

Standardization of the nomenclature for genetic characteristics of wild-type rubella viruses

A number of genetically distinct groups of rubella viruses currently circulate in the world. The molecular epidemiology of rubella viruses will be useful primarily in rubella control activities for tracking transmission pathways, for documenting changes in the viruses present in particular regions over time and for documenting interruption of transmission of rubella viruses. Use of molecular epidemiology in control activities is now more relevant since the WHO European Region and the Region of the Americas have adopted congenital rubella infection prevention and elimination goals respectively, targeted for 2010. However, a systematic nomenclature for wild-type rubella viruses has been lacking, which has limited the usefulness of rubella virus genetic data in supporting rubella control activities.

A meeting was organized at WHO headquarters on 2–3 September 2004 to discuss standardization of a nomenclature for describing the genetic characteristics of wild-type rubella viruses. This report summarizes the outcomes of that meeting. The main goals of the meeting were to provide guidelines for describing the major genetic groups of rubella virus and to establish uniform genetic analysis protocols. In addition, development of a uniform

Normalisation de la nomenclature des caractéristiques génétiques des virus rubéoleux sauvages

Plusieurs groupes génétiquement distincts de virus rubéoleux circulent actuellement dans le monde. L'épidémiologie moléculaire des virus rubéoleux servira avant tout aux activités de lutte anti-rubéoleuse pour repérer les filières de transmission, documenter les modifications dans le temps des virus présents dans des régions particulières et documenter l'interruption de la transmission des virus. Le recours à l'épidémiologie moléculaire dans les activités de lutte est désormais plus utile car la Région européenne et la Région des Amériques de l'OMS ont fixé des objectifs respectivement pour la prévention et l'élimination de l'infection rubéoleuse congénitale pour 2010. Toutefois, une nomenclature systématique des virus rubéoleux de type sauvage fait défaut, ce qui a limité l'utilité des données génétiques sur les virus rubéoleux pour les activités de lutte contre la maladie.

Une réunion a été organisée au Siège de l'OMS les 2 et 3 septembre 2004 pour examiner la normalisation de la nomenclature visant à décrire les caractéristiques génétiques des virus rubéoleux sauvages. Le présent rapport fait le point des résultats de la réunion. La réunion visait principalement à définir des recommandations pour la description des principaux groupes génétiques de virus rubéoleux et à établir des protocoles standardisés pour l'analyse génétique. En outre, on a examiné la mise au point d'un système standar-

virus naming system, rubella virus strain banks and a sequence database were discussed. Standardization of a nomenclature should allow more efficient communication between the various laboratories conducting molecular characterization of wild-type strains and will provide public health officials with a consistent method for describing viruses associated with cases, outbreaks and epidemics.

Epidemiology and logistics

Rubella viral surveillance is an important part of rubella surveillance during the elimination and control phases; the goal should be to obtain a virus sequence from each chain of transmission during the elimination phase or from representative samples from each outbreak during the control phase. Attempts should be made to obtain specimens for virus isolation and/or sequencing at the same time that serum samples are obtained for serological diagnosis. Epidemiological data should accompany viral surveillance specimens in order to increase the usefulness of the molecular information. These data should include geographical location, date and type of specimen, date of rash onset, clinical presentation (postnatal rubella or congenital rubella syndrome (CRS)), age of patient, vaccine history and vaccination date.

WHO is developing a laboratory manual for measles and rubella that will include detailed instructions for obtaining, processing and analysing specimens for viral surveillance. It is imperative that appropriate epidemiological data be collected and linked to specimens.

Existing WHO measles and rubella global specialized laboratories and measles virus strain banks at the Centers for Disease Control and Prevention, Atlanta, United States, and the Health Protection Agency, London, England, will also maintain rubella virus strain banks. The functions of the strain banks will include: identification or validation of new genotype(s), reporting of genotype data and collection of strains. Information for candidate genotypes will be submitted directly to a strain bank for identification or validation.

Genotype information will be reported within the framework of the established global laboratory network, including timeliness indicators. Timeliness indicators for reporting rubella genotypes will be the same as those recommended for reporting of measles genotypes (within 3 months of obtaining a sequence(s)). Sequence information will be made available through the WHO strain banks, through web sites and via GenBank.¹ Updates to the recommended nomenclature will be published periodically in the *Weekly Epidemiological Record*. The WHO Global Measles and Rubella Laboratory Network will support efforts to obtain isolates of currently circulating wild-type rubella viruses and will facilitate referral to designated laboratories able to perform genetic analysis.

