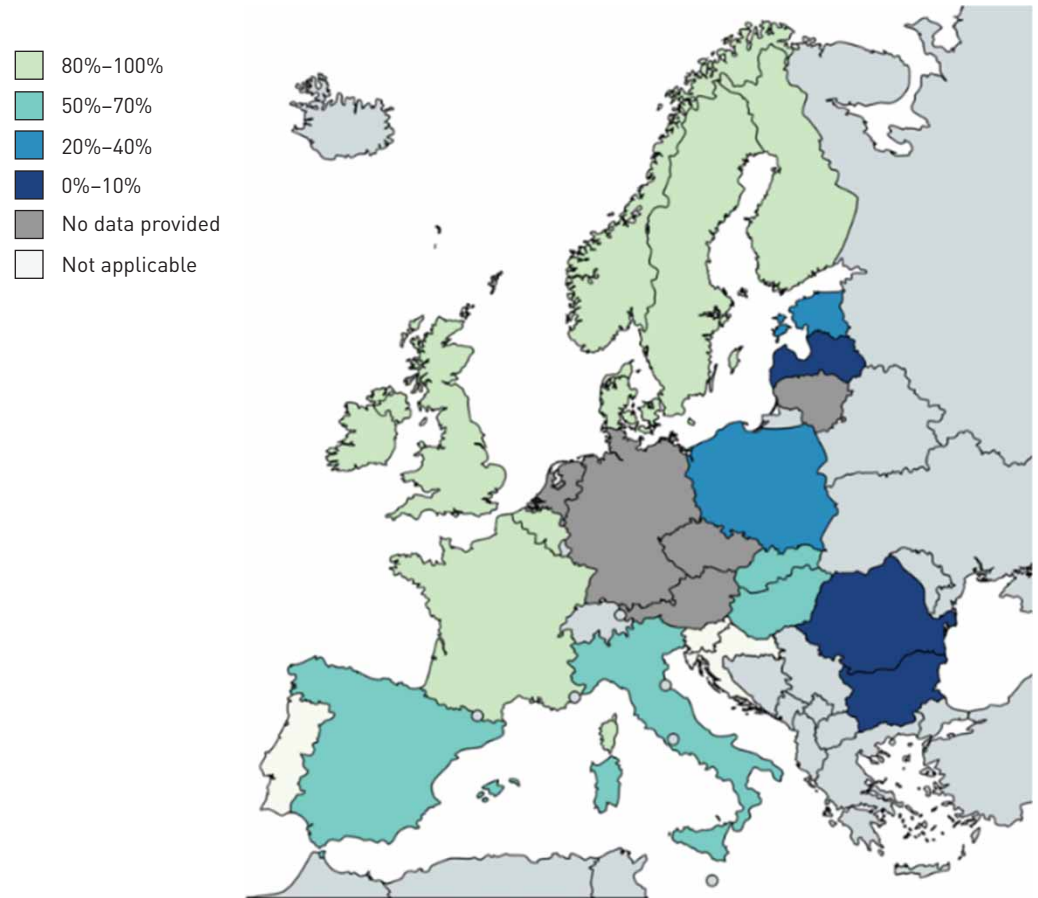


FIGURE 3 Map showing the percentage of rifampicin-resistant/multidrug-resistant tuberculosis cases of foreign origin in cross-border clusters per European Union/European Economic Area country. The map was generated using the online tool available at www.mapchart.net/.

analytical pipeline to all the collected isolates whose sequence data fulfilled pre-defined quality criteria. The relevance of this WGS-based surveillance project was widely recognised across EU/EEA countries, allowing us to reach a high coverage with respect to the MDR-TB cases notified to ECDC; therefore, our data are representative of the RR/MDR-MTBC isolates currently circulating in the EU/EEA. Importantly, this study allowed EU/EEA countries lacking WGS capacity to genotype their RR/MDR-MTBC strains, thus generating a database to study national MTBC transmission patterns and building WGS analytical capacity.

Despite the successful implementation of the pilot study several challenges were identified and lessons were learnt to improve further implementation of WGS for TB in the EU/EEA. First, although WGS has the potential to provide data for close to “real-time surveillance”, the intrinsic delays due to culture, batching and referral of the MTBC isolates to the sequencing laboratories affected our capacity to identify clusters promptly and to perform a prospective type of analysis. Although beyond the scope of the study, this limited the impact of our results on public health, as successfully shown in other local settings [29, 30]. The availability of more affordable and portable sequencing platforms could contribute to reducing this delay by increasing accessibility to this technology. Similarly, when feasible routinely, direct sequencing from clinical samples will further shorten the time to results, thus allowing real-time surveillance.

