Expanded tracking of a Beijing *Mycobacterium tuberculosis* strain involved in an outbreak in France

Charlotte Genestet[1,2×], Joao Perdigao[3×], Marta Herranz[4,5], Sandra R. Maus[4,5], Jean-Luc Berland[1,2], Alvaro Chiner-Oms[6], Iñaki Comas[7,8], Patricia Muñoz[4,5,9,10], Isabel Portugal[3], Oana Dumitrescu[1,2+*], Laura Pérez Lago[4,5+*], Darío García de Viedma[4,5,9+*]

[1]CIRI - Centre International de Recherche en Infectiologie, Ecole Normale Supérieure de Lyon, Université Claude Bernard Lyon-1, Inserm U1111, CNRS UMR5308, France

[2]Hospices Civils de Lyon, Institut des Agents Infectieux, Lyon, France

[3]Research Institute for Medicines (iMed.ULisboa), Faculdade de Farmácia, Universidade de Lisboa, Lisboa, Portugal

[4]Servicio de Microbiología Clínica y Enfermedades Infecciosas, Hospital General Universitario Gregorio Marañón

[5]Instituto de Investigación Sanitaria Gregorio Marañón (IiSGM), Madrid, Spain

[6]Centro Superior de Investigación en Salud Pública (FISABIO)-Universitat de València, Valencia, Spain

[7]Instituto de Biomedicina de Valencia-CSIC, Valencia, Spain

[8]CIBER Salud Pública (CIBERESP), Spain

[9]CIBER Enfermedades Respiratorias (CIBERES), Madrid, Spain

[10]Departamento de Medicina, Universidad Complutense, Madrid

***Corresponding authors:**

dgviedma2@gmail.com; lperezg00@gmail.com

Phone: (+34) 91426510.

Postal address: C/Dr. Esquerdo 46 (28007), Madrid, Spain

oana.dumitrescu@chu-lyon.fr

92 Rue Pasteur, 69007 Lyon, France

+: these authors contributed equally as senior authors

×: These authors contributed equally as first authors

Summary

Advanced epidemiological surveillance supported on genomic epidemiology were necessary to fully describe a *Mycobacterium tuberculosis* Beijing strain (ARA, Auvergne-Rhône-Alpes region) outbreak in France. The index case was a migrant from Cape Verde with a two-year history of tuberculosis, high bacillary load, and frequent trips within France and to Portugal. We designed an ARA-specific PCR to complete the tracking of this outbreak strain. The ARA-specific PCR was applied on 160 Beijing isolates from independent cases from the Auvergne-Rhône-Alpes region and on 25 cases from Cape Verde migrants in the *Île-de-France* region. No more cases were found, indicating that the previous surveillance had been exhaustive enough to capture the whole dimension of the outbreak in France. Next, we performed a cross-border surveillance, applying the PCR on 38 Beijing isolates in Portuthe gal. In all four cases (all migrants from Cape Verde) the ARA strain was identified. Whole genome sequences analysis of all ARA isolates representatives indicated a diversification of the strain into two variants present one in France and one in Portugal.

Running title: Expanded tracking of an outbreak *strain of M. tuberculosis*

Key words: tuberculosis, outbreak, Beijing, cross-border, WGS, strain-specific PCR

The current epidemiological scenario of tuberculosis (TB) is a consequence of international migratory movements. The overlapping between recent transmission in the host country with independent importations of related variants from strains prevalent at the countries of origin of migrants, challenges epidemiological surveillance (Abascal et al., 2019). In this complex setting, standard molecular epidemiology is not discriminatory enough to properly dissect TB transmission, whilst whole genome sequencing (WGS) has demonstrated to be the most suitable approach (Abascal et al., 2019; Jajou et al., 2018; Stucki et al., 2016). Besides the refinement of genomic epidemiology, expanded surveillance beyond the most active local hotspots to other related settings and even related countries is necessary to completely understand the dynamics of TB transmission (Abascal et al., 2019; Fiebig et al., 2017).

