Title: Increase of diversity of mumps virus genotype G SH variants circulating among a highly immunized population: Spain, 2007-2019

A.M. Gavilán[1,7], F. Díez-Fuertes[1,8], J.C. Sanz[2,7], A. M. Castellanos[1,7], N. López-Perea[3,7], S.M. Jiménez[4], C. Ruiz-Sopeña[5], J. Masa-Calles[3,7], L. García-Comas[6], F. de Ory[1,7], M. Pérez-Olmeda[1,8], A. Fernández-García[1,7]* and J. E. Echevarría[1,7]*

1. Centro Nacional de Microbiología, Instituto de Salud Carlos III. Madrid, Spain.
2. Laboratorio Regional de Salud Pública de la Comunidad de Madrid. Madrid, Spain.
3. Centro Nacional de Epidemiología, Instituto de Salud Carlos III. Madrid, Spain.
4. Servicio de Microbiología. Hospital de Segovia. Segovia, Spain.
5. Servicio de Epidemiología. Consejería de Sanidad de Castilla y León. Valladolid, Spain.
6. Servicio de Epidemiología. Consejería de Sanidad de la Comunidad de Madrid. Madrid, Spain.
7. Consorcio de Investigación Biomédica en Red de Epidemiología y Salud Pública. Madrid, Spain.
8. Consorcio de Investigación Biomédica en Red de Enfermedades Infecciosas. Madrid, Spain.

*Equal contribution as co-senior authors

Corresponding author:

Aurora Fernández-García

Phone Number: +34 918223724

E-mail: aurorafg@isciii.es

Running Title: Mumps diversity in immunized population

Abstract

MuV caused three epidemic waves in Spain since genotype G emerged in 2005, despite high vaccination coverage. SH gene sequencing according to WHO protocols allowed the identification of seven relevant variants and 88 haplotypes. While the originally imported MuVi/Sheffield.GBR/1.05/-variant prevailed during the first two waves, it was subsequently replaced by other variants originated by either local evolution or importation, according to the additional analysis of hypervariable NCRs. The time of emergence of the MRCA of each MuV variant clade was concordant with the data of the earliest sequence. The analysis of Shannon entropy showed an accumulation of variability on six particular positions as the cause of the increase on the number of circulating SH variants. Consequently, SH gene sequencing needs to be complemented with other more variable markers for mumps surveillance immediately after the emergence of a new genotype, but the subsequent emergence of new SH variants turns it unnecessary.

Key words:

mumps, mumps virus, SH variants, genetic diversity, phylogeny, molecular epidemiology, laboratory surveillance.

Abreviated summary (36 words)

Increase on diversity of mumps variants circulating within a highly immunized population in Spain from 2007 to 2019 improved the performance of the WHO recommended molecular tools for mumps virus surveillance based on SH gene sequencing.

Introduction

Mumps is a vaccine-preventable disease caused by mumps virus (MuV), which is a member of the Orthorubulavirusgenus (*Paramyxoviridae* family). The MuV virion is an enveloped particle containing a non-segmented negative strand RNA molecule of 15,384 nucleotides (nt)[1].

Mumps is a highly transmissible disease whose main clinical symptom is swelling of the parotid glands. Previously, nonspecific symptoms can appear such as fever, headache, malaise and anorexia. Other less frequent symptoms include orchitis, mastitis, oophoritis and pancreatitis. After infection, rare complications can arise, such as aseptic meningitis and encephalitis [2]. Infected people can transmit the virus through respiratory droplets from two before to five days onset of parotitis [2]. Asymptomatic and subclinical infected people can also transmit the virus.

Mumps vaccination was introduced into the Spanish childhood immunization schedule in combination with measles and rubella virus as a part of Measles Mumps Rubella Vaccine (MMR vaccine) in 1981 [3]. As a result of the high vaccine coverage, mumps incidence dropped from 211 cases per 100,000 inhabitants in 1982 to 3-31 cases per 100,000 inhabitants from 1997. However, in Spain, as in other countries, the virus still generates outbreaks among highly vaccinated population and causes cyclic epidemic waves that peak every 4–7 years, most recently in 1996, 2000, 2007, 2013 and 2019 [3][4]. The origins of the capacity of MuV to circulate among immunized people are not well established. The use of the Rubini vaccine strain from 1992 to 1999 in Spain, which was later withdrawn due to reduced effectiveness, has been pointed as a major cause [4]. Additional hypothetical causes include waning of vaccine-induced immunity [2] [5], incomplete cross-neutralization between heterologous mumps virus genotypes [6], and a decreased capacity of the antibodies induced by the vaccine to protect against presently circulating wild-type MuV as a result of long-term antigenic drift [7]. Despite this continuous circulation, the prevalence of severe complications and hospitalizations due to MuV have decreased significantly [4].

WHO recommends mumps genotyping by sequencing of the SH gene as a part of the epidemiologic surveillance system required to control the disease [8]. This molecular tool

enables us to identify the general patterns of viral circulation, as well as to trace the source of the outbreaks and the chains of transmission. According to this molecular method, 12 genotypes have been recognized: A, B, C (including former genotype E), D, F, G, H, I, J, K (including former genotype M), L and N [9].

The K genotype of mumps was the first reported to be circulating in Spain in the Basque country (Northern Spain) between 1987 and 1990 [10]. Genotype H was predominant from 1996 to 2003, until it was replaced by genotype G in 2005, while genotypes A, C, D and J were reported in several isolated cases or limited outbreaks [11]. Genotype G is still predominant after 15 years. A single SH variant (MuVi/Sheffield.GBR/1.05/) of genotype G was predominant from 2005 to 2015, although cryptic strain replacement between epidemic waves was demonstrated by the use of additional molecular tools based on genomic non coding regions (NCRs) [12].

