

The role of extended and whole genome sequencing for tracking transmission of measles and rubella viruses: report from the Global Measles and Rubella Laboratory Network meeting, 2017

The lack of an endemic genotype of measles and rubella is an essential criterion for verification of elimination of disease transmission.¹ Molecular surveillance for measles and rubella is supported by the WHO Global Measles and Rubella Laboratory Network (GMRLN) as a means to characterize outbreaks and track transmission of these viruses. GMRLN has defined the genomic target sequences used to genotype measles and rubella viruses, which are submitted to the 2 global databases, MeaNS and RubeNS,² maintained by Public Health England. The genetic characterization of viruses has proven to be a very powerful approach in tracking the global circulation of measles and rubella, and in documenting elimination or continuing endemic transmission.

For measles virus, GMRLN recommends sequencing of the 450 nucleotide (nt) coding for the COOH terminal 150 amino acids of the nucleoprotein (N450) as the primary target for genotyping, and the entire coding region of the hemagglutinin (H) gene (1854 nt) as the secondary target (*Figure 1a*). These targets identify 24 measles genotypes (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). The standard target for genotyping rubella virus is a 739 nt sequence within the E1 protein (*Figure 1b*). Rubella viruses are classified into 2 clades: clade 1 and clade 2, which include 12 genotypes (1B, 1C, 1D, 1E, 1F, 1G, 1H, 1I, 1J, 2A, 2B and 2C) and 1 provisional genotype (1a).³

Increases in vaccination coverage and the elimination of endemic transmission from some countries and regions have considerably decreased both the number of measles cases and the diversity of the circulating measles viruses.⁴ During 2016–2017, only 5 measles genotypes were detected (B3, D4, D8, D9 and H1). Characterization of defined sequence variants, known as “named strains”, is used to track transmission patterns for viruses within a single genotype. Named strains represent epidemiologically significant sequences that are circulating in multiple countries. Circulating strains can be linked to named strains based on the identity of the N450 sequence.⁵ This allows for a finer mapping of measles transmission than the use of genotyping alone. However, in countries with low numbers of linked cases, or where measles is repeatedly imported from the ongoing large outbreaks, the sequences of the standard N450 target vary by no more than 1 nt. Therefore, N450 sequences lose some of their ability to discriminate between continuing endemic

Rôle du séquençage génomique étendu et complet pour suivre la transmission des virus rougeoleux et rubéoleux: rapport de la réunion du Réseau mondial de laboratoires de la rougeole et de la rubéole, 2017

L'absence de génotype endémique des virus de la rougeole et de la rubéole constitue un critère essentiel de vérification de l'élimination de la transmission de ces maladies.¹ La surveillance moléculaire de la rougeole et de la rubéole, appuyée par le Réseau mondial OMS de laboratoires de la rougeole et de la rubéole (GMRLN), permet de caractériser les flambées et de suivre la transmission des virus. Le GMRLN a défini les séquences génomiques cibles qui sont utilisées pour le génotypage des virus rougeoleux et rubéoleux; les séquences sont soumises dans 2 bases de données mondiales gérées par Public Health England: MeaNS et RubeNS.² La caractérisation génétique des virus est une approche qui s'est avérée très efficace pour suivre la circulation mondiale des virus rougeoleux et rubéoleux et rendre compte de l'élimination ou de la persistance de la transmission endémique.

