Running title: First complete genome Measles virus Brazil

Title: Complete genome of a Measles virus from Roraima state, Brazil

Cátia Alexandra Ribeiro Meneses[1#], Valdinete Alves do Nascimento[2#], Victor Costa de Souza[2#], Rodrigo Melo Maito[1], Marconi Aragão Gomes[1], Claudeth Rocha Santa Brígida Cunha[1], Ilma de Aguiar Antony[1], Maria Eliane Oliveira e Silva[1], Daniela Palha de Souza Campos[3], André de Lima Guerra Corado[2], Karina Pinheiro Pessoa[2], Dana Cristina da Silva Monteiro[2], Osnei Okumoto[4], Marília Coelho Cunha[5], Flávia Caselli Pacheco[5], Felipe Gomes Naveca[2]

[#] contributed equally to this work.

[1] Laboratório Central de Saúde Pública de Roraima, Secretaria de Estado da Saúde de Roraima, Boa Vista, Roraima, Brasil.

[2] Instituto Leônidas e Maria Deane, Fiocruz, Manaus, Amazonas, Brasil.

[3] Coordenadoria Geral de Vigilância em Saúde, Secretaria de Estado da Saúde de Roraima, Boa Vista, Roraima, Brasil.

[4] Secretaria de Vigilância em Saúde - SVS, Ministério da Saúde, Brasília, Distrito Federal, Brasil.

[5] Coordenação Geral de Laboratórios de Saúde Pública – CGLAB, SVS, Ministério da Saúde, Brasília, Distrito Federal, Brasil.

[+]Address correspondence to Felipe Gomes Naveca, felipe.naveca@fiocruz.br

SUMMARY

Measles is a human infectious disease of global concern caused by the measles virus. In this study, we report the complete genome sequencing of one measles isolate, genotype D8, obtained in Boa Vista city, the capital of the Roraima State, Brazil, directly from the urine sample. Phylogenetics reconstruction grouped the genome described in this study with samples from Australia, Italy, United Kingdom, and the USA. To our knowledge, this is the first complete genome of a wild-type measles virus from Latin America. Therefore, the present data strengthens the current knowledge about the molecular epidemiology of measles worldwide.

Key words: Measles virus; Brazil; Genotype; Genome, Viral.

Sponsorships: CNPq / CAPES / MS-DECIT / Fiocruz

Measles is a highly contagious airborne disease that begins as an acute febrile illness with fever, coryza, conjunctival hyperemia and cough, which are accompanied by a maculopapular skin rash. Some measles patients, mainly those immunocompromised, may evolve to severe complications including blindness and life-threating forms as severe diarrhea and pneumonia. Two highly severe, but rare neurologic forms, are also related to measles virus infection, the measles inclusion-body encephalitis MIBE and the subacute sclerosing panencephalitis (SSPE) (1).

The measles virus is an enveloped, single-stranded, negative-sense RNA virus, with a genome around 16Kb that belongs to the *Paramyxoviridae* family, genus *Morbillivirus.* A total of eight proteins, six structural (nucleoprotein, phosphoprotein, matrix, fusion, haemagglutinin, and the large polymerase protein) and two non-structural (V and C proteins) are encoded by the viral genome (1). Due to the lack of proofreading activity of RNA polymerases and the higher viral titers achieved during the acute phase of illness, RNA viruses are more likely to accumulate variability in their genome's sequences, in comparison to DNA viruses (2). Regarding the measles virus, this diversity is reflected in 24 genotypes known up to date (A, B1, B2, B3, C1, C2, D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, E, F, G1, G2, G3, H1, and H2), but from October 2017 to September 2018, only five genotypes (B3, D4, D8, D9, and H1) were reported (http://www.who-measles.org/Public/Data_Mnt/who_map.php).

Measles is a vaccine-preventable disease which has decreased mortality over the last decades. Since the 1980s, the total number of deaths related to measles illness dropped from more than 2 million to approximately 100,000 per year due to the improvement of social indicators (e.g., better nutrition), as well as the establishment of the global efforts to increase vaccination coverage (1). Nevertheless, timely surveillance of suspected measles cases with highly sensitive and specific molecular diagnostic tools, together with genetic characterization of isolates, is of paramount importance for the global efforts of virus elimination (1, 3).