A cluster of a pansusceptible Beijing SIT1 strain was described in France (in the Auvergne-Rhône-Alpes (ARA) region) involving 14 cases (June 2017-April 2018) (Genestet et al., 2019). The index case was a Cape Verdean migrant with a two-year history of TB and high bacillary load. Before living in the ARA region, the index case had lived in the Île-de-France region and reported frequent travels to Portugal. The initial cluster members lived in the same city of the ARA region and were identified (MtbC15-9 type) by standard molecular epidemiology. However, WGS analysis and extended epidemiological contact tracing guided by WGS data and supported on interviews with the involved subjects were needed to complete the true magnitude of the cluster. This allowed us to detect other cases living in four distant cities.

The fact that the index case made frequent trips and had a high bacillary load alerted us of additional potential exposures. This justified evaluating whether the exhaustive surveillance efforts made to complete the description of the cluster had been enough to fully rule out the existence of probable new cases.

The aim of this study was to implement a rapid, low-cost extended surveillance strategy to quickly track the ARA Beijing strain using strain-specific PCR. Similar approaches have shown to be successful for tracing high-risk strains and to rapidly update the presence of a strain within a population (Perez-Lago et al., 2019; Perez-Lago et al., 2017). Thirty-five strain-specific SNPs were identified by comparing the SNPs extracted from WGS data from six of the 14 cases involved in the cluster previously sequenced and deposited (Genestet et al., 2019) with those from an in-house database (IBV-CSIC, Valencia, Spain) that contains SNPs from 4,761 strains and thus illustrates the geographic and phylogenetic diversity of the MTB complex (Comas et al., 2013).

The strain-specific PCR (Table 1) consisted of a four-plex-allele-specific-oligonucleotide (ASO)-PCR targeting four SNPs, three from the ARA-Beijing strain (selecting from the 35 specific SNPs the only two mapping in essential genes, the remaining one mapping in a non-essential gene but involving a synonymous change, with the aim of looking for stability of the markers) and the marker SNP for the Beijing lineage. The primers for the PCR (Supplementary file) were designed to be homologous for either the ARA-Beijing alleles (SNPs 1 and 3) or the alternative allele (SNP2) to offer three alternative amplification patterns depending on whether the isolates corresponded to the ARA-Beijing, to another Beijing strain different from the ARA-Beijing, or to any another non-Beijing strain (Figure 1).

We aimed to determine ARA-Beijng strain-related potential additional cases around the areas where the index case had moved, in France (ARA and Île-de-France) but also in Portugal (door of entrance from migrants from Cape Verde to Europe), which added a cross-border nature to our tracking.

The ARA-Beijing PCR was applied on 160 Beijing isolates (as per spoligotyping) from independent cases from the ARA region and on 25 cases from Cape

Verde migrants in the Île-de-France region (2016-18). It was performed directly on boiled crude extracts of the stored isolates; all the analyses were carried out in two weeks at a total cost of 2.5 euro per determination. No additional cases infected by the ARA strain were found, indicating that the previous surveillance had been exhaustive enough to capture the whole dimension of the outbreak.

The methodology was then transferred to Portugal and the PCR was applied on a Lisbon-based convenience sample of 27 Beijing isolates (2008-14) representing all Beijing isolates in Portugal for that period, identified by spoligotyping. The three-band pattern expected for the ARA-Beijing strain was not obtained for any of these isolates. However, an amplification pattern indicating the presence of two (SNPs 2 and 3) of the three ARA-Beijing specific SNPs and the Beijing marker SNP were observed for four isolates (Figure 2). Interestingly, all four isolates corresponded to Cape Verde migrants (years 2009-10).

The four isolates with a partial ARA-Beijing pattern were analyzed by WGS as described elsewhere (Perez-Lago et al., 2014). The short genetic distances (Figure 3) with the ARA-Beijing French representatives confirmed they were same strain, although lacking SNP 1. The Portugal representatives were closely related among them, two showing a distance of 0 SNPs and the other two differing in 1 and 2 SNPs in comparison to the first two cases.