Pour le virus de la rougeole, le GMRLN recommande de séquencer les 450 nucléotides codant pour les 150 acides aminés de l'extrémité COOH de la nucléoprotéine (N450) comme cible principale du génotypage, et la région complète du gène codant pour l'hémagglutinine (H) (1854 nucléotides) comme cible secondaire (*Figure 1a*). Ces cibles permettent d'identifier 24 génotypes de la rougeole (A, B1, B2, B3, C1, C2, D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, E, F, G1, G2, G3, H1 et H2). Pour le génotypage du virus de la rubéole, une séquence de 739 nucléotides de la protéine E1 constitue la cible standard (*Figure 1b*). Les virus rubéoleux sont classés dans 2 clades – clade 1 et clade 2 – comportant 12 génotypes (1B, 1C, 1D, 1E, 1F, 1G, 1H, 1I, 1J, 2A, 2B et 2C) et 1 génotype provisoire (1a).³

L'amélioration de la couverture vaccinale et l'élimination de la transmission endémique dans un certain nombre de pays et de régions ont entraîné une baisse considérable du nombre de cas de rougeole, ainsi que de la diversité des virus rougeoleux circulants.⁴ Dans la période 2016–2017, seuls 5 génotypes rougeoleux ont été détectés (B3, D4, D8, D9 et H1). Des variants de séquences définis, appelés «souches désignées», sont caractérisés pour suivre le profil de transmission des virus d'un génotype donné. Les souches désignées représentent les séquences épidémiologiquement significatives circulant dans plusieurs pays. Le lien entre les souches circulantes et les souches désignées peut être établi sur la base de l'identité de la séquence N450.⁵ On peut ainsi dresser une cartographie plus précise de la transmission de la rougeole qu'avec le seul génotypage. Cependant, dans les pays qui n'ont qu'un faible nombre de cas liés ou qui subissent une importation répétée de virus rougeoleux provenant de flambées de grande ampleur, les séquences de la cible standard N450 varient au maximum de 1 nucléotide. Ces séquences N450 ont dès lors une moindre capacité à distinguer une transmission

¹ See No. 9, 2013, pp. 89–98.

² See No. 32, 2013, pp. 337–343.

³ See No. 3, 2013, pp. 29–36.

⁴ See http://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en/

⁵ See No. 30, 2015, pp. 373–380.

¹ Voir N° 9, 2013, pp. 89-98.

² Voir N° 32, 2013, pp. 337-343.

³ Voir N° 3, 2013, pp. 29-36.

⁴ Voir http://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en/.

⁵ Voir N° 30, 2015, pp. 373-380.

transmission and multiple importations of the same viral lineage from measles endemic regions.

Sequence analysis of a larger fragment of the measles genome will enhance the resolution of the molecular epidemiologic analysis for measles. In addition to the N450 and H gene targets, extended sequencing, including whole genome sequencing (WGS), provides more robust sequence data. To evaluate the utility and scope of extended and whole genome sequencing, GMRLN formed the “Next generation, Extended window and Whole genome sequencing working group” (N.E.W.). This working group met in Berlin in March 2016, and in Geneva in June 2016 and July 2017, to evaluate the available technology and to define how it could be used to improve the resolution of molecular surveillance of measles and rubella viruses. This report is a summary of the main conclusions of N.E.W.

There are many reports which demonstrate the utility of N450 genotyping in the analysis of outbreaks and for verification of elimination, including how MeaNS can be used to track the worldwide distribution of the variants. WHO periodically publishes a map showing the global distribution of measles and rubella genotypes.⁶

Recent studies, and unpublished results from several laboratories of the GMRLN, clearly show that extended window or WGS enhances the resolution of measles outbreak investigations. For example, analysis of WGS and the sequences of the hypervariable, noncoding region between the matrix (M) and fusion (F) genes (MF-NCR) improved the resolution of the circulation patterns of genotype D8 viruses in the United Kingdom.⁷ Based on the homogenous clustering of the WGS, the authors concluded that all cases were due to local transmission rather than separate importations. In addition, sequencing of the MF-NCR region suggested resolution into separate clusters of 2 outbreaks associated with genotypes D8 and B3 in Sweden,⁸ which was consistent with the epidemiological data. Two reports showed that WGS could help track outbreaks due to multiple genotypes of measles in British Columbia, Canada in 2010,⁹ and the sequences of the MF-NCR showed that a cluster of measles cases in Ontario, Canada were part of the same chain of transmission despite the lack of epidemiological linkage.¹⁰

The consensus of N.E.W. was that sequencing the N450 region should continue to be the standard method for routine genotyping of measles viruses. However, extended window or WGS sequencing should be considered in countries and regions in which measles has been eliminated or in countries near to elimination. In these countries, outbreaks with the same recurring N450 variant

endémique persistante d’une importation répétée de la même lignée virale en provenance de régions d’endémie de la rougeole.