In the present work we show the implications for molecular surveillance of the emergence of new predominant SH variants after an increase of the genomic diversity of mumps virus circulating in Spain during the period of study from 2007 to 2019.

Materials and Methods

1. Samples and sequences

A total of 8.742 samples were tested from 2008 to 2019 at the National Center for Microbiology (CNM) as part of the activities of diagnosis and surveillance of mumps, of which 2464 were found to be positive by RT-PCR. Generally, every positive case was genotyped, except during high incidence periods, when one case per location and epidemiological week was randomly selected. The MuV genotype was investigated in 1153 of them and 1125 were genotype G, which constitutes the aim of this work, together with 121 Spanish sequences from 2007 to 2019 obtained from GenBank [13].

The first cases of relevant emerging and re-emerging variants were selected for sequencing of hypervariable non coding regions (NCRs) together with concomitant cases of previously circulating variants of the same locations (see results).

2. RNA extraction, genetic amplification and sequencing

Total nucleic acids were extracted with an automatic extractor (QIAsymphony, Qiagen) using a commercial kit (QIAsymphony Virus/Bacteria Midi Kit (96); Qiagen). Four hundred microliters of sample were extracted to obtain 40 µL of final eluate. The whole SH gene sequence (316 nt) was obtained from all samples according to a published protocol [14]. Genomic fragments including the NCRs located between the nucleoprotein (N) and phosphoprotein (P) coding sequences (CDSs) (N-P, 768 nt) and the Matrix protein (M) and Fusion protein (F) CDSs (M-F, 452 nt), respectively, were amplified in a subset of samples (see below), as previously described [12]. PCR-amplified products were purified by an enzymatic reaction using Illustra ExoProStar 1-Step (GE Health Care Life Science, Freiburg, Germany) and sequenced using the Sanger method with the ABI Big Dye Terminator Cycle Sequencing Kit (Applied Biosystems, Branchburg, NJ, United States) using the corresponding forward and reverse primers.

3. Phylogenetic analysis, identification of mutations and variant classification

Sequences were edited using BioEdit v.7.2.5 [15] and aligned with MAFFT v.7 [16]. Phylogenetic analysis were performed by the method of Maximum Likelihood (ML) using the IQ-TREE software via the webserver (W-IQ-TREE) [17], with the best evolutionary model previously selected in the model selection tool, based on the Akaike information criterion. The assessment of nodes was supported by ultrafast bootstrap approach (UFboot) [18]. Phylogenetic trees were edited using MEGA v.7 [19].

Every SH sequence was named in accordance with the WHO´s standard nomenclature [8]. Haplotypes (a set of identical sequences) were identified using the phylogenetic analysis and Basic Local Alignment Search Tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) [20] to identify any identical sequence in GenBank. Haplotypes were named using the name of the earliest sequence. Haplotypes which showed an extensive circulation were considered variants. The specific conditions for variant assignation were continuous detection for six months or more and/or spreading to three or more Spanish provinces.

4. Phylodynamic analysis

The time of emergence of the most recent common ancestor (MRCA) of the identified variants were estimated using SH sequences with the Bayesian Markov chain Monte Carlo (MCMC) coalescent method implemented in BEAST v1.10.4. [21]. Regression of root-to-tip genetic distances against sampling time was estimated using Tempest. [22]. BEAST analysis was carried out using a SRD06 codon-based evolutionary model and an uncorrelated relaxed molecular clock model with a lognormal rate distribution and the Bayesian Skyline Plot population growth model. The coalescent model was selected over other models because it assumes that a small random sample from a large population is included in the dataset. BEAST was running for at least 100 million MCMC steps, sampling every 5,000 steps, and removing 10% as burn-in. The convergence of MCMC chains was checked using Tracer v.1.7.1 [23], ensuring that the effective sample size (ESS) values were greater than 100 for each estimated parameter. The maximum clade credibility (MCC) tree was obtained from the sampled trees using TreeAnnotator v1.10.4, then visualized and edited with FigTree v1.4.4 [24].

5. Shannon entropy analysis

Nucleotide variability was measured as entropy using the Shannon entropy-two tool (https://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy.html) [25] looking for the difference in entropies between two MuV SH gene alignments in each position. The first alignment included sequences from the 2010-2015 epidemic wave and was used as background. The second one collected sequences from the 2016-2019 epidemic wave and was used as query. Shannon entropy score was calculated by summing the individual values of each position.

6. GenBank Accession Numbers

The sequences obtained in this study have been deposited in GenBank under accession numbers MN436853-MN437318, MN308288-MN308363, MN489001-MN489004, MN567307-MN567478, MT906866-MT907264, MT919088-MT919099, MT858762-MT858792, MT670354-MT670376, MT26985-MT27001, MW044679-MW044695 and MW657773-MW657788.

7. Ethics statement

The samples used in this work were obtained in the context of the Mumps Microbiological Surveillance Programme of the CNM and used in accordance with the requirements of Spanish biomedical research law (Ley 14/2007 de Investigación Biomédica). The protocol was approved by the Comité de Ética de la Investigación del Instituto de Salud Carlos III (approval no. Reference code: CEI PI 35–2015).

Results

An ML phylogenetic analysis was made with all the available MuV SH Spanish sequences from genotype G (1246) during the period of study, using the generalized time-reversible (GTR) as an evolutionary model (data not shown). In total, 88 haplotypes were identified from 2007 to 2019 (supplementary table). Seven of these haplotypes, comprising 1072 sequences (86.0%), met the criteria to be considered variants. Four hundred thirty-seven sequences (35.0%) belonged to the previously described MuVi/Sheffield.GBR/1.05/ variant (EU597478) [9].