Two groups were observed in the network of relationships when all isolates were included, based on the presence/absence of two signature SNPs (including the previously mentioned one). These two SNPs (C126818T and T4304361C) were shared by the 14 isolates in France but were absent in those from Portugal. This may indicate two diversification routes, one in each country. Similarly, in another event studied by our group, two closely related independently distributed variants of a strain exported

from Peru were involved, one detected in Florence and the second in other Italian/Spanish cities (Acosta et al., 2019). Isolate sampling in Portugal was carried out three years before the sampling in France. This may indicate the existence of a recent common ancestor that circulated in Portugal before the emergence of a variant harboring the two SNPs that were later successfully transmitted in France. One of the SNPs is synonymous and the other one (C126818T) led to an aminoacid substitution (Ala1242Thr) in the gene Rv0107c that encodes a potential cation-transporting ATPase I (CtpI).

We screened a local database with SNPs of ~5,300 Beijing strains from publicly available WGS data obtained from 41 different countries at the European Nucleotide Archive until the end of July 2017. One isolate (from Azerbaijian, year of isolation 2015) showed 25 of the 26 SNPs that were specific for the Beijing ARA strain. No explanation was found for the connection with Azerbaijian.

The fact that all cases detected in Portugal infected with this strain were migrants from Cape Verde, which coincides with the origin for the index case in the France outbreak, suggests the strain originated in Cape Verde and is likely a prevalent strain in that country. Performing strain-specific PCR locally in Cape Verde may offer valuable information on this last aspect.

The systematic use of a simple, low-cost strategy based on strain-specific PCR for a Beijing outbreak strain allowed us to confirm the appropriateness of the exhaustive surveillance performed to capture all outbreak cases without the need of additional WGS analysis. New cases infected by a variant of the outbreak strain revealed the likely cross-border nature of the outbreak strain and the diversification of the strain into two variants, one in each country. This study illustrates how nationwide efforts can easily be

complemented to understand the complexity of transmission in the new global scenario of TB.

Acknowledgements

This work was supported by ERANet-LAC [ELAC2015/T08-0664] and *Instituto de Salud Carlos III* [AC16/00057, FIS15/01554, FIS13/01207, CP15/00075, PI16/01449, /01PI19/00331] and cofounded by European Regional Development Funds from the European Commission: "A way of making Europe". Miguel Servet Contract CP15/00075) to LPL. JP [CEECIND/00394/2017] is supported by FCT through *Estímulo Individual ao Emprego Cientifico*. The authors are grateful to Dainora Jaloveckas (cienciatraducida.com) for editing and proofreading assistance.

Ethical statement: The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to.

Conflict of interest: No author declare conflict of interest.

Figures

Figure 1. A) Amplification patterns expected for ARA-strain PCR (*in silico* analysis). B) Experimental patterns obtained for a selection of representative French isolates.

Figure 2. ARA-strain PCR amplification patterns for ARA- *Île-de-France* (left panel) and Portugal (right panel) representative isolates

Figure 3. Network of relationships based on whole genome sequencing data that include ARA-strain representatives from France and Portugal. Each white dot represents a SNP. The size of the circles is proportional to the number of isolates in each node. Dark gray circles represent France isolates and light gray circles represent Portugal isolates. The data that support the findings of this study are openly available in [ENA] at https://www.ebi.ac.uk/ena/browser/home], reference number yet not available. Mv1: it refers to a median vector, which represents a node in the network that should be occupied by an isolate, but it was not sampled in our study.

References

Abascal, E., Perez-Lago, L., Martinez-Lirola, M., Chiner-Oms, A., Herranz, M., Chaoui, I., . . . Garcia-de-Viedma, D. (2019). Whole genome sequencing-based analysis of tuberculosis (TB) in migrants: rapid tools for cross-border surveillance and to distinguish between recent transmission in the host country and new importations. *Euro Surveill, 24*(4). doi: 10.2807/1560-7917.ES.2019.24.4.1800005

Acosta, F., Agapito, J., Cabibbe, A. M., Caceres, T., Sola, C., Perez-Lago, L., . . . Garcia de Viedma, D. (2019). Exportation of MDR TB to Europe from Setting with Actively Transmitted Persistent Strains in Peru. *Emerg Infect Dis, 25*(3), 596-598. doi: 10.3201/eid2503.180574