Second, complete basic clinical and epidemiological data were only available from 11 of the countries contributing to the study; few countries have a centralised TB surveillance system or a system that allows the National TB Reference Laboratory staff to directly access this information. We acknowledge that the use of WGS data alone likely results in an over-estimation of transmission and that only by complementing it with detailed clinical and epidemiological data is it possible to fully interpret transmission dynamics. This requires linking of clinical, epidemiological and laboratory data. Given the need for data protection, it will be imperative for future WGS-based surveillance systems to ensure data handling according to GDPR and local regulations and to secure it through an adequate protection system.

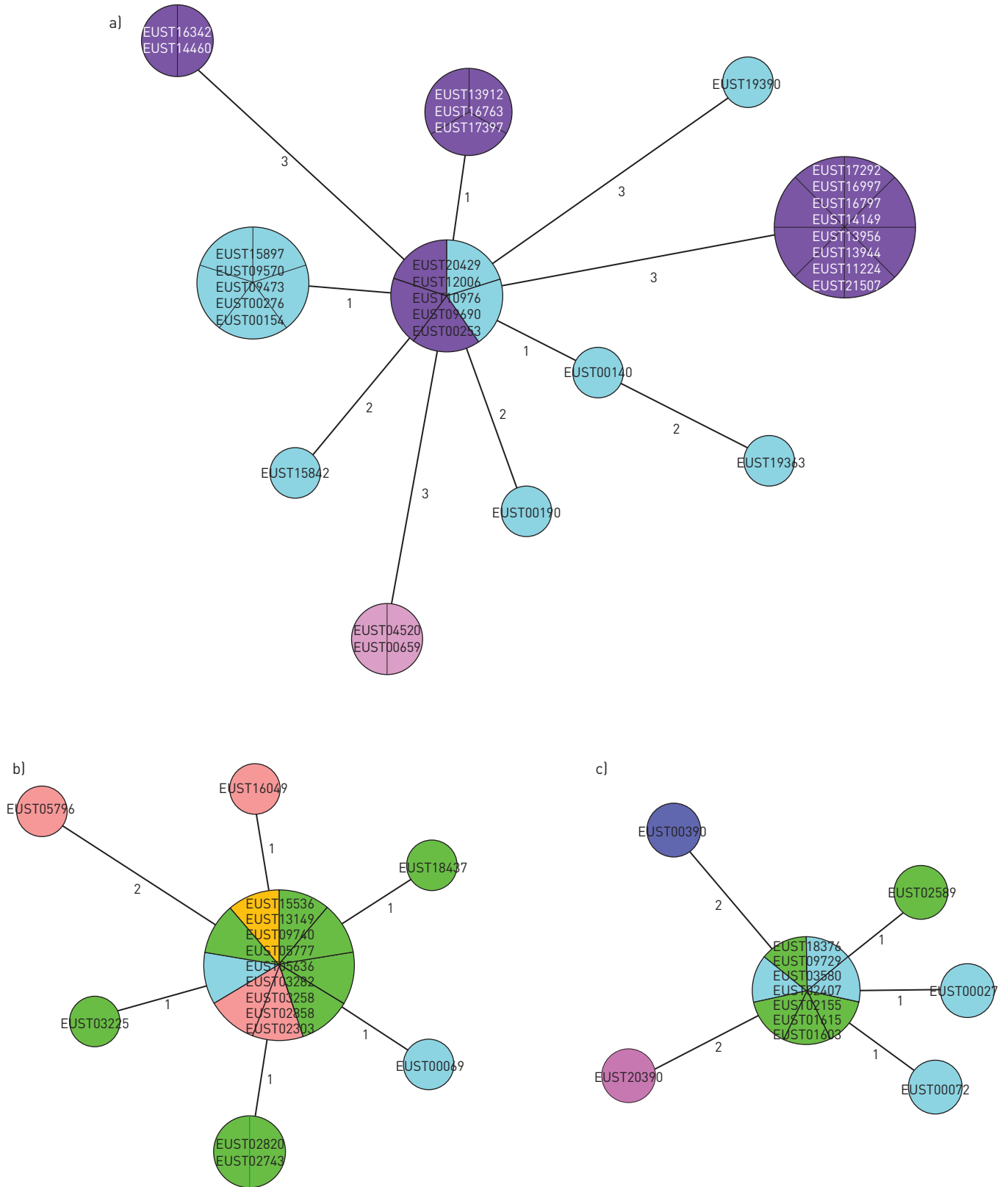


FIGURE 4 Minimum spanning tree of a) snpCL1, b) snpCL3 and c) snpCL8. The numbers on the branches indicate the genetic distance in single nucleotide polymorphism (SNP) differences to the nearest isolate. SNP differences between distant isolates can be reconstituted by summing the number of SNPs on the branches. Different colours indicate the different European Union/European Economic Area countries where the strains were isolated.

Third, participating countries were able to assess the WGS data in the pilot study database and to proactively perform additional analysis through the study webserver interface. This increased equal accessibility of this technology for countries with health systems lacking WGS in-country capacity. Last, upon the identification of cross-border clusters, it will be important to define specific criteria, such as resistance to second-line and new anti-TB drugs or high likelihood of recent transmission, to justify an international epidemiological investigation and to use the available resources most effectively.

The results of the pilot study and the lessons learnt will direct further work on the establishment of an operational platform combining WGS-based TB typing and epidemiological data. The aim is to establish an EU/EEA centralised WGS-typing/epi-data database for identification of cross-border TB and MDR-TB clusters and to investigate relevant clusters with all countries involved. The pilot study and future work will further contribute to standardisation of WGS for TB and build capacity in the EU/EEA.

Acknowledgements: We would like to thank all the staff at Regional Reference Laboratories for the collection and referral of RR/MDR-MTBC isolates, and the operational contact point for epidemiology for providing valuable data.