The earliest sequence of each haplotype was used to construct an ML tree using the GTR as evolutionary model (figure 1). Most of the haplotypes were from the same genetic lineage and related to MuVi/Sheffield.GBR/1.05/, according to the topology of the phylogenetic tree (figure 1) and the nt differences (supplementary table). Six variants grouped into clades with their related haplotypes, sharing characteristic mutations in the SH sequences: MuVs/NewYork.USA/45.15/ (C248T); MuVs/Avila.ESP/11.16/ (A226T) and its derived variant MuVs/Madrid.ESP/50.16/2 subclade (A226T, T247C); MuVs/Avila.ESP/51.18/ (C13A); MuVs/Salamanca.ESP/24.19/ (T100C); MuVs/New_Jersey.USA/20.10/ (C67T). The only exception was a sequence (MuV/Segovia. ESP/39.19/2) of the MuVs/New_Jersey.USA/20.10/ clade, which also showed the characteristic mutation of the MuVs/Tarragona.ESP/20.11/ haplotype (G159T). Most of the haplotypes showed one to three mutations regarding the MuVi/Sheffield.GBR/1.05/ variant. Nine haplotypes had four or more mutations, two of which were grouped into variant clades: MuVs/Guadalajara.ESP/22.19/ (11 nt differences), which shares the C13A mutation with MuVs/Avila.ESP/51.18/, and

MuVs/Soria.ESP/48.19/ (4 nt differences), which shows the T100C mutation characteristic of the MuVs/Salamanca.ESP/24.19/ variant.

A total of 44 out of 88 (50%) of the haplotypes showed non-synonymous changes. Most of them have single amino acid mutation, however nine of them accumulate more than three. Three cases out of 1246 (0.24%) contained sequences of abnormal length. Two of them lost a stop codon and extended nineteen positions to reach a total of 76 amino acids instead of 57. Another one lost the final thirteen amino acids showing an SH sequence of 44 amino acids.

The characteristic mutations of the variants are silent in most cases (supplementary table). However, MuVs/Salamanca.ESP/24.19/ presents a L17P substitution and MuVs/NewJersey.USA/20.10/ has a P6L mutation.

An increase in the diversity of MuV SH variants and their derived haplotypes was observed throughout successive epidemic waves (figure 2). The MuVi/Sheffield.GBR/1.05/ variant was the only one identified in the 2005-2009 and 2010-2015 epidemic waves. However, in the last epidemic wave (2016-2019), several new circulating MuV variants were identified. The MuVs/Avila.ESP/11.16/ variant, was the first one, replacing the MuVi/Sheffield.GBR/1.05/ variant, which became undetectable by mid-2016 and disappeared for one year. Meanwhile, two new MuV variants (MuVs/NewYork.USA/45.15/ and MuVs/Madrid.ESP/50.16/2) co-circulated with MuVs/Avila.ESP/11.16/. In the mid-2017, the MuVi/Sheffield.GBR/1.05/ variant reappeared, co-existing with the latter. During 2018 and 2019, the dominant variants, MuVs/NewYork.USA/45.15/ and MuVs/Avila.ESP/11.16/, were substituted by three new ones: MuVs/Avila.ESP/51.18/, MuVs/Salamanca.ESP/24.19/ and MuVs/New_Jersey.USA/20.10/ (figure 2).

A phylodynamic analysis was implemented in BEAST to estimate the time to the MRCA of each identified MuV variant clade. The estimated time of emergence of each clade was consistent with the epidemiological data (figure 3).

Entropy Shannon analysis revealed an entropy score along the SH gene of 6,83 for 2010-2015, while 4,72 for 2016-2019. The difference in entropy was more concentrated in some particular nucleotide positions in the second period, which correspond to the characteristic mutations of the new variants (C13A, C67T, T100C, A226T, T247C, and C248T) (figure 4).

Additional molecular tools based on sequencing of MuV genomic fragments including NCR were used to increase the phylogenetic resolution [12], to establish whether the new MuV variants evolved from those circulating previously or were introduced by importation events. With this purpose, a phylogenetic tree was derived using the concatenated N-P, M-F NCR and SH sequences from the first five cases of the emerging MuVs/Avila.ESP/11.16/ variant and the last six cases of the previously dominant MuVi/Sheffield.GBR/1.05/ variant (figure 5, panel A). Sequences from cases of the MuVs/Avila.ESP/11.16/ variant grouped in a separate/different branch of cases of the MuVi/Sheffield.GBR/1.05/ variant, suggesting an importation event. In order to study the re-emergence of the MuVi/Sheffield.GBR/1.05/ variant in 2017, a phylogenetic tree was derived using the concatenated N-P, M-F NCR and SH sequences from 13 cases of the MuVi/Sheffield.GBR/1.05/ variant identified in 2017 and 2018, and simultaneous circulating variants. MuVi/Sheffield.GBR/1.05/ variant sequences grouped in a different clade from MuVs/Avila.ESP/11.16/ and MuVs/NewYork.USA/45.15/ (figure 5, panel B) suggesting importation. Finally, sequences of the MuVs/New_Jersey.USA/20.10/ variant, which produced a big outbreak in Segovia in mid-2019 were analyzed. According to the topology of the phylogenetic tree (figure 5, panel C), the concatenated N-P, M-F NCR and SH sequences from cases of the MuVs/New_Jersey.USA/20.10/ variant grouped in a clade. MuVi/Sheffield.GBR/1.05/ variant cases, which circulated earlier in 2019, were part of a different clade, but one sequence was at the base of the clade, suggesting that this variant evolved from existing one.