Sequencing and sequence analysis

Isolation of rubella virus from clinical specimens should be attempted, with the goal being to identify representative isolates from each chain of transmission during the elimi-

disé de désignation uniforme des virus, ainsi que l'établissement de banques de souches de virus rubéoleux et d'une base de données de séquences. La normalisation de la nomenclature devrait permettre une communication plus efficace entre les divers laboratoires qui caractérisent les souches sauvages d'un point de vue moléculaire et fournira aux responsables de la santé publique une méthode systématique pour décrire les virus associés aux cas, aux flambées et aux épidémies.

Epidémiologie et logistique

La surveillance des virus rubéoleux constitue un aspect important de la surveillance de la rubéole à chaque phase de la lutte et de l'élimination; le but doit être d'obtenir une séquence virale de chaque chaîne de transmission au cours de la phase d'élimination ou des échantillons représentatifs de chaque flambée au cours de la phase de lutte. Il faut s'efforcer d'obtenir des échantillons pour l'isolement et/ou le séquençage de virus en même temps que les prélèvements du sérum pour la sérologie. Les échantillons destinés à la surveillance virale de données épidémiologiques doivent être accompagnés afin de rendre les informations moléculaires plus utiles. Ces données doivent indiquer le lieu géographique, la date et le type de prélèvement, la date de début de l'éruption, le tableau clinique (rubéole postnatale ou syndrome de rubéole congénitale (SRC)), l'âge du malade, les antécédents vaccinaux et la date de la vaccination.

L'OMS met au point un manuel de laboratoire pour la rougeole et la rubéole qui comprendra des instructions détaillées pour obtenir, traiter et analyser les échantillons aux fins de la surveillance virale. Il est indispensable d'accompagner les échantillons des données épidémiologiques appropriées.

Les laboratoires spécialisés mondiaux existants de l'OMS pour la rougeole et la rubéole et les banques de souches de virus rougeoleux aux *Centers for Diseases Control and Prevention* à Atlanta (Etats-Unis) et à la *Health Protection Agency* à Londres (Royaume-Uni) maintiendront aussi des banques de souches de virus rubéoleux. Elles auront notamment pour fonction d'identifier ou de valider les nouveaux génotypes, de notifier les données génotypiques et de collecter les souches. Les informations concernant les éventuels génotypes seront soumises directement à une banque de souches pour identification ou validation.

L'information génotypique sera notifiée dans le cadre du réseau de laboratoires mondiaux avec notamment des indicateurs des délais à respecter. Ces indicateurs pour la notification des génotypes des virus rubéoleux seront les mêmes que ceux recommandés pour la notification des génotypes rougeoleux (3 mois à compter de l'obtention d'une ou de plusieurs séquences). L'information sur les séquences sera disponible par le biais des banques de souches de l'OMS, sur des sites Web et par l'intermédiaire de GenBank.¹ Des mises à jour de la nomenclature recommandée seront publiées périodiquement dans le *Relevé épidémiologique hebdomadaire*. Le réseau mondial de laboratoires de la rougeole et de la rubéole de l'OMS appuyera les efforts visant à obtenir des isolats des virus rubéoleux sauvages actuellement en circulation et facilitera l'acheminement vers des laboratoires désignés en mesure d'effectuer des analyses génétiques.

Séquençage et analyse des séquences

On essaiera d'isoler le virus de la rubéole à partir des prélèvements cliniques, dans le but d'identifier des isolats représentatifs de chacune des chaînes de transmission au cours de la phase d'élimi-

¹ GenBank is the genetic sequence database of the National Institutes of Health, Bethesda, Maryland, United States, for all publicly available DNA sequences.

¹ GenBank est la base de données de séquences génétiques pour toutes les séquences d'ADN accessibles au public des *National Institutes of Health*, Bethesda, Maryland (Etats-Unis).

nation phase or each outbreak during the control phase. Rubella virus isolation should be performed in Vero cells and should be confirmed by indirect fluorescent assay or reverse transcription (RT)-PCR, as the cytopathic effect is minimal. Rubella isolation can also be performed using Vero/SLAM cells in laboratories that are attempting to isolate wild-type measles viruses using these cells. Use of RT-PCR to amplify rubella sequences directly from clinical material is highly sensitive, and sequence information derived from the products can be used for molecular epidemiological studies. Protocols for isolation, direct RT-PCR, sequencing and sequence analysis will be available from WHO.