In August 2018, the Central Laboratory of the Roraima State (LACEN-RR) started testing measles suspected samples using the Real-Time PCR protocol developed by CDC-USA. Procedures followed the annex 6.2 of the Manual for the Laboratory-based

Surveillance of Measles, Rubella, and Congenital Rubella Syndrome (Third edition, June 2018) freely available at http://www.who.int/immunization/monitoring_surveillance/burden/laboratory/Annex_6.2.pdf?ua=1. Since then, 172 samples were tested, and 39 were positive.

Thus, as a request of both Roraima's state and the Brazilian health surveillance authorities, we select one positive sample with high viral load (Ct 22.0) to submit for a protocol for the entire genome amplification and nucleotide sequencing. Initially, we aligned all measles D8 genomes with MAFFT v7.388 (4). The consensus sequence was used to design primers over the entire genome of measles D8 with the aid of Primal software (5) spanning around 400bp distances. Two other primers targeting the initial 28 bases of the 5' UTR region and the final 25 bases of the 3' UTR were designed with a modified version of Primer3 v2.3.7 embedded in Geneious software v10.2.6 (6). This primer design may be used for NGS sequencing, as previously described for Zika virus (5), or capillary sequencing as we performed in this study. Primer sequences and the details for the RT-PCR scheme used for entire genome amplification and nucleotide sequencing are in the supplemental files 1 and 2 of this manuscript.

Briefly, four overlapping amplicons encompassing the entire measles genome were generated using Superscript IV One-Step RT-PCR System (ThermoFisher Scientific). Amplicons were precipitated with molecular biology grade PEG8000 and used as a template for nucleotide sequencing with BigDye terminator v3.1 in an ABI 3130 genetic analyzer.

A total of 63 trace files were trimmed for quality and used for assembly employing Geneious's map to reference tool and the NCBI measles reference sequence, available in

GenBank under accession number NC_001498.1. The final whole genome sequence of MVs/Roraima.BRA/31.18[D8] isolate contains 15,894 nucleotides with no ambiguity, or unidentified (N) bases, a Q40 score of 99.8% and a GC content of 47.4 %.

Initially, the sequence reported here was genotyped using both the Nucleoprotein (N) gene and the Hemagglutinin (H) gene with the measles genotyping tool available at http://www.who-measles.org/Public/Tool/genotype_tool.php, and the D8 genotype was confirmed. Subsequently, we conducted a maximum likelihood phylogenetic analysis with PhyML (7) using the N gene region of the Roraima sample, the genotypes reference sequences available at MeaNS http://www.who-measles.org/ and four other sequences representing the genotype D8 lineages (FJ765078.1 – MVi/Villupuram.Ind/03.07; JX486001.1 - MVi/Hulu-Langat.MYS/26.11; KF683445.1 - MVs/Frankfurt_Main.DEU/17.11; KT588030.1 - MVs/Republic_of_Komi.RUS/35.13). This approach confirmed that the Roraima sample belongs to the lineage MVi/Hulu-Langat.MYS/26.11 (Supplemental file 3).

Secondly, a BLAST search was conducted using the megablast algorithm against the entire NR database. The first five closest matches are samples from Brisbane, Australia (MH638233, score 29137, 15834/15862 identical sites); Ancona, Italy (MH173047, score 29000, 15816/15872 identical sites); California, USA (KY969480, score 28755, 15773/15874 identical sites); Virginia, USA (JN635404, score 28731, 15782/15894 identical sites) and London, England (KT732231, score 28476, 15592/15678 identical sites).