Discussion

Genotype G was the most frequently detected in Spain during the period of the study, as well as in other western countries and Japan [9] [11]. The MuVi/ Sheffield.GBR.1.05/ variant seemed

to be dominant, based on the sequences available in GenBank as previously described for the United States of America (USA) [26]. In Spain, MuV was able to circulate among the highly vaccinated population and generated successive epidemic waves [4][10][13]. It might be expected that periods of lower incidence could act as evolutionary bottlenecks in which some strains could disappear. The replacement of genotype H by the MuVi/Sheffield.GBR/1.05/ variant of genotype G in 2005 is an example of this [11]. However, according to the SH gene analysis, the MuVi/Sheffield.GBR/1.05/ variant was able to remain the dominant strain throughout two subsequent epidemic waves (2005-2009, 2010-2015) with no apparent replacement. Additional molecular tools, based on sequencing of MuV genomic fragments including NCRs, were necessary to reveal the hidden strain replacement between the two epidemic waves, that remained unnoticed with the WHO recommended genotyping tool [12]. However, in the course of the last epidemic wave (2016-2019), SH gene analysis was sufficient to reveal an increase of circulating variants of the G genotype mumps virus (figure 2). During this period, seven variants were identified, based on criteria detailed in the Materials and Methods, all of which belonged to the MuVi/Sheffield.GBR/1.05/ lineage. Haplotypes related to each variant grouped into the same clades of the ML phylogenetic tree (figure 1) and shared one or two specific mutations (supplementary table), suggesting that most of them were derived from each variant by genetic evolution. The number of mutations identified were concordant with the evolution rate previously estimated for the MuV SH gene ($1.71x10^{-3}$ substitutions/site/year) [27].

The increase of new variants seems to be related to the emergence of new strains of MuV by local evolution over the time, which can eventually establish a sustained circulation. These variants subsequently generate derived haplotypes that are generally unable to spread through the population. However, one of them, MuVs/Madrid.ESP/50.16/2 succeeded in establish circulation as a new variant. Nevertheless, an importation event was probably the origin of the MuVs/Avila.ESP/11.16/ variant circulation in Spain, as suggested by the phylogenetic analysis of the concatenated SH, N-P and M-F NCR regions (figure 4, panel A). It could be the also the

case for MuVs/NewYork/45.15/ variant, which was circulating extensively during the previous weeks in other countries, such as the USA [26]. This variant is characterized by mutation C248T and was previously named as Sheffield-C248 T by Mac Nall *et al*.[26]. As suggested by the NCR analysis (Figure 4, panel B), the re-emergence of the MuVi/Sheffield.GBR/1.05/ variant in 2017 was probably due to an importation event, rather than local evolution by reversion of the characteristic mutation from the co-circulating variants. In addition to importation events, local evolution was implied in the emergence of MuVs/New_Jersey.USA/20.10/ variant and its derived haplotypes, responsible for causing an important outbreak in Segovia in 2019 (figure 4, panel C). The high number of mutations identified in nine haplotypes, many of them non synonymous (supplementary table), could not be explained by the local evolution of sequences, according to the evolution rate estimated for the MuV SH gene, and probably corresponded to distinct self-limited importation events. Two of them (MuVs/Guadalajara.ESP/22.19/ and MuVs/Soria.ESP/48.19/) grouped into variant clades because shared the specific characteristic mutations.

The topology of the phylogenetic tree of MuV variants and haplotypes circulating in Spain is consistent, with the temporal data of the sequences by the epidemiological week and year in which they were detected (figure 2), and shows their sequential emergence over the time. These results were confirmed, using the MCMC coalescent method implemented in BEAST, which estimated the times of emergence of the MRCA of each identified variant, (figure 3), which always occurred before the first week of detection of each variant.

Although the total Shannon entropy score obtained along the SH sequence was higher for 2010-2015 than for 2016-2019 epidemic wave, suggesting more genetic diversity in the first wave, an unique variant (MuVi/Sheffield.GBR/1.05/) was the predominant and most of the haplotypes detected were not successful. The total Shannon entropy score during the 2016-2019 epidemic wave was lower, but the difference in the entropy value was concentrated in six nucleotide positions (figure 4), concordant with the specific variant point mutations (C13A, C67T, T100C, A226T, T247C, and C248T), resulting in an increase of the number of circulating variants. Co-

circulation of different haplotypes has been previously described [26]. However, the number of circulating SH variants associated to characteristic mutations described here was not found in the USA, where C248T seemed to be the only one selected [26]. The longer period of study of the present work could account as an explanation for this difference. Surprisingly, we found a lower frequency (0,24% versus 1,79 %) of SH protein sequences with modified length[26], all three from different locations and years.

The co-circulation of several variants of one predominant lineage that emerged by local evolution, and from multiple independent introductions has previously been reported in the USA, from combining molecular tools based on the full genome and epidemiological analysis [28]. Full genome sequencing is replacing partial sequencing for molecular epidemiology studies of mumps [28][29][30] and other viral diseases. However, the recommendation of full genome sequencing as a global tool for laboratory surveillance of mumps is not feasible at present, due to the lack of availability of deep-sequencing technology in many countries. Consequently, current WHO protocols for mumps genotyping will continue to be used in the immediate future. In a previous study [12], we showed that additional markers based on the sequencing of genomic regions including NCRs, are required for the laboratory surveillance of mumps, in addition to the WHO protocols based on the SH sequence, following the emergence of a new dominant genotype. Other authors have used these new molecular markers to improve the resolution in molecular epidemiology studies [31], although they recommend the use of the complete genome to track chains of transmission [29].

Consequently, our results show the WHO recommended protocol for mumps genotyping based on the sequencing of the SH gene could not be discriminative enough to trace the circulation patterns of MuV in the years following the emergence of a new dominant genotype, as suggested in a previous work [12]. Under these circumstances, additional auxiliary markers as NCR sequences are needed. However, as genetic variability accumulated, SH sequence became discriminative enough for this purpose, although additional markers are still needed to establish the origin of the new variants as the result of local evolution or importation. Initially, genetic

variation seems to spread randomly through the SH sequence while the original imported variant persist as the dominant. After several years of continuous circulation variability seems to concentrate on some particular positions that seems to be fixed to became new variants able to establish circulation, until they are replaced by new ones. However, such mutations are synonymous in most cases. On the contrary, mutants in other positions fail to spread and are detected as sporadic haplotypes, even when some changes are not synonymous. This fact suggests that SH gene would not be the target for positive selection. The comparative study of the full genomes of SH MuV haplotypes that were or were not able to establish circulation, could allow to identify features which could eventually facilitate virus-spreading among highly-immunized populations.