Most of the genetic studies of wild-type rubella viruses have been conducted by sequencing the full length or portions of the coding region of the E1 envelope protein. Sequencing other regions of the rubella virus genome for molecular epidemiology was discussed, but no compelling arguments for changing from the E1 region were identified. At present, different regions (windows) in the E1 coding region are used for genetic characterization. After evaluation of these windows, a window of 739 nucleotides (nts) – nts 8731–9469 – was recommended for routine molecular epidemiological analysis. This window results from combining 2 commonly used, overlapping sequence windows, designated 601 and 552 (see below).

Use of any one of several widely available computer programs for phylogenetic analysis is acceptable. For example, computer programs based on maximum likelihood, maximum parsimony or other methods are acceptable, provided the analysis is done using the set of accepted reference sequences. The use of reference viruses in an analysis will control for ambiguities in the assignment of viruses. Analysis must be performed with at least the recommended, minimum acceptable window (739 nts); the analysis will be considered valid if the accepted set of reference viruses falls into the accepted groups. Analysis of each data set with multiple phylogenetic analysis programs is desirable.

A phylogenetic program called MrBayes was used to generate the figures in this report. This program provides clade credibility values for the various nodes in phylogenetic trees; clade credibility values are analogous to bootstrap values, which can be obtained from maximum parsimony and maximum likelihood analyses.

The major phylogenetic groups of rubella viruses, which differ by 8–10% in nucleotide sequence, are easily separated and were designated as clades 1 and 2 at the meeting. This is a departure from the terminology previously used, but makes the rubella nomenclature closer to that used to describe the genetic characteristics of wild-type measles viruses. Reference viruses for 7 intraclade groups, called genotypes, were accepted and designated with upper-case letters (1B, 1C, 1D, 1E, 1F, 2A and 2B) (*Table 1, Fig. 1*). In addition, 3 provisional genotypes (1a, 1g and 2c) were discussed (*Table 1, Fig. 1*).

A provisional genotype will become an established genotype when reference viruses are obtained and phylogenetic relationships between the provisional genotype and other genotypes become clear. Genotype 1a was considered provisional because the phylogeny of this group of viruses was complex and poorly understood. Genotype 1a includes viruses from the 1960s (e.g. early vaccine strains), but this

nation ou de chacune des flambées pendant la phase de lutte. L'isolement du virus rubéoleux se fera sur cellules Vero et sera confirmé par fluorescence indirecte ou PCR transcriptase inverse (RT-PCR), l'effet cytopathique étant très peu important. L'isolement peut également être pratiqué au moyen de cellules Vero/SLAM dans les laboratoires qui isolent les virus rougeoleux sauvages au moyen de ces cellules. La RT PCR utilisée pour amplifier les séquences du virus rubéoleux directement à partir des prélèvements cliniques est très sensible, et les données concernant les séquences obtenues à partir de ces produits peuvent servir dans les études d'épidémiologie moléculaire. Des protocoles concernant l'isolement, la RT PCR directe, le séquençage et l'analyse des séquences seront disponibles auprès de l'OMS.

La plupart des études génétiques sur les virus rubéoleux sauvages ont été réalisées par séquençage de la totalité ou de portions de la région codante de la protéine d'enveloppe E1. Le séquençage d'autres régions du génome du virus rubéoleux pour l'épidémiologie moléculaire a été envisagé, mais aucun argument indiscutable n'indique qu'il faille séquencer d'autres régions que la région E1. Actuellement, différentes régions ou «fenêtres» de la région codante E1 sont utilisées pour la caractérisation génétique. Après évaluation de plusieurs de ces fractions, il a été recommandé d'examiner une «fenêtre» de 739 nucléotides (8731–9469) dans les analyses d'épidémiologie moléculaire de routine. Cette «fenêtre» est obtenue par l'association de 2 séquences couramment utilisées qui se recoupent, désignées par 601 et 552 (voir ci dessous).

N'importe lequel des programmes informatisés largement disponibles applicables à l'analyse phylogénétique est utilisable. Par exemple, les programmes informatiques qui reposent sur le maximum de vraisemblance, le maximum de parcimonie ou d'autres méthodes sont acceptables, à condition que l'analyse soit pratiquée sur la série de séquences de référence acceptées. L'utilisation de virus de référence dans l'analyse doit permettre d'éliminer les ambiguïtés de la catégorisation un virus. L'analyse doit être réalisée avec au moins la région minimale acceptable recommandée (739 nucléotides); elle sera considérée comme valable si la série de virus de référence acceptée correspond aux groupes acceptés. Il est souhaitable d'analyser les données au moyen de plusieurs programmes d'analyse phylogénétique.