The ECDC molecular surveillance project participants are as follows. Ewa Augustynowicz-Kopeć: Dept of Microbiology, National Tuberculosis and Lung Diseases Research Institute, Warsaw, Poland; Elizabeta Bachiyska: National Center of Infectious and Parasitic Diseases, Sofia, Bulgaria; Agnes Bakos: Koranyi National Institute for Pulmonology, Budapest, Hungary; Roxana Coriu: Pulmonology, Marius Nasta Institute, Bucharest, Romania; Věra Dvořáková: National Reference Laboratory for Mycobacteria, Prague, Czech Republic; Lanfranco Fattorini: Dept of Infectious Diseases, National Institute of Health, Rome, Italy; Margaret Fitzgibbon: Irish Mycobacteria Reference Laboratory, St. James's Hospital, Dublin, Ireland; Agnieszka Głogowska: Dept of Microbiology, National Tuberculosis and Lung Diseases Research Institute, Warsaw, Poland; Ramona Groenheit: Public Health Agency of Sweden, Solna, Sweden; Marjo Haanperä: Dept of Health Security, Finnish Institute for Health and Welfare (THL), Helsinki, Finland; Laura Herrera León: Instituto de Salud Carlos III, Madrid, Spain; Daniela Homorodean: Clinical Hospital of Pneumology, Cluj-Napoca, Romania; Alexander Indra: Austrian Reference Laboratory for Mycobacteria, Austrian Agency for Health and Food Safety, Vienna, Austria; Sarah Jackson: Health Protection Surveillance Centre, Dublin, Ireland; Tiina Kummik: Dept of Mycobacteriology, Tartu University Hospital, Tartu, Estonia; Troels Lillebaek: International Reference Laboratory of Mycobacteriology, Statens Serum Institut, Copenhagen, Denmark; Rita Macedo: Tuberculosis National Reference Laboratory, Portuguese National Institute of Health, Lisbon, Portugal; Vanessa Mathys: Unit Bacterial Diseases Service, Infectious Diseases in Humans, Sciensano, Brussels, Belgium; Anne Torunn Mengshoel: National Reference Laboratory for Mycobacteria, Dept of Bacteriology, Norwegian Institute of Public Health, Oslo, Norway; Matthias Merker: German Center for Infection Research, partner site Borstel-Hamburg-Lübeck-Riems, Borstel, Germany; Molecular and Experimental Mycobacteriology, National Reference Center for Mycobacteria, Research Center Borstel, Borstel, Germany; Darshaalini Nadarajan: National and WHO Supranational Reference Center for Mycobacteria, Research Center Borstel, Borstel, Germany; Anders Norman: International Reference Laboratory of Mycobacteriology, Statens Serum Institut, Copenhagen, Denmark; Ilva Pole: Dept of Mycobacteriology, Center of Tuberculosis and Lung Diseases, Riga East University Hospital, Riga, Latvia; Igor Porvaznik: Clinical Microbiology Dept, National Institute for Tuberculosis, Lung Diseases and Thoracic Surgery, Vyšné Hágy, Slovakia; Wladimir Sougakoff: Laboratory of Bacteriology, National Reference Centre for Mycobacteria (CNR-MyRMA), AP-HP, Paris, France; Andrea Spitaleri: Emerging Bacterial Pathogens Unit, Division of Immunology, Transplantation and Infectious Diseases, IRCCS San Raffaele Scientific Institute, Milan, Italy; Sara Truden: National Reference Laboratory for Mycobacteria, University Clinic for Respiratory and Allergic Diseases, Golnik, Slovenia; Laima Vasiliauskaitė: Dept of Physiology, Biochemistry, Microbiology and Laboratory Medicine, Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University, Vilnius, Lithuania; Centre of Laboratory Medicine, Tuberculosis Laboratory, Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania; Ljiljana Žmak: National Reference Laboratory for Tuberculosis, Croatian Institute of Public Health, Zagreb, Croatia.