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K. E., Kato-Maeda, M., . . . Gagneux, S. (2013). Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nat Genet, 45*(10), 1176-1182. doi: 10.1038/ng.2744

Fiebig, L., Kohl, T. A., Popovici, O., Muhlenfeld, M., Indra, A., Homorodean, D., . . . Haas, W. (2017). 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, 22*(2). doi: 10.2807/1560-7917.ES.2017.22.2.30439

Genestet, C., Tatai, C., Berland, J. L., Claude, J. B., Westeel, E., Hodille, E., . . . Lyon, T. B. Study Group. (2019). Prospective Whole-Genome Sequencing in Tuberculosis Outbreak Investigation, France, 2017-2018. *Emerg Infect Dis, 25*(3), 589-592. doi: 10.3201/eid2503.181124

Jajou, R., de Neeling, A., Rasmussen, E. M., Norman, A., Mulder, A., van Hunen, R., . . . van Soolingen, D. (2018). A Predominant Variable-Number Tandem-Repeat Cluster of Mycobacterium tuberculosis Isolates among Asylum Seekers in the Netherlands and Denmark, Deciphered by Whole-Genome Sequencing. *J Clin Microbiol, 56*(2). doi: 10.1128/JCM.01100-17

Perez-Lago, L., Campos-Herrero, M. I., Canas, F., Copado, R., Sante, L., Pino, B., . . . Samper, S. (2019). A Mycobacterium tuberculosis Beijing strain persists at high rates and extends its geographic boundaries 20 years after importation. *Sci Rep, 9*(1), 4687. doi: 10.1038/s41598-019-40525-6

Perez-Lago, L., Comas, I., Navarro, Y., Gonzalez-Candelas, F., Herranz, M., Bouza, E., & Garcia-de-Viedma, D. (2014). Whole genome sequencing analysis of intrapatient microevolution in Mycobacterium tuberculosis: potential impact on the inference of tuberculosis transmission. *J Infect Dis, 209*(1), 98-108. doi: 10.1093/infdis/jit439

Perez-Lago, L., Izco, S., Herranz, M., Tudo, G., Carcelen, M., Comas, I., . . . Garcia de Viedma, D. (2017). A novel strategy based on genomics and specific PCR reveals how a multidrug-resistant Mycobacterium tuberculosis strain became prevalent in Equatorial Guinea 15 years after its emergence. *Clin Microbiol Infect, 23*(2), 92-97. doi: 10.1016/j.cmi.2016.10.006

Stucki, D., Ballif, M., Egger, M., Furrer, H., Altpeter, E., Battegay, M., . . . Fenner, L. (2016). Standard Genotyping Overestimates Transmission of Mycobacterium tuberculosis among Immigrants in a Low-Incidence Country. *J Clin Microbiol, 54*(7), 1862-1870. doi: 10.1128/JCM.00126-16

Table

Table 1. Multiplex ASO-PCRs design, primers, allele targets and PCR products sizes.

SNP	Gene	Nucleotide change	Position	Amplicon size (bp)	Target	Primer sequence
1	Rv3829	T/C	4304361	386	ARA	Rv3829F:5´- CGATCCGGTGGTCCAATTCA-3´ Rv3829R:5´- CCTCCGGGACACCCAAC-3´
2	Rv0697	T/C	797094	297	No-ARA	Rv0697F:5´- GCCGCTGACCTACCTAGTG-3´ Rv0697R:5´ TCGGCAGTTGCAACCCA-3´
3	Rv1303	C/A	1460199	225	ARA	Rv1303F:5´-´- AGGAACCGGTCGCAACTTATTA-3´ Rv1303R:5´- CGAAGTCACTTTGGCGCG-3´
4	Rv2952	G/A	3304966	139	Beijing	Rv2952F:5´- GCCGAAAACCTGCCCTTC-3´ Rv2952R:5´- CGGCGTATGGGAAGTACCT -3´