Study limitations

This study has two main limitations. The work made use of the samples received at the CNM through the specific surveillance mumps program and the Spanish MuV SH sequences available in GenBank. Although most Spanish provinces are represented in the dataset, they are not proportionally representative of their population. Fewer samples were received in 2012 and 2013 than expected, given the incidence of mumps (figure 2), probably due to the coexistence of large measles outbreaks that placed great pressure on the work of the surveillance system.

References

1. Rubin S, Eckhaus M, Rennick LJ, Bamford CGG, Duprex WP. Molecular biology, pathogenesis and pathology of mumps virus. J Pathol [Internet]. John Wiley and Sons Ltd; **2015** [cited 2020 Apr 23]; 235(2):242–252. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25229387

2. Lam E, Rosen JB, Zucker JR. Mumps: An update on outbreaks, vaccine efficacy, and genomic diversity. Clin Microbiol Rev. **2020**; 33(2):1–16.

3. Situación de la Parotiditis en España [Internet]. 1982. Available from: http://www.isciii.es/ISCIII/es/contenidos/fd-servicios-cientifico-tecnicos/fd-vigilancias-alertas/fd-enfermedades/fd-enfermedades-prevenibles-vacunacion/pdf_2017/Situacion_Parotiditis_ES_1982_2016.pdf

4. López-Perea N, Masa-Calles J, Torres de Mier M de V, et al. Shift within age-groups of mumps incidence, hospitalizations and severe complications in a highly vaccinated population. Spain, 1998–2014. Vaccine. **2017**; 35(34):4339–4345.

5. Ata Ur Rasheed M, Hickman CJ, McGrew M, et al. Decreased humoral immunity to mumps in young adults immunized with MMR vaccine in childhood. Proc Natl Acad Sci U S A. **2019**; 116(38):19071–19076.

6. Nö Jd J, Tecle T, Samuelsson A. Mumps virus neutralizing antibodies do not protect against reinfection with a heterologous mumps virus genotype [Internet]. Vaccine. 2001. Available from: www.elsevier.com/locate/vaccine

7. Šantak M, Lang-Balija M, Ivancic-Jelecki J, Košutič-Gulija T, Ljubin-Sternak S, Forcic D. Antigenic differences between vaccine and circulating wild-type mumps viruses decreases neutralization capacity of vaccine-induced antibodies. Epidemiol Infect. **2013**; 141(6):1298–1309.

8. Meeting NU. Weekly epidemiological record Relevé épidémiologique hebdomadaire. **2012**; 27(22):217–224.

9. Jin L, Örvell C, Myers R, et al. Genomic diversity of mumps virus and global distribution of the 12 genotypes. Rev Med Virol. **2015**; 25(2):85–101.

10. Cilla G, Montes M, Zapico MS, et al. Genetic characterization of historical epidemic mumps

viruses in northern Spain, 1987-1990. Infect Genet Evol [Internet]. Elsevier; **2014** [cited 2020 Apr 23]; 28:5–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25176599

11. Echevarría JE, Castellanos A, Sanz JC, et al. Circulation of mumps virus genotypes in Spain from 1996 to 2007. J Clin Microbiol. **2010**; 48(4):1245–1254.

12. Gavilán AM, Fernández-García A, Rueda A, et al. Genomic non-coding regions reveal hidden patterns of mumps virus circulation in Spain, 2005 to 2015. Eurosurveillance. European Centre for Disease Prevention and Control (ECDC); **2018**; 23(15).

13. Barrabeig I, Antón A, Torner N, Pumarola T, Costa J, Domínguez À. Mumps: MMR vaccination and genetic diversity of mumps virus, 2007-2011 in Catalonia, Spain. BMC Infect Dis. BMC Infectious Diseases; **2019**; 19(1):1–11.

14. Royuela E, Castellanos A, Sánchez-Herrero C, Sanz JC, Ory F De, Echevarria JE. Mumps virus diagnosis and genotyping using a novel single RT-PCR. J Clin Virol [Internet]. Elsevier B.V.; **2011**; 52(4):359–362. Available from: http://dx.doi.org/10.1016/j.jcv.2011.09.007

15. Hall TA. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. Nucleic Acids Symp Ser. **1999**; 41:95–98.

16. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. **2013**; 30(4):772–780.

17. Trifinopoulos J, Nguyen LT, Haeseler A von, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. **2016**; 44(W1):W232–W235.

18. Hoang DT, Chernomor O, Haeseler A von, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular biology and evolution. Mol Biol Evol [Internet]. **2018**; 35(2):518–522. Available from: http://www.iqtree.org.

19. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. **2016**; 33(7):1870–1874.

20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. **1990**; 215(3):403–410.

21. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the

BEAST 1.7. Mol Biol Evol. **2012**; 29(8):1969–1973.

22. Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. **2016**; 2(1):1–7.

23. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol. **2018**; 67(5):901–904.

24. Rambaut. http://tree.bio.ed.ac.uk/software/figtree/.

25. No Title. Available from: https://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy.html

26. McNall RJ, Wharton AK, Anderson R, et al. Genetic characterization of mumps viruses associated with the resurgence of mumps in the United States: 2015–2017. Virus Res. Elsevier B.V.; **2020**; 281.

27. Cui A, Rivailler P, Zhu Z, et al. Evolutionary analysis of mumps viruses of genotype F collected in mainland China in 2001-2015. Sci Rep. Springer US; **2017**; 7(1):1–7.