Le programme utilisé pour obtenir les figures de ce rapport est appelé MrBayes. Ce programme donne des valeurs de la crédibilité des différentes branches à chaque noeud des arbres phylogénétiques; ces valeurs sont comparables aux valeurs données par la méthode des «bretelles», que l'on peut obtenir par l'analyse de la parcimonie maximale et de méthode du maximum de vraisemblance.

Les groupes phylogénétiques de virus rubéoleux les plus importants, qui diffèrent par 8 à 10% de leurs séquences nucléotidiques, ont été facilement distingués et désignés lors de la réunion sous le nom de clades 1 et 2. Cette terminologie s'écarte de celle qui était employée auparavant, mais rapproche la nomenclature du virus rubéoleux de celle qui est utilisée pour décrire les caractéristiques génétiques des virus rougeoleux sauvages. Des virus de référence ont été acceptés pour 7 groupes intra clades, appelés génotypes, et désignés par des lettres majuscules (1B, 1C, 1D, 1E, 1F, 2A et 2B) (*Tableau 1; Figure 1*). En outre, 3 génotypes provisoires (1a, 1g, 2c) ont été examinés (*Tableau 1, Figure 1*).

Un génotype provisoire devient établi lorsque l'on a obtenu des virus de référence et que les relations phylogénétiques entre le génotype provisoire et d'autres génotypes sont clairement définis. Le génotype 1a a été considéré comme provisoire, la phylogénie de ce groupe de virus étant complexe et mal connue. Le génotype 1a inclut des virus des années 60 (les premières souches vaccinales, par exemple), mais ce génotype a rarement été observé ces

Table 1 Reference strains to be used for genetic analysis of wild-type rubella viruses, 2004^aTableau 1 Souches de référence à utiliser pour l'analyse génétique des virus rubéoleux sauvages, 2004^a

Genotype – Génotype	Reference strains ^b – Souches de référence ^b	Reference strains ^c – Souches de référence ^c	Accession numbers – Numéros d'accèsion
1a	RVi/BEL/63	Cendehill BEL 63 ^d	AF188704
	RVi/Con.USA/61	HPV77 US 61 ^d	M30776
	RVi/Toyama.JPN/67	TO-336 WT JP 67 ^e	AB047330
1B	RVi/ISR/75[1B]	I-9 IS 75	AY968207
	RVi/ISR/88[1B]	I-34 IS 88	AY968209
	RVi/ISR/79[1B]	I-13 IS 79	AY968208
1C	RVi/Los Angeles.Cal.USA/91[1C]	BUR US 91	AY968212
	RVi/SLV/02[1C]	QUI ELS 02	AY968211
	RVi/PAN/99[1C]	P-31 PAN 99	AY968217
1D	RVi/Daly City.Cal.USA/97[1D]CRS	SAL-CA US 97	AY968206
	RVi/Tokyo.JPN/90[1D]CRS	NC JP 90	AY968214
	RVi/Saitama.JPN/94[1D]	SAI-1 JP 94	AY968216
1E	RVi/Shandong.CHN/02[1E]	T14 CH 02	AY968210
	RVi/MYS/01[1E]	M-1 MAL 01	AY968221
1F	RVi/Shandong.CHN/00[1F]	TS10 CH 00	AY968213
	RVi/Anhui.CHN/00[1F]	TS 38 CH 00	AY968215
2A	RVi/Beijing.CHN/79[2A]	BRD1 CH 79	AY258322
	RVi/Beijing.CHN/80[2A]	BRD2 CH 80 ^d	AY258323
2B	RVi/TelAviv.ISR/68[2B]	I-11 IS 68	AY968219
	RVi/Seattle.Wash.USA/16.00[2B]	TAN IND 00	AY968220
	RVi/Anhui.CHN/00/2[2B]	TS34 CH 00	AY968218
Vaccine ^f	RVi/USA/64	RA27/3 US 64 ^d	L78917 ^g

^a Sequence information for the structural protein (SP) coding region (C, E2 and E1) is available at the United States Centres for Disease Control and Prevention (CDC). Reference viruses for provisional genotypes 1g and 2c are not available yet. – Les données concernant la séquence de la région codante des protéines de structure (C, E2 et E1) sont disponibles auprès des *United States Centres for Disease Control and Prevention* (CDC). Il n'existe pas encore de virus de référence pour les génotypes provisoires 1g et 2c.