Figure 1

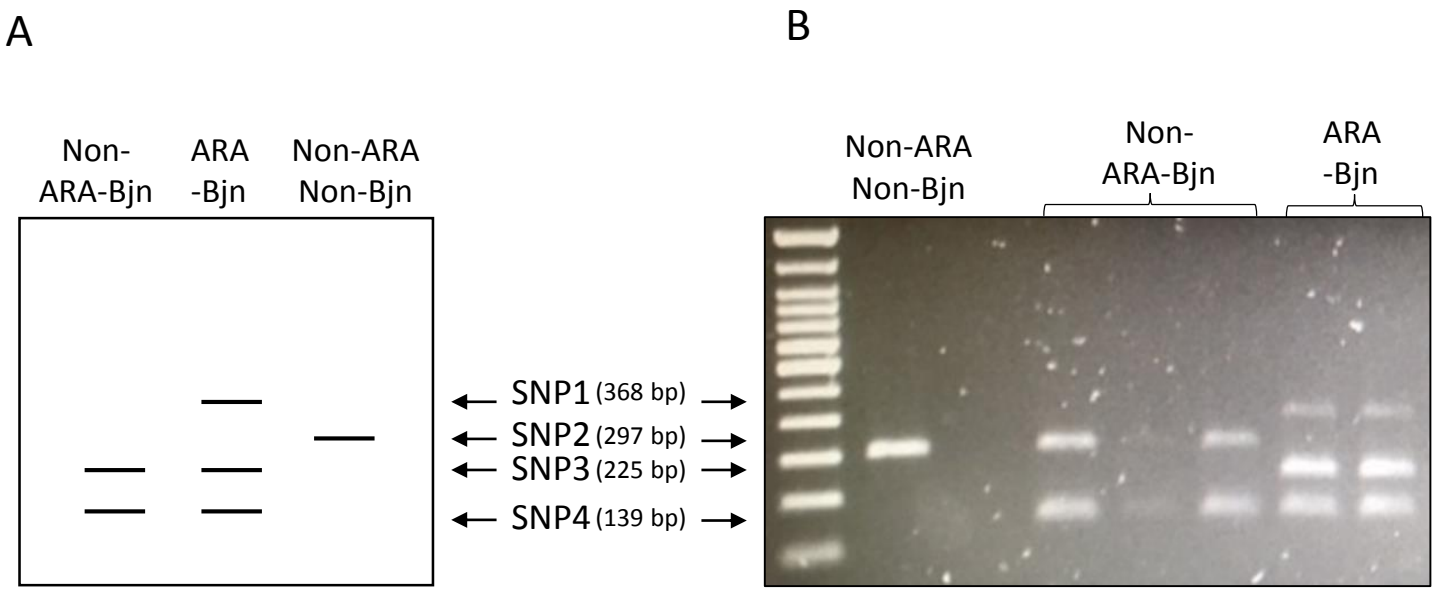

A
Non-
ARA-Bjn
ARA
-Bjn
Non-ARA
Non-Bjn
SNP1 (368 bp)
SNP2 (297 bp)
SNP3 (225 bp)
SNP4 (139 bp)
B
Non-ARA
Non-Bjn
Non-
ARA-Bjn
ARA
-Bjn