28. Wohl S, Metsky HC, Schaffner SF, et al. Combining genomics and epidemiology to track mumps virus transmission in the United States. PLoS Biol. **2020**; 18(2):e3000611.

29. Bodewes R, Reijnen L, Kerkhof J, et al. Molecular epidemiology of mumps viruses in the Netherlands, 2017-2019. PLoS One. Public Library of Science; 2020.

30. Magaña LC, Espinosa A, Marine RL, et al. Complete Genome Sequences of Mumps and Measles Virus Isolates from Three States in the United States. Am Soc Microbiol [Internet]. **2017** [cited 2020 Jun 22]; . Available from: https://doi.org/10.1128/genomeA

31. Bodewes R, Rooijen K van, Cremer J, Veldhuijzen IK, Binnendijk R van. Optimizing molecular surveillance of mumps genotype G viruses. Infect Genet Evol. Elsevier B.V.; **2019**; 69:230–234.

Figure legends

Figure 1. Phylogenetic tree of MuV SH sequences haplotypes. Analysis employed the ML method in IQ-tree, using TNe as the evolutionary model. The MuV G genotype reference sequence (MuVi/Gloucester.GBR/32.96 [G], AF280799) was used as outgroup. Colored branches denote different variant clades and the earliest SH variant sequence of each variant is indicated by colored dots. An asterisk next to the name of the SH sequences indicates those with ≥4 nt differences with respect to the MuVi/Sheffield.GBR/1.05/ variant. UltrafastBootstrap values > 80 are shown.

Figure 2. MuV genotype G SH variants with relevant circulation in Spain from 2007 to 2019.

Figure 3. Time-scale tree of MuV SH haplotype and variant sequences. The time of emergence of the MRCA is estimated in weeks, taking the last week of 2019 as the temporal reference. The week of each haplotype is included at the end of the name. The week of the MRCA of each variant is shown at the corresponding node. The MuV variant clades are identified by the previously used color code. Posterior probabilities ≥0.8 are shown as black spots.

Figure 4. Shannon entropy differences between 2010-2015 and 2016-2019 epidemic waves MuV SH sequences alignments. MuV SH sequences belonging to 2010-2015 epidemic waves were used as background (blue) and were represented at the positive Y-axis, whereas, sequences belonging to 2016-2019 were used as query (orange) and were represented at the negative Y-axis.

Figure 5. Phylogenetic tree of MuV genome regions including N-P, M-F NCRs and SH concatenated sequences. Analysis involved the ML method in W-IQ-TREE, using JC as evolutionary model. The MuV G genotype reference sequence (MuVi/Gloucester.GBR/32.96 [G], AF280799) was used as outgroup. Panel A. Analysis to study the origin of the MuVs/Avila.ESP/11.16/ variant. Green spots correspond to the MuVs/Avila.ESP/11.16/ variant and red spots to the MuVi/Sheffield.GBR/1.05/ variant. Panel B. The study of the origin of the

re-emergence of the MuVi/Sheffield.GBR/1.05/ variant. Green dots correspond to the MuVs/Avila.ESP/11.16/ variant, grey dots to the MuVs/NewYork.USA/45.15/ variant and uncoloured spots correspond to the re-emerged MuVi/Sheffield.GBR/1.05/ variant. Panel C. Analysis to study the origin of the MuVs/New_Jersey.USA/20.10/ variant. Pink dots correspond to the MuVs/New_Jersey.USA/20.10/ variant, while derived haplotypes (MuVs/Segovia.ESP/39.19/2, MuVs/Segovia.ESP/39.19/12 and MuVs/Segovia.ESP/39.19/32) are indicated by empty circles. Red samples correspond to the MuVi/Sheffield.GBR/1.05/ variant.

Footnote page

Conflict of interest

None declared.

Acknowledgements

We would like to thank Sara Ruiz and the Genomics Unit of the CNM-ISCIII for their technical assistance with sequencing. This work was supported by the "Instituto de Salud Carlos III" (PI15CIII/00023, PI19ICIII/0041). Ana M Gavilán was funded by CIBER de Epidemiología y Salud Pública (CIBERESP), ISCIII.

Authors' contributions

Ana María Gavilán: technical work, data analysis and writing as main author. Aurora Fernández-García and Juan Emilio Echevarría: design of the study, data analysis, and writing as main authors. Francisco Díez-Fuertes: technical bioinformatics, data analysis, drafting and revision of the manuscript. Noemi López-Perea: data analysis and drafting and revision of the manuscript. Ana M Castellanos: technical work. Juan Carlos Sanz, Silvia M Jiménez, Cristina Ruiz-Sopeña, and L. García-Comas: identification and confirmation of mumps cases and review and assistance in the editing the final version of the manuscript. J. Masa Calles, Mayte Pérez-Olmeda and Fernando de Ory: review and assistance in editing the final version of the manuscript.

Meeting where the information has previously been presented

23rd Annual Conference of the European Society for clinical virology. Increase in variability of mumps virus genotype G strains in Spain, 2008-2019. Manchester, 2021.