Finally, a dataset containing all 42 complete measles genomes belonging to the genotype D8 available in GenBank in 06-Nov- 2018, and the sequence obtained in this

study, were analyzed with jModeltest 2.1.7 v20150220 and likelihood scores were computed for five substitutions schemes (40 models). The GTR+I model was selected by AICc and used for Bayesian phylogenetic reconstruction using two runs with 20mi generations with MrBayes 3.2.6 (8). This analysis showed two main clades, one containing 31 sequences from the United Kingdom 2012-2013, one from Netherlands (2013), one from Canada (2010), one from Germany (2013), two from Vietnam (2014), and one sequence from Texas, USA (2007), as the outgroup. The second clade contains the same five sequences returned from the BLAST analysis with the highest scores, from Italy (2017), Australia (2015), United Kingdom (2012), and two from USA, California (2010) and Virginia (2009), been this last one the outgroup of this clade. Furthermore, the Bayesian tree showed that the isolate MVs/Roraima.BRA/31.18[D8] grouped in this second clade, close together with the sample from Brisbane, Australia, collected in 2015 and the sample from Ancona, Italy collected in November 2017, with a high posterior probability (1.0) (Figure).

Surprisingly, we did not find any measles complete genome record from Latin America on the public databases including GenBank, Virus Pathogen Resource (ViPR - www.viprbrc.org), and MeaNS. This fact hindered a more detailed analysis of the complete genome reported in this study in the context of the local transmission, strengthening the necessity of sequencing more fully genomes from this region.

Previous studies in Brazil have primarily concentrated on genotyping a small fragment of 450bp of the measles nucleoprotein (9,10). Although, these studies have undoubtedly contributed to the molecular epidemiology of measles, sequencing larger genome fragments or better yet, complete viral genomes, is pivotal for a better

understanding of the epidemic dynamics of any emerging or remerging viral diseases (11,12).

Nucleotide sequence accession number:

The complete genome sequence of the MVs/Roraima.BRA/31.18[D8] isolate is available at GenBank under the accession number MK161348.

ACKNOWLEDGMENTS

FGN is funded by Fundação de Amparo à Pesquisa do Estado do Amazonas – FAPEAM (www.fapeam.am.gov.br, call 001/2013 – PPSUS / 062.00656/2014 and call 001/2014 – PROEP / 062.01939/2014); Conselho Nacional de Desenvolvimento Científico e Tecnológico (http://www.cnpq.br, grant 440856/2016-7) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (http://www.capes.gov.br, grants 88881.130825/2016-01 and 88887.130823/2016-00) call MCTIC/FNDCT -CNPq / MEC-CAPES/ MS-Decit 14/2016 – Prevenção e Combate ao vírus Zika. The authors thank the Program for Technological Development in Tools for Health-PDTIS FIOCRUZ for use of nucleotide sequencing facilities at ILMD—Fiocruz Amazônia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Moss WJ. Measles. Lancet. 2017; 390(10111): 2490-2502.
2. Steinhauer DA, Holland JJ. Rapid evolution of RNA viruses. Annu Rev Microbiol. 1987; 41: 409-33.
3. Rota PA, Moss WJ, Takeda M, de Swart RL, Thompson KM, Goodson JL. Measles. Nat Rev Dis Primers. 2016; 2: 16049.
4. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30(4): 772-80.
5. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, Burton DR, Lewis-Ximenez LL, de Jesus JG, Giovanetti M, Hill SC, Black A, Bedford T, Carroll MW, Nunes M, Alcantara LC Jr, Sabino EC, Baylis SA, Faria NR, Loose M, Simpson JT, Pybus OG, Andersen KG, Loman NJ. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017; 12(6): 1261-1276.
6. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28(12): 1647-9.
7. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59(3): 307-21.

8. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012; 61(3): 539-42.
9. Oliveira MI, Afonso AM, Adelaide CF, Lemos XR, Almeida J, Frugis Yu AL, Curti SP. Genetic diversity of measles virus-resurgence of new genotype D8 in São Paulo, Brazil. Rev Inst Med Trop Sao Paulo. 2014; 56(4): 366.
10. Oliveira MI, Rota PA, Curti SP, Figueiredo CA, Afonso AM, Theobaldo M, Souza LT, Liffick SL, Bellini WJ, Moraes JC, Stevien KE, Durigon EL. Genetic homogeneity of measles viruses associated with a measles outbreak, São Paulo, Brazil, 1997. Emerg Infect Dis. 2002; 8(8): 808-13.
11. Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, Park DJ, Ladner JT, Arias A, Asogun D, Bielejec F, Caddy SL, Cotten M, D'Ambrozio J, Dellicour S, Di Caro A, Diclaro JW, Duraffour S, Elmore MJ, Fakoli LS, Faye O, Gilbert ML, Gevao SM, Gire S, Gladden-Young A, Gnirke A, Goba A, Grant DS, Haagmans BL, Hiscox JA, Jah U, Kugelman JR, Liu D, Lu J, Malboeuf CM, Mate S, Matthews DA, Matranga CB, Meredith LW, Qu J, Quick J, Pas SD, Phan MVT, Pollakis G, Reusken CB, Sanchez-Lockhart M, Schaffner SF, Schieffelin JS, Sealfon RS, Simon-Loriere E, Smits SL, Stoecker K, Thorne L, Tobin EA, Vandi MA, Watson SJ, West K, Whitmer S, Wiley MR, Winnicki SM, Wohl S, Wölfel R, Yozwiak NL, Andersen KG, Blyden SO, Bolay F, Carroll MW, Dahn B, Diallo B, Formenty P, Fraser C, Gao GF, Garry RF, Goodfellow I, Günther S, Happi CT, Holmes EC, Kargbo B, Keïta S,

Kellam P, Koopmans MPG, Kuhn JH, Loman NJ, Magassouba N, Naidoo D, Nichol ST, Nyenswah T, Palacios G, Pybus OG, Sabeti PC, Sall A, Ströher U, Wurie I, Suchard MA, Lemey P, Rambaut A. Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature. 2017; 544(7650): 309-315.

12. Faria NR, Kraemer MUG, Hill SC, Goes de Jesus J, Aguiar RS, Iani FCM, Xavier J, Quick J, du Plessis L, Dellicour S, Thézé J, Carvalho RDO, Baele G, Wu CH, Silveira PP, Arruda MB, Pereira MA, Pereira GC, Lourenço J, Obolski U, Abade L, Vasylyeva TI, Giovanetti M, Yi D, Weiss DJ, Wint GRW, Shearer FM, Funk S, Nikolay B, Fonseca V, Adelino TER, Oliveira MAA, Silva MVF, Sacchetto L, Figueiredo PO, Rezende IM, Mello EM, Said RFC, Santos DA, Ferraz ML, Brito MG, Santana LF, Menezes MT, Brindeiro RM, Tanuri A, Dos Santos FCP, Cunha MS, Nogueira JS, Rocco IM, da Costa AC, Komninakis SCV, Azevedo V, Chieppe AO, Araujo ESM, Mendonça MCL, dos Santos CC, dos Santos CD, Mares-Guia AM, Nogueira RMR, Sequeira PC, Abreu RG, Garcia MHO, Abreu AL, Okumoto O, Kroon EG, de Albuquerque CFC, Lewandowski K, Pullan ST, Carroll M, de Oliveira T, Sabino EC, Souza RP, Suchard MA, Lemey P, Trindade GS, Drumond BP, Filippis AMB, Loman NJ, Cauchemez S, Alcantara LCJ, Pybus OG. Genomic and epidemiological monitoring of yellow fever virus transmission potential. Science. 2018; 361(6405): 894-899.