Figure 2

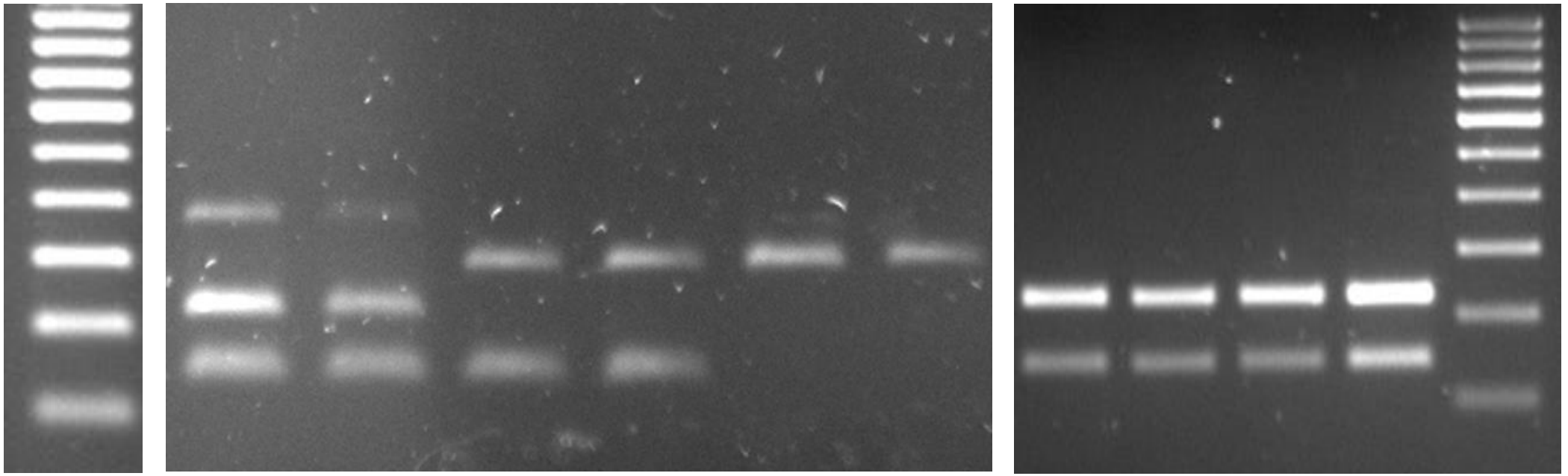
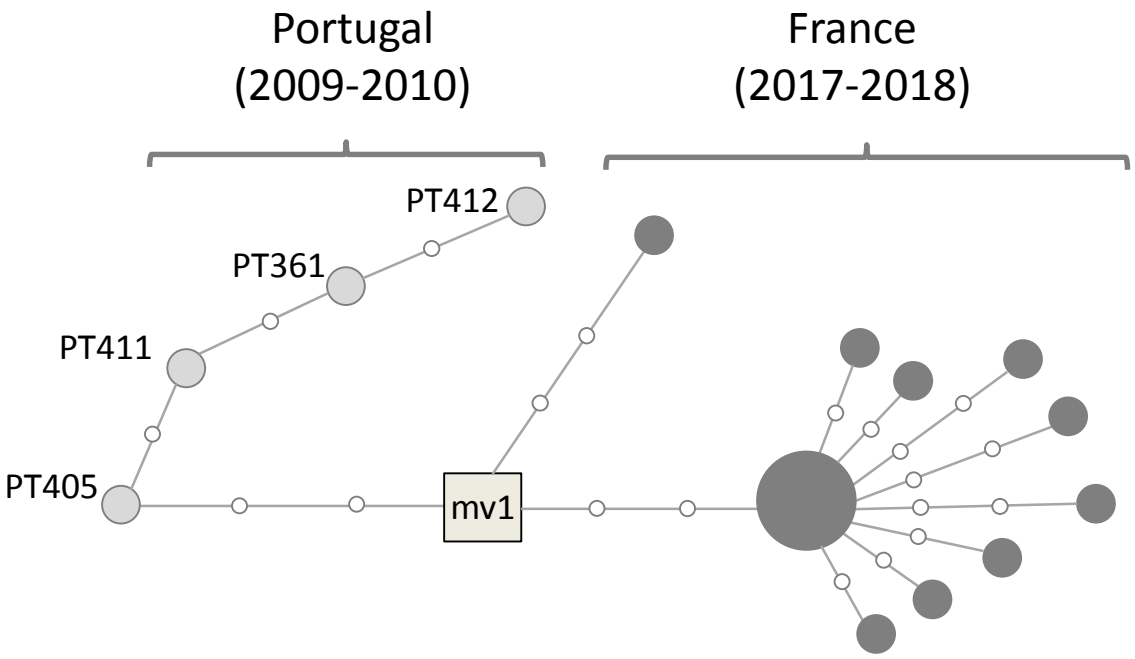

Mw
ARA-BJN
Non-ARA-BJN
Non-ARA-non-BJN
ARA-BJN
Mw
France isolates pattern
Portugal isolates pattern

Figure 3


Portugal
(2009-2010)
France
(2017-2018)
PT412
PT361
PT411
PT405
mv1