Corresponding author contact information

Aurora Fernández-García (aurorafg@isciii.es)

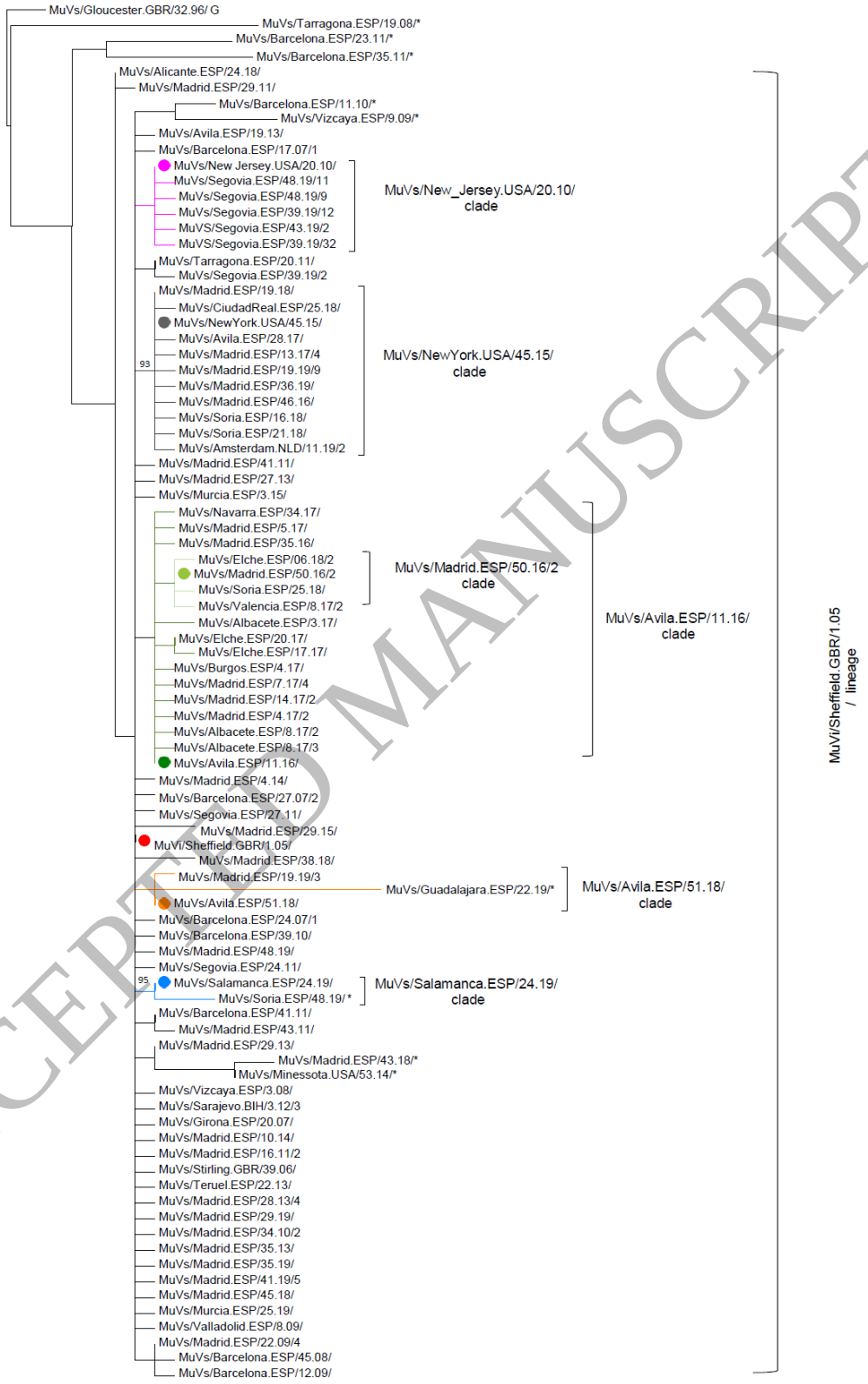

Figure 1
150x233 mm (.25 x DPI)

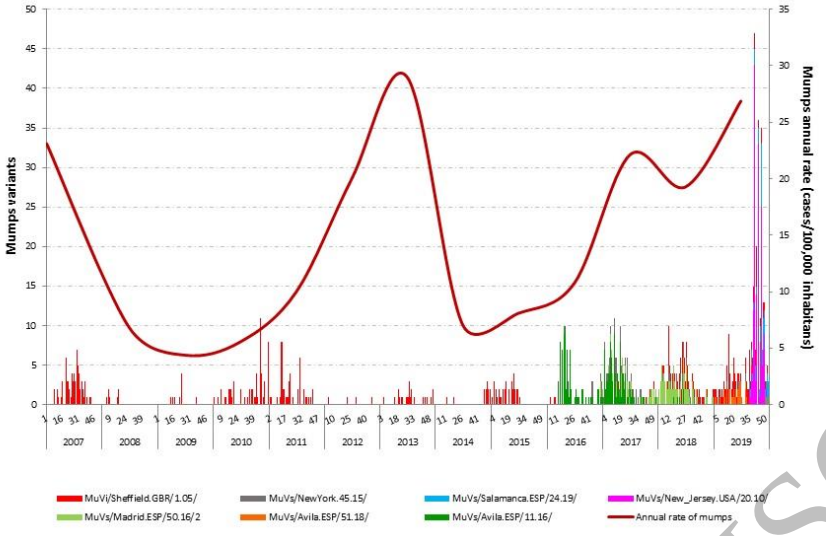

Figure 2
150x84 mm (.25 x DPI)