^b Strains are named using the new convention described in this report. Missing information for some strains is being obtained or is unavailable. – Les souches sont désignées conformément à la nouvelle convention décrite dans ce rapport. Les informations manquent pour certaines souches ; elles sont en cours d'obtention ou n'existent pas.

^c Strains are named using old or previously published identifiers. – Souches désignées conformément aux critères anciens ou publiés antérieurement.

^d Attenuated vaccine virus for which original wild-type virus has been lost. – Virus vaccinal atténué pour lequel le virus sauvage d'origine a été perdu.

^e Progenitor wild-type virus for attenuated vaccine virus TO-336 Vac. – Virus sauvage à l'origine du virus vaccinal atténué TO-336 Vac.

^f The genotype of this virus is considered to be 1a. – On considère que le génotype de ce virus est 1a.

^g The CDC SP sequence is at least 2 nucleotides different from this published sequence. In the CDC sequence, nucleotide 9266 is changed from an A to a G and nucleotide 9386 is changed from a C to a G. – La séquence de la protéine de structure des CDC diffère au moins par 2 nucléotides de la séquence publiée ici. Dans la séquence des CDC, le nucléotide 9266 (A) est remplacé par G, et le nucléotide 9386 (C) est remplacé par G.

genotype has been found infrequently in recent years, limiting the number of new genotype 1a viruses and thus limiting the opportunities to further characterize the complexity of this group of viruses. Recently, however, new genotype 1a viruses have been found in Mongolia and Myanmar and this may allow further characterization of this group of viruses. Given the historical significance of this group, reference sequences were designated for this provisional group (Table 1). Genotype 1g was considered provisional both because reference sequences were not available and because the relationship between genotype 1g and genotype 1B was not clear. The clustering of viruses from genotype 1B and 1g is more sensitive to the sequence window used than that of viruses from genotypes 1C, 1D, 1E, 1F, 2A, 2B or 2c (Internet Fig. 2).² Genotype 2c was considered provisional only because reference viruses were not available; information reported at the meeting indicated that 2c reference sequences will soon be available.

The phylogenetic trees for reference viruses using different windows illustrate why the 739 nt window was recommended (Internet Fig. 2).² The structural protein window includes the entire coding region for the C, E2 and E1 pro-

dernières années, ce qui limite le nombre des nouveaux virus de génotype 1a et, par conséquent, également les possibilités de caractérisation de ce groupe complexe de virus. On vient cependant d'observer des nouveaux virus de génotype 1a en Mongolie et au Myanmar, ce qui permettra de mieux caractériser le groupe. En raison de l'importance historique de ce groupe, des séquences de référence ont été désignées pour ce groupe provisoire (Tableau 1). Le génotype 1g a été considéré comme provisoire car il n'existe pas de séquences de référence et la relation entre le génotype 1g et le génotype 1B n'est pas claire. Le groupement des virus de génotype 1B et 1g est plus sensible au choix de la séquence de nucléotides utilisée que les génotypes 1C, 1D, 1E, 1F, 2A, 2B ou 2c (Figure 2 sur le site Internet).² Le génotype 2c a été considéré comme provisoire pour la seule raison qu'il n'existe pas de virus de référence ; d'après les informations fournies lors de la réunion, des séquences de référence pour le génotype 2c seront bientôt disponibles.