L’analyse de la séquence d’un fragment plus long du génome rougeoleux permet d’améliorer la résolution de l’analyse épidémiologique moléculaire de la rougeole. Le séquençage étendu au-delà des cibles génétiques N450 et H, y compris le séquençage du génome complet, fournit des informations plus fiables sur les séquences. Pour évaluer l’utilité et le champ d’application du séquençage étendu et complet, le GMRLN a formé un groupe de travail sur le séquençage de nouvelle génération, le séquençage étendu et le séquençage du génome complet (N.E.W, de l’anglais «Next generation, Extended window and Whole genome sequencing»). Ce groupe de travail s’est réuni à Berlin en mars 2016, puis à Genève en juin 2016 et en juillet 2017, afin d’évaluer les technologies existantes et de déterminer comment elles pourraient être employées pour améliorer la résolution de la surveillance moléculaire des virus rougeoleux et rubéoleux. Le présent rapport récapitule les principales conclusions du groupe de travail N.E.W.

De nombreux articles témoignent de l’utilité du génotypage de la séquence N450 pour l’analyse des flambées et la vérification de l’élimination, montrant notamment comment la base de données MeaNS peut être utilisée pour suivre la répartition mondiale des variants. L’OMS publie régulièrement une carte illustrant la répartition mondiale des différents génotypes rougeoleux et rubéoleux.⁶

De récentes études, ainsi que les résultats non publiés de plusieurs laboratoires du GMRLN, démontrent clairement que le séquençage étendu ou complet du génome accroît la résolution des enquêtes sur les flambées de rougeole. Par exemple, au Royaume-Uni, l’analyse des séquences du génome complet et de la région hypervariable non codante entre le gène matriciel (M) et le gène chimère (F) (région MF-NCR) a permis d’identifier le profil de circulation du génotype des virus D8 avec une résolution accrue.⁷ Sur la base d’un regroupement homogène (ou «clustering») des séquences du génome complet, les auteurs ont conclu que tous les cas étaient imputables à une transmission locale, plutôt qu’à des importations distinctes. En Suède, un séquençage de la région MF-NCR a en outre indiqué une résolution dans des grappes distinctes de 2 flambées associées aux génotypes D8 et B3,⁸ un résultat qui est compatible avec les données épidémiologiques. Au Canada, deux rapports ont montré que le séquençage du génome complet était susceptible de faciliter le suivi de flambées occasionnées par plusieurs génotypes du virus de la rougeole en Colombie-Britannique en 2010⁹ et que les séquences de la région MF-NCR indiquaient qu’une grappe de cas de rougeole en Ontario relevait de la même chaîne de transmission, malgré l’absence de lien épidémiologique.¹⁰

De l’avis unanime du groupe de travail N.E.W, le séquençage de la région N450 doit rester la méthode standard employée pour le génotypage de routine des virus rougeoleux, mais le séquençage étendu ou complet du génome devrait être envisagé dans les pays et régions où la rougeole a été éliminée, ainsi que dans les pays s’approchant du stade de l’élimination. Dans ces pays, il est souvent impossible de distinguer les flambées dues à un

⁶ See http://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en/

⁷ Penedos AR et al. *PLoS One*. 2015;10:e0143081.

⁸ Harvala H et al. *J Infect Dis*. 2016; 213:592–599.

⁹ Gardy JL et al. *J Infect Dis*. 2015;212:1574–1578.

¹⁰ Thomas S et al. *Emerg Infect Dis*. 2017; 23:1063–1069.