Author contributions: All study participants contributed significantly to the results presented in this manuscript.

Conflict of interest: R. Anthony reports grants from ECDC public tender OJ/2017/OCS/7766, during the conduct of the study. T.A. Kohl has nothing to disclose. A. de Neeling reports grants from ECDC public tender OJ/2017/OCS/7766, during the conduct of the study. V. Nikolayevskyy reports grants from ECDC, during the conduct of the study. C. Ködmön has nothing to disclose. F.P. Maurer has nothing to disclose. S. Niemann reports grants from the German Center for Infection Research, Excellenz Cluster Precision Medicine in Chronic Inflammation EXC 2167, Leibniz Science Campus Evolutionary Medicine of the LUNG (EvoLUNG) and ECDC public tender OJ/2017/OCS/7766, during the conduct of the study. D. van Soolingen has nothing to disclose. M.J. van der Werf has nothing to disclose. D.M. Cirillo reports grants from ECDC public tender OJ/2017/OCS/7766, during the conduct of the study. E. Tagliani reports grants from ECDC public tender OJ/2017/OCS/7766, during the conduct of this study.

Support statement: This project was funded by the European Centre for Disease Prevention and Control (framework contact ECDC/2017/012). The funder of the study was involved in the revision of the manuscript. Funding information for this article has been deposited with the Crossref Funder Registry.

References

- 1 European Centre for Disease Prevention and Control/World Health Organization Regional Office for Europe. Tuberculosis Surveillance in Europe 2020–2018 Data. Stockholm, European Centre for Disease Prevention and Control, 2020.