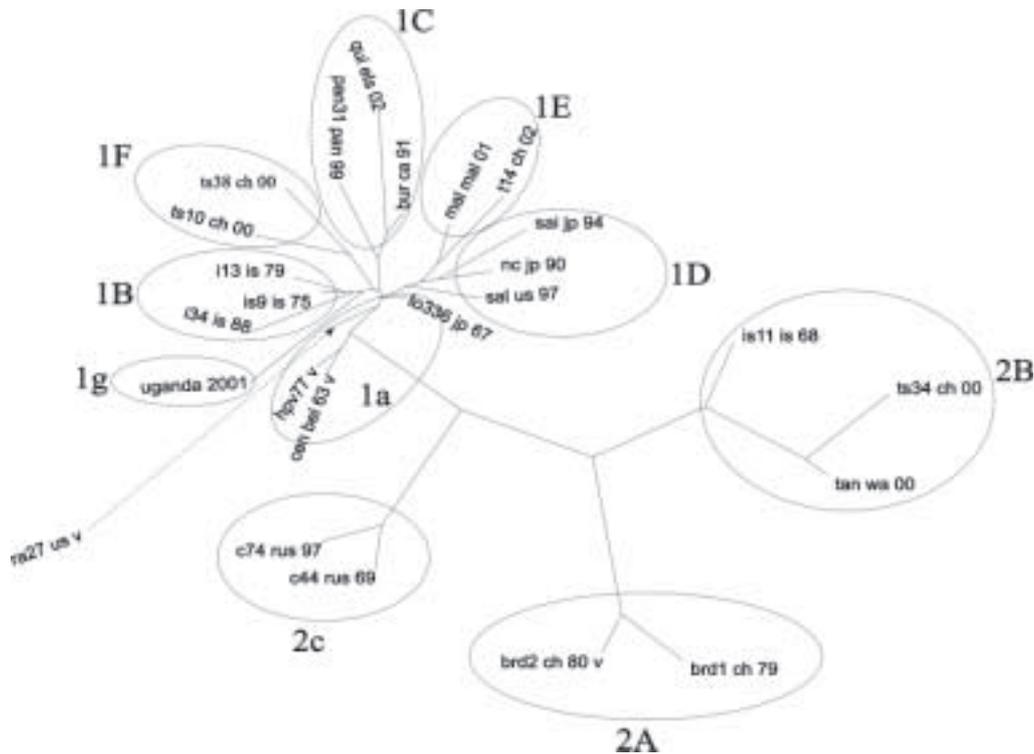
Les arbres phylogénétiques des virus de référence établis au moyen de différentes fenêtres montrent pourquoi la fenêtre de 739 nucléotides a été recommandée (Figure 1 sur le site Internet).² Cette fenêtre contient la région codante entière des protéines C, E2 et E1 et sert de

² See <http://www.mayeticvillage.com/who-rubellagenotype>, user name: "rubella"; password: "rubella".

² Les informations sont disponibles sur le site Internet associé au présent rapport: <http://www.mayeticvillage.com/who-rubellagenotype>, nom d'utilisateur: «rubella»; mot de passe: «rubella».

Fig. 1 **Phylogenetic tree of rubella virus reference sequences listed in Table 1, using nts 8731-9469^a**

Fig. 1 **Arbre phylogénétique des séquences de référence des virus rubéoleux présentées dans le tableau, d'après la séquence de nucléotides 8731-9469^a**



^a The Bayesian inference program, MrBayes, was used with settings of 200 000 Ngens; Samplefreq, 100; Nchains, 4; Burnin, 10. The position of 1 sequence (uganda 2001), which is probably in genotype 1g, is shown to represent this provisional genotype. The positions of 2 genotype 2c viruses are shown using published E1 sequences. See also Table 1, and Internet Table 2, Internet Fig. 2 and Internet Fig. 3 (available at <http://www.mayeticvillage.com/who-rubellagenotype>, user name: "rubella"; password: "rubella"). – Le programme MrBayes utilisant les méthodes d'inférence bayésienne a été appliqué avec les critères suivants : Ngens 200 000; Samplefreq 100; Nchains 4; Burnin 10. La position d'1 séquence (Uganda 2001), qui pourrait être de genotype 1g, est indiquée pour représenter ce génotype provisoire. La position de 2 virus de genotype 2c est indiquée en utilisant les séquences publiées de E1. Voir également le Tableau 1 et le Tableau 2 sur le site Internet ainsi que les Figures 2 et 3 sur le site Internet (disponible sur <http://www.mayeticvillage.com/who-rubellagenotype>, user name: "rubella"; password: "rubella").

teins and is the standard against which the other windows were evaluated. In the analysis presented, the 601 window did not give the proper groupings for the 1B and 1g viruses, and the 552 window did not give high clade credibility values for the 1F genotype. However, the combined window – 739 – gave the proper groupings of reference viruses for the 1B genotype and high clade credibility values for viruses from the 1F genotype.

Proposals for new genotypes should be submitted to the strain banks for evaluation. When possible, the earliest isolates should be used as reference viruses.

référence pour évaluer les autres fenêtres. Dans l'analyse présentée, la fenêtre 601 n'a pas permis de grouper correctement les virus 1B et 1g, et la fenêtre 552 et la crédibilité des clades est faible pour le génotype 1F. La fenêtre combinée – 739 – a permis d'obtenir un groupement correct des virus de référence du génotype 1B et une crédibilité forte des clades pour les virus du génotype 1F.

Les propositions concernant les nouveaux génotypes doivent être soumises aux banques de souches qui les évalueront. On utilisera si possible comme virus de référence les virus isolés antérieurement.

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_29934