⁶ Voir http://www.who.int/immunization/monitoring_surveillance/burden/vpd/surveillance_type/active/measles_monthlydata/en/

⁷ Penedos AR et al. *PLoS One*. 2015;10:e0143081.

⁸ Harvala H et al. *J Infect Dis*. 2016; 213:592–599.

⁹ Gardy JL et al. *J Infect Dis*. 2015;212:1574–1578.

¹⁰ Thomas S et al. *Emerg Infect Dis*. 2017; 23:1063–1069.

often cannot be distinguished from ongoing endemic transmission or repeat importation events; therefore the use of extended sequencing may be useful in resolving the outbreaks.

Two options for sequencing measles virus were deemed useful in practice. The first is to sequence the MF-NCR (*Figure 1a*) which N.E.W defined as the region between the stop codon of the M gene and ending at the start codon of the F gene (i.e. between positions 4443 and 5460 of the prototype Edmonston strain, GenBank accession number AF266288). The MF-NCR usually includes 1018 nt, but the length may vary because some measles strains contain insertions or deletions in the MF-NCR.^{11, 12} It has been shown that the variability of this region reflects the diversity of the whole measles genome and it is therefore a very useful surrogate for WGS. Sequencing the MF-NCR requires only reverse transcription polymerase chain reaction amplification and standard Sanger sequencing from a clinical sample. If the MF-NCR sequence is not sufficiently diverse, the WGS, or a truncated WGS, can be obtained by Sanger sequencing or, more recently, by next generation sequencing. The sequence for the truncated WGS (WGS-t) starts at the start codon of the N gene and ends at the stop codon of the L gene (*Figure 1*). The genomic termini cannot be easily sequenced with next-generation or Sanger methods and the highly conserved sequences of the genomic termini do not contribute to the diversity of the measles genome. The MeaNS database now accepts submission of measles MF-NCR and WGS-t sequences.

Methods for WGS and sequencing the MF-NCR region have been developed but have not yet been standardized for routine application. Analysis of different viral genotypes may require reoptimization of the protocols and *ad hoc* primer design. While WGS from measles virus isolates has consistently produced good results, direct sequencing from clinical specimens is not always possible and may require an additional step of Sanger sequencing to fill the gaps. Since viral isolates are a better sample for extended sequencing and WGS, GMRLN will continue to rely on the capacity for virus isolation with the network. N.E.W. can provide the laboratories involved in measles and rubella surveillance with reference material and advice on the practical application of these techniques.

The database of MF-NCR sequences and WGS is limited and does not reflect the global distribution of measles strains. Reference sequences for the whole genome or the MF-NCR are available, but reference sequences for all genotypes and many of the currently circulating named strains are not yet available. To expand the database for MN-NCR sequences and WGS there is a need to collect adequate specimens from as many outbreaks as possible.

même variant N450 récurrent et les événements liés à une transmission endémique persistante ou à des importations répétées; le recours au séquençage étendu peut donc être utile à la résolution de ces flambées.

Deux possibilités de séquençage du virus de la rougeole ont été jugées utiles dans la pratique. La première consiste à séquencer la région MF-NCR (*Figure 1a*), que le groupe N.E.W. a définie comme étant la région allant du codon de terminaison du gène M jusqu'au codon d'initiation du gène F (c'est-à-dire entre les positions 4443 et 5460 de la souche prototype Edmonston, numéro d'accès GenBank AF266288). La région MF-NCR compte normalement 1018 nucléotides, mais sa longueur peut varier, car certaines souches du virus de la rougeole contiennent des insertions ou des délétions dans la région MF-NCR.^{11, 12} Il a été démontré que la variabilité de cette région reflète la diversité de l'ensemble du génome du virus rougeoleux et que l'analyse de cette région peut donc servir de substitut intéressant au séquençage du génome complet. Le séquençage de la région MF-NCR exige simplement une amplification en chaîne par polymérase après transcription inverse et un séquençage de Sanger standard à partir d'un échantillon clinique. Si la diversité de la séquence MF-NCR est insuffisante, le séquençage du génome complet, ou d'un génome tronqué, peut être réalisé par la méthode de Sanger ou par une méthode plus récente de séquençage de nouvelle génération. La séquence du génome tronqué commence au codon d'initiation du gène N et se termine au codon de terminaison du gène L (*Figure 1*). Les terminaisons génomiques ne sont pas aisément séquencées par les méthodes de Sanger ou de nouvelle génération et les séquences hautement conservées des terminaisons génomiques ne contribuent pas à la diversité du génome du virus rougeoleux. La base de données MeaNS accepte désormais les soumissions de séquences couvrant la région MF-NCR et le génome tronqué du virus de la rougeole.