- 2 European Centre for Disease Prevention and Control. Molecular Typing for Surveillance of Multidrug-Resistant Tuberculosis in the EU/EEA - March 2017. Stockholm, European Centre for Disease Prevention and Control, 2017.
- 3 De Beer JL, Kodmon C, van der Werf MJ, *et al.* Molecular surveillance of multi- and extensively drug-resistant tuberculosis transmission in the European Union from 2003 to 2011. *Euro Surveill* 2014; 19: 20742.
- 4 Wylie DH, Davidson JA, Grace Smith E, *et al.* A quantitative evaluation of MIRU-VNTR typing against whole-genome sequencing for identifying *Mycobacterium tuberculosis* transmission: a prospective observational cohort study. *EBioMedicine* 2018; 34: 122–130.
- 5 Meehan CJ, Goig GA, Kohl TA, *et al.* Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. *Nat Rev Microbiol* 2019; 17: 533–545.
- 6 Walker TM, Kohl TA, Omar SV, *et al.* Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis* 2015; 15: 1193–1202.
- 7 Roetzer A, Diel R, Kohl TA, *et al.* Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med* 2013; 10: e1001387.
- 8 Gröschel MI, Walker TM, van der Werf TS, *et al.* Pathogen-based precision medicine for drug-resistant tuberculosis. *PLoS Pathog* 2018; 14: e1007297.
- 9 Pankhurst LJ, Del Ojo Elias C, Votintseva AA, *et al.* Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. *Lancet Respir Med* 2016; 4: 49–58.
- 10 Walker TM, Merker M, Knoblauch AM, *et al.* A cluster of multidrug-resistant *Mycobacterium tuberculosis* among patients arriving in Europe from the Horn of Africa: a molecular epidemiological study. *Lancet Infect Dis* 2018; 18: 431–440.
- 11 Fiebig L, Kohl TA, Popovici O, *et al.* A joint cross-border investigation of a cluster of multidrug-resistant tuberculosis in Austria, Romania and Germany in 2014 using classic, genotyping and whole genome sequencing methods: lessons learnt. *Euro Surveill* 2017; 22: 30439.
- 12 European Centre for Disease Prevention and Control. Multidrug-Resistant Tuberculosis in Migrants, Multicountry Cluster – 13 April 2017. Stockholm, European Centre for Disease Prevention and Control, 2017.
- 13 European Centre for Disease Prevention and Control. Extensively Drug-Resistant (XDR) Tuberculosis – Multi-Country Cluster, Romania. 21 October 2016. Stockholm, European Centre for Disease Prevention and Control, 2016.
- 14 Nikolayevskyy V, Kranzer K, Niemann S, *et al.* Whole genome sequencing of *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks: a systematic review. *Tuberculosis (Edinb)* 2016; 98: 77–85.
- 15 Hatherell HA, Colijn C, Stagg HR, *et al.* Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. *BMC Med* 2016; 14: 21.
- 16 Nikolayevskyy V, Niemann S, Anthony R, *et al.* Role and value of whole genome sequencing in studying tuberculosis transmission. *Clin Microbiol Infect* 2019; 25: 1377–1382.
- 17 European Centre for Disease Prevention and Control. Expert Opinion on Whole Genome Sequencing for Public Health Surveillance. Stockholm, European Centre for Disease Prevention and Control, 2016.
- 18 Tagliani E, Cirillo DM, Ködmön C, *et al.* EUSeqMyTB to set standards and build capacity for whole genome sequencing for tuberculosis in the EU. *Lancet Infect Dis* 2018; 18: 377.
- 19 Kohl TA, Harmsen D, Rothgänger J, *et al.* Harmonized genome wide typing of tubercle bacilli using a web-based gene-by-gene nomenclature system. *EBioMedicine* 2018; 34: 131–138.
- 20 Kohl TA, Utpatel C, Schleusener V, *et al.* MTBseq: a comprehensive pipeline for whole genome sequence analysis of *Mycobacterium tuberculosis* complex isolates. *PeerJ* 2018; 6: e5895.
- 21 Walker TM, Merker M, Kohl TA, *et al.* Whole genome sequencing for M/XDR tuberculosis surveillance and for resistance testing. *Clin Microbiol Infect* 2017; 23: 161–166.
- 22 Miotto P, Tessema B, Tagliani E, *et al.* A standardised method for interpreting the association between mutations and phenotypic drug resistance in *Mycobacterium tuberculosis*. *Eur Respir J* 2017; 50: 1701354.
- 23 Allix-Béguec C, Arandjelovic I, Bi L, *et al.* Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. *N Engl J Med* 2018; 379: 1403–1415.
- 24 Homolka S, Projahn M, Feuerriegel S, *et al.* High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. *PLoS One* 2012; 7: e39855.
- 25 Merker M, Blin C, Mona S, *et al.* Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet* 2015; 47: 242–249.
- 26 Kadura S, King N, Nakhoul M, *et al.* Systematic review of mutations associated with resistance to the new and repurposed *Mycobacterium tuberculosis* drugs bedaquiline, clofazimine, linezolid, delamanid and pretomanid. *J Antimicrob Chemother* 2020; 75: 2031–2043.
- 27 Coll F, McNerney R, Guerra-Assunção JA, *et al.* A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Comm* 2014; 5: 4812.
- 28 WHO Regional Office for Europe/European Centre for Disease Prevention and Control. Tuberculosis Surveillance and Monitoring in Europe 2019–2017 Data. Copenhagen, World Health Organization Regional Office for Europe, 2019.
- 29 Packer S, Green C, Brooks-Pollock E, *et al.* Social network analysis and whole genome sequencing in a cohort study to investigate TB transmission in an educational setting. *BMC Infect Dis* 2019; 19: 154.
- 30 Genestet C, Tatai C, Berland JL, *et al.* Prospective whole-genome sequencing in tuberculosis outbreak investigation, France, 2017–2018. *Emerging Infect Dis* 2019; 25: 589–592.