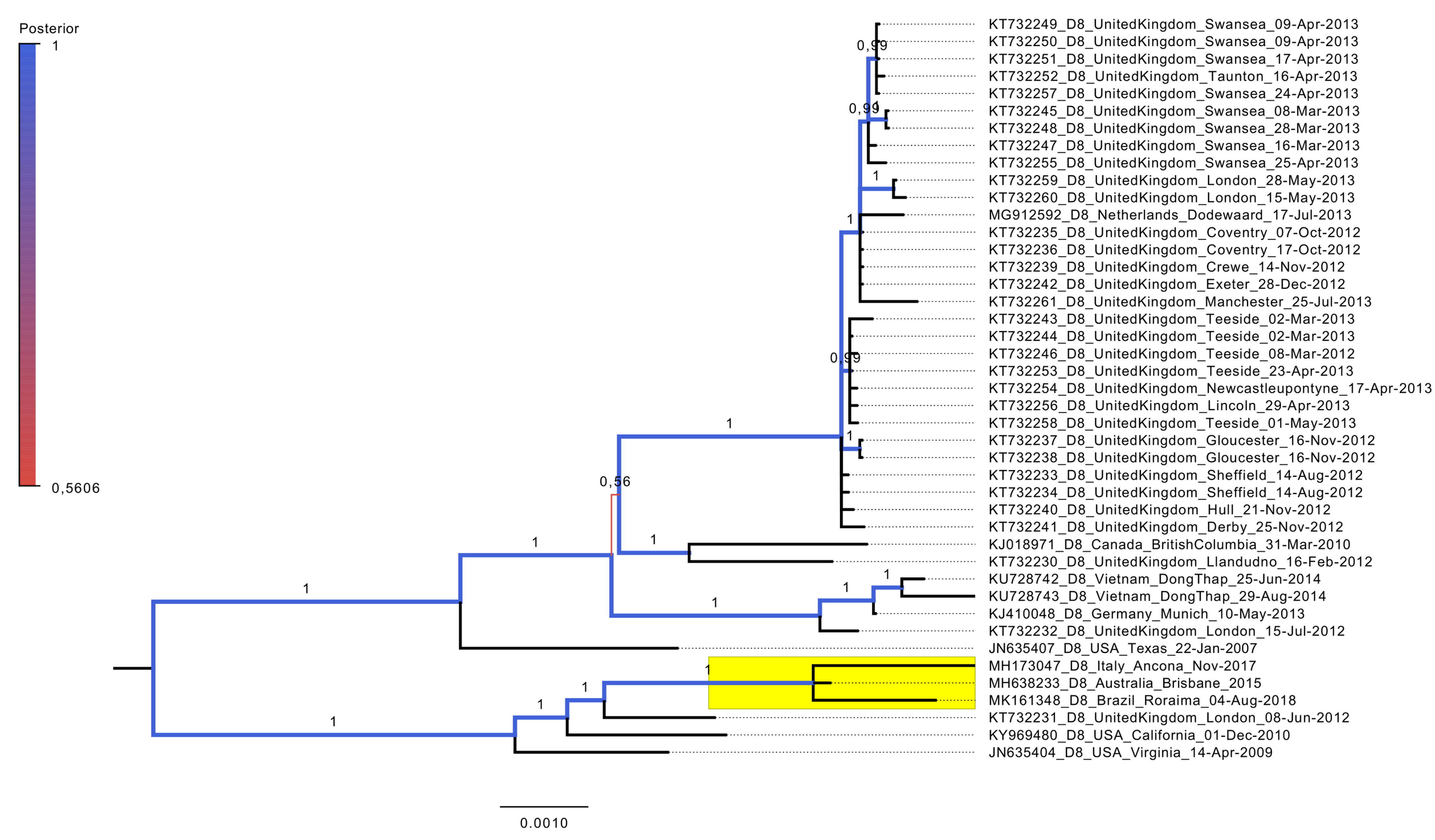
Figure: Phylogenetic tree of complete measles virus genomes. A mid-rooted Bayesian tree with increasing node order was constructed with MrBayes 3.2.6 and 43 taxa (complete CDS and intergenic regions from position 108 to 15,785 related to the reference sequence NC_001498.1) representing all the complete measles (genotype D8) genomes available at GenBank in 06-Nov-2018 and the sequence reported in this study. Branches are colored by posterior probabilities, according to the legend, and the specific posterior probabilities values are shown. The clade containing the sequence described in this study is highlighted in yellow. The scale bar represents nucleotide substitutions per site.

Supplemental file 1. Primers designed to amplify and to sequence the entire genome of measles Genotype D8.