Des méthodes de séquençage du génome complet et de la région MF-NCR ont été mises au point, mais n'ont pas encore été standardisées en vue d'une application de routine. L'analyse de différents génotypes viraux peut nécessiter une nouvelle optimisation des protocoles et la conception d'amorces *ad hoc*. Bien que le séquençage du génome complet ait systématiquement donné de bons résultats lorsqu'il est réalisé à partir d'isolements du virus de la rougeole, le séquençage direct à partir d'échantillons cliniques n'est pas toujours possible et peut nécessiter une étape supplémentaire de séquençage Sanger pour combler les lacunes. Étant donné que les isolements de virus constituent de meilleurs échantillons pour le séquençage étendu et complet, le GMRLN continuera de s'appuyer sur les capacités d'isolement des virus au sein du réseau. Le groupe de travail N.E.W. peut fournir des documents de référence et des conseils sur l'application pratique de ces techniques aux laboratoires qui participent à la surveillance de la rougeole et de la rubéole.

La base de données des séquences de la région MF-NCR et du génome complet est limitée et ne reflète pas la répartition mondiale des souches de virus rougeoleux. Certaines séquences de référence sont fournies pour le génome complet et la région MF NCR, mais on ne dispose pas encore de séquences de référence pour tous les génotypes, ni pour de nombreuses souches désignées qui circulent actuellement. Afin d'enrichir cette base de données, il faudra recueillir suffisamment d'échantillons provenant d'un nombre aussi important que possible de flambées.

¹¹ Bankamp B et al. PLoS One. 2014;9:e95470.