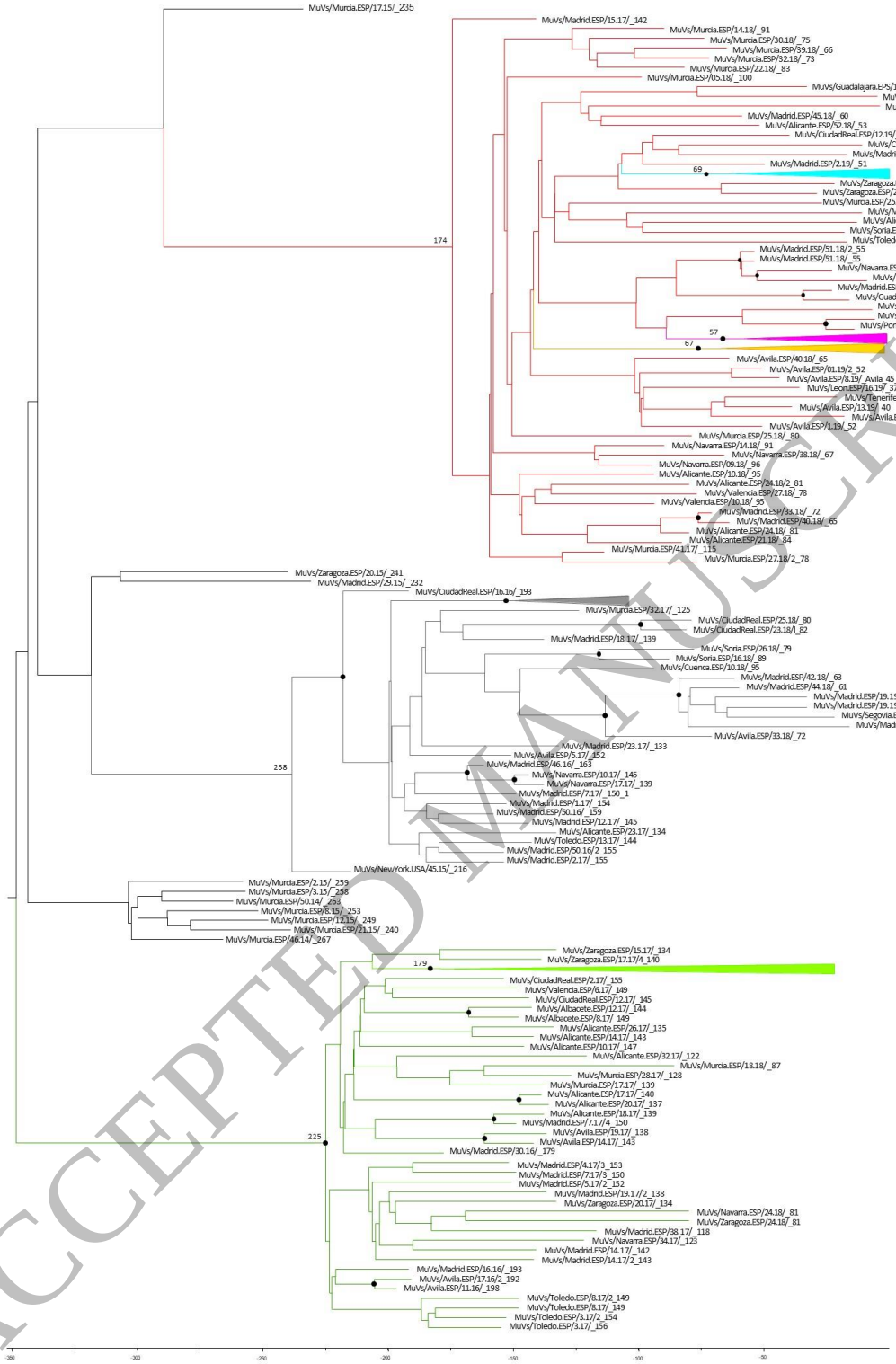

Figure 3
150x201 mm (.25 x DPI)

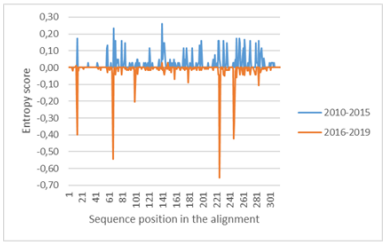

Figure 4
150x84 mm (.25 x DPI)

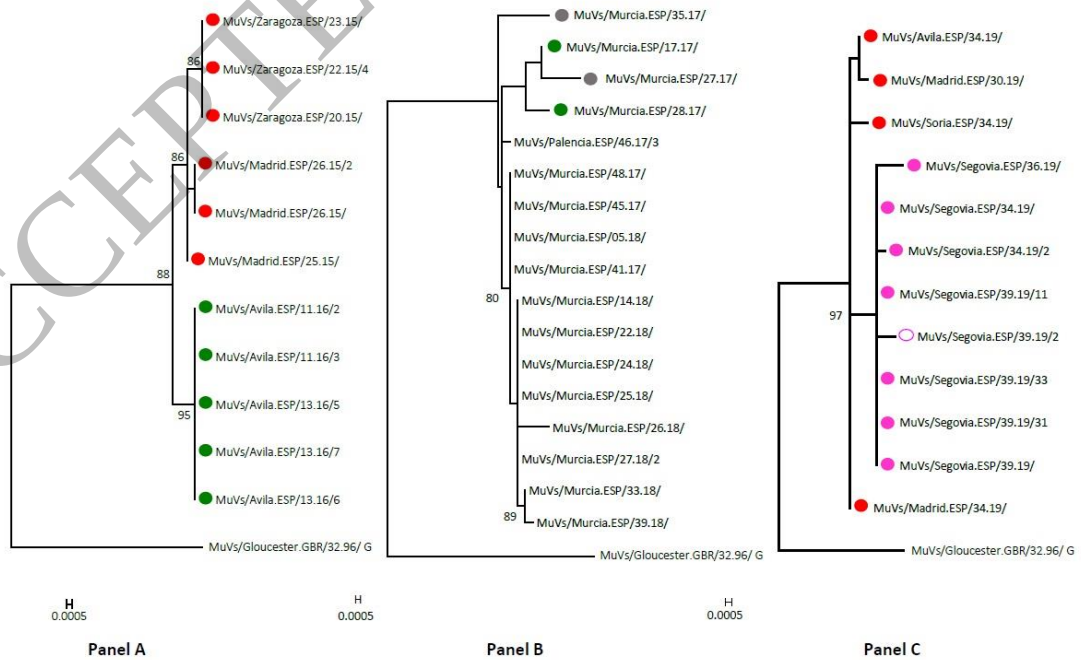

Figure 5
150x101 mm (.25 x DPI)