Supplemental file 2. Primers sets used to amplify and to sequence the entire measles genome. Four amplicons were generated: Amplicon 1 (primers 5UTR + 11R, 3614bp); Amplicon 2 (11L + 26R, 5003bp); Amplicon 3 (24L + 40R, 5405bp); Amplicon 4 (37L + 3UTR, 4682bp). Each amplicon was sequenced with the primers listed below each scheme.

Supplemental file 3. Phylogenetic tree of measles genotypes based in the N gene. A mid-rooted ML tree with increasing node order was constructed with PhyML online server available at http://www.atgc-montpellier.fr/phyml/. All 24 genotypes and the four

lineages of genotype D8 are represented. The D8 sequences are grouped in the blue clade, whereas the clade containing the Roraima sample is highlighted in yellow.


Posterior
1
0,5606
KT732249_D8_UnitedKingdom_Swansea_09-Apr-2013
KT732250_D8_UnitedKingdom_Swansea_09-Apr-2013
KT732251_D8_UnitedKingdom_Swansea_17-Apr-2013
KT732252_D8_UnitedKingdom_Taunton_16-Apr-2013
KT732257_D8_UnitedKingdom_Swansea_24-Apr-2013
KT732245_D8_UnitedKingdom_Swansea_08-Mar-2013
KT732248_D8_UnitedKingdom_Swansea_28-Mar-2013
KT732247_D8_UnitedKingdom_Swansea_16-Mar-2013
KT732255_D8_UnitedKingdom_Swansea_25-Apr-2013
KT732259_D8_UnitedKingdom_London_28-May-2013
KT732260_D8_UnitedKingdom_London_15-May-2013
MG912592_D8_Netherlands_Dodewaard_17-Jul-2013
KT732235_D8_UnitedKingdom_Coventry_07-Oct-2012
KT732236_D8_UnitedKingdom_Coventry_17-Oct-2012
KT732239_D8_UnitedKingdom_Crewe_14-Nov-2012
KT732242_D8_UnitedKingdom_Exeter_28-Dec-2012
KT732261_D8_UnitedKingdom_Manchester_25-Jul-2013
KT732243_D8_UnitedKingdom_Teeside_02-Mar-2013
KT732244_D8_UnitedKingdom_Teeside_02-Mar-2013
KT732246_D8_UnitedKingdom_Teeside_08-Mar-2012
KT732253_D8_UnitedKingdom_Teeside_23-Apr-2013
KT732254_D8_UnitedKingdom_Newcastleupontyne_17-Apr-2013
KT732256_D8_UnitedKingdom_Lincoln_29-Apr-2013
KT732258_D8_UnitedKingdom_Teeside_01-May-2013
KT732237_D8_UnitedKingdom_Gloucester_16-Nov-2012
KT732238_D8_UnitedKingdom_Gloucester_16-Nov-2012
KT732233_D8_UnitedKingdom_Sheffield_14-Aug-2012
KT732234_D8_UnitedKingdom_Sheffield_14-Aug-2012
KT732240_D8_UnitedKingdom_Hull_21-Nov-2012
KT732241_D8_UnitedKingdom_Derby_25-Nov-2012
KJ018971_D8_Canada_BritishColumbia_31-Mar-2010
KT732230_D8_UnitedKingdom_Llandudno_16-Feb-2012
KU728742_D8_Vietnam_DongThap_25-Jun-2014
KU728743_D8_Vietnam_DongThap_29-Aug-2014
KJ410048_D8_Germany_Munich_10-May-2013
KT732232_D8_UnitedKingdom_London_15-Jul-2012
JN635407_D8_USA_Texas_22-Jan-2007
MH173047_D8_Italy_Ancona_Nov-2017
MH638233_D8_Australia_Brisbane_2015
MK161348_D8_Brazil_Roraima_04-Aug-2018
KT732231_D8_UnitedKingdom_London_08-Jun-2012
KY969480_D8_USA_California_01-Dec-2010
JN635404_D8_USA_Virginia_14-Apr-2009
0,99
0,99
1
1
0,99
1
1
0,56
1
1
1
1
1
1
1
1
1
1
1
0.0010