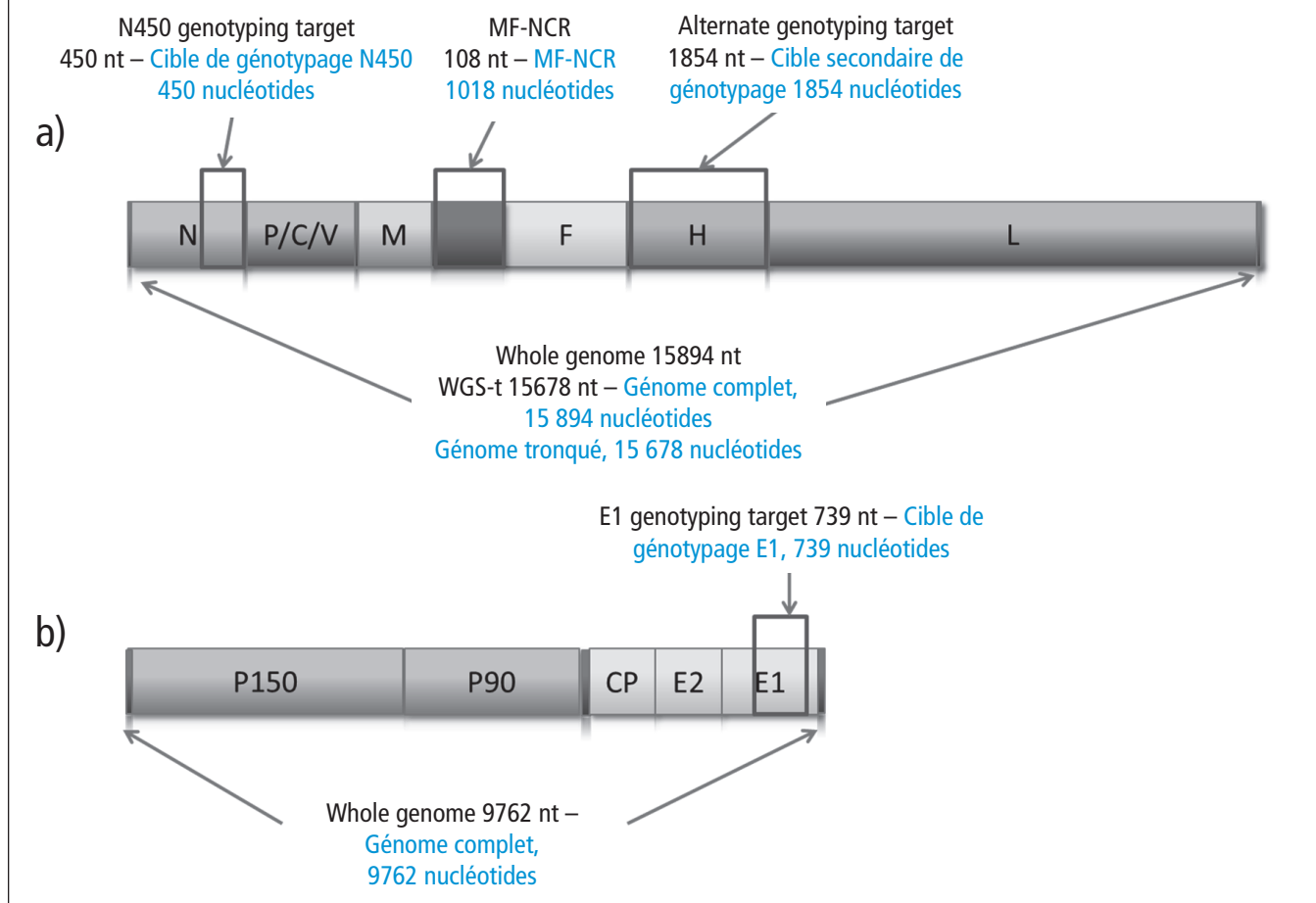
¹² Thomas S et al. Emerg Infect Dis. 2017; 23:1063–1069.

¹¹ Bankamp B et al. PLoS One. 2014;9:e95470.

¹² Thomas S et al. Emerg Infect Dis. 2017; 23:1063–1069.

Figure 1 **Schematic representation of the measles (Edmonston strain) (a) and rubella (RA-27 strain) (b) showing the targets for genotyping and sequencing discussed in this report**

Figure 1 **Représentation schématique des virus de la rougeole (souche Edmonston) (a) et de la rubéole (souche RA-27) (b), illustrant les cibles de génotypage et de séquençage évoquées dans ce rapport**



The current genotyping window for wild-type rubella virus is sufficient to track imported rubella virus. However, the window is not sufficient to differentiate closely related viruses, such as vaccine variants or viruses within an outbreak. Several commercial reagents are now available to pursue WGS of viral and other pathogens. Rubella virus WGS can be achieved by either metagenomics or targeted sequenc-

La fenêtre actuelle de génotypage des virus rubéoleux de type sauvage est suffisante pour suivre les importations du virus de la rubéole. Toutefois, elle est insuffisante pour distinguer les virus étroitement apparentés, comme les virus et les variants de la souche vaccinale dans le cadre d'une flambée. Plusieurs réactifs commerciaux sont aujourd'hui disponibles aux fins du séquençage complet du génome des agents pathogènes viraux ou autres. Le séquençage génomique complet du virus de la

预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=5_26005

