Molecular Epidemiology of Measles Virus: Identification of Pathways of Transmission and Implications for Measles Elimination

Jennifer S. Rota, Janet L. Heath, Paul A. Rota, Gail E. King,* Maria L. Celma, Juan Carabañá, Rafael Fernandez-Muñoz, David Brown, Li Jin, and William J. Bellini

The nucleotide sequences of either the hemagglutinin or nucleoprotein genes from wild type measles viruses isolated in the United States between 1989 and 1992 differed by <0.5%. This suggests that the majority of viruses associated with resurgence of measles in the United States belonged to a single indigenous genotype. In contrast, wild type viruses isolated from sporadic outbreaks of measles in the United States during 1994 were genetically heterogeneous. These viruses were more closely related to wild type viruses previously circulating in Europe, Africa, or Japan and were epidemiologically linked to importations or no known source. In addition to demonstrating the utility of genetic analysis in understanding the epidemiology of measles, these data suggest that the transmission of the indigenous virus was interrupted after the 1989–1992 epidemic. Measures to further reduce the incidence of measles in the United States should include efforts to control importation and subsequent spread of measles.

Although the public health impact of measles has been effectively reduced by vaccination, previous attempts to eliminate the indigenous transmission of measles in the United States were unsuccessful [1]. One reason for the resurgence of measles cases in the United States between 1989 and 1992 was the failure to maintain high vaccination rates, particularly in urban areas [2]. This resulted in multiple generations of cases after introduction of the virus from either indigenous or imported sources. The present two-dose vaccination schedule and aggressive childhood vaccination programs were designed to achieve and sustain the very high vaccination rates necessary to prevent accumulation of susceptible persons.

The success of improved measles control measures was apparent in 1993 when a historic low number of cases (312) was reported. For a consecutive 6-week period in 1993, no indigenous measles cases were identified [3]. In 1994, >900 cases of measles occurred in the United States, but the characteristics of these more recent outbreaks differed from those of the resurgent period in that the cases occurred primarily in older age groups and away from major urban areas [4]. For many of the outbreaks in 1994, it was possible to link the index case to an importation of measles virus. However, in other instances, no index case was identified and the source of infection was unknown. Without epidemiologic evidence to the contrary, the cases involved in an outbreak with an unknown source were designated as indigenous [5]. Accurate identification of the source of outbreaks is necessary to assess the status of measles control programs.

The nucleoprotein (N) and hemagglutinin (H) genes of measles virus display a high degree of genetic variability, and sequence analyses of these genes have identified several distinct genotypes of wild type virus that are currently cocirculating worldwide [6–9] (unpublished data). Wild type measles virus can also be differentiated from vaccine strains by sequence comparison [10].

In this work, we used sequence analysis of H and N genes to identify the source of wild type measles viruses associated with outbreaks in the United States during 1994. These data facilitated the classification of measles cases as imported or indigenous as well as the identification or confirmation of epidemiologic links (or lack thereof) between cases or outbreaks (or both). The results also suggested that transmission of indigenous virus was interrupted following the resurgence of measles in the United States between 1989 and 1992.

Materials and Methods

Viruses and RNA preparation. The wild type measles viruses isolated in the United States during the years 1989–1992 are described in table 1. The group of viruses isolated in 1994 is described in table 2. Measles sequences that were used as reference sequences in this work were from viruses isolated in or during the...
Table 1. Background of the 12 US measles virus isolates used to construct the 1989–1992 US consensus sequence.

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Location</th>
<th>Provider*</th>
<th>Collection date</th>
<th>Condition</th>
<th>Age</th>
<th>Clinical specimen</th>
</tr>
</thead>
<tbody>
<tr>
<td>TX-1 1989</td>
<td>Houston</td>
<td>PW</td>
<td>1/89</td>
<td>Pneumonia (pregnant)</td>
<td>21 y</td>
<td>NP</td>
</tr>
<tr>
<td>TX-2 1989</td>
<td>Houston</td>
<td>PW</td>
<td>2/89</td>
<td>Pneumonia</td>
<td>4 y</td>
<td>NP</td>
</tr>
<tr>
<td>CA 1989</td>
<td>San Diego</td>
<td>SHD</td>
<td>2/89</td>
<td>Pneumonia</td>
<td>2 y</td>
<td>NP</td>
</tr>
<tr>
<td>IL-1 1989</td>
<td>Chicago</td>
<td>MS</td>
<td>9/89</td>
<td>Uncomplicated</td>
<td>10 y</td>
<td>NP</td>
</tr>
<tr>
<td>IL-2 1989</td>
<td>Chicago</td>
<td>MS</td>
<td>9/89</td>
<td>Liver transplant (AIDS)</td>
<td>9 y</td>
<td>Lung</td>
</tr>
<tr>
<td>IL-3 1989</td>
<td>Chicago</td>
<td>MS</td>
<td>6/89</td>
<td>Encephalitis</td>
<td>14 mo</td>
<td>NP</td>
</tr>
<tr>
<td>IL-4 1989</td>
<td>Chicago</td>
<td>MS</td>
<td>8/89</td>
<td>Uncomplicated</td>
<td>8 mo</td>
<td>NP</td>
</tr>
<tr>
<td>CA 1990</td>
<td>Berkeley, CA</td>
<td>EL</td>
<td>3/90</td>
<td>Pneumonia</td>
<td>8 mo</td>
<td>TS</td>
</tr>
<tr>
<td>PA-1 1990</td>
<td>Allentown, PA</td>
<td>SHD</td>
<td>6/90</td>
<td>Uncomplicated</td>
<td>20 y</td>
<td>TS</td>
</tr>
<tr>
<td>PA-2 1990</td>
<td>Philadelphia</td>
<td>AD</td>
<td>11/90</td>
<td>Uncomplicated</td>
<td>?</td>
<td>NW</td>
</tr>
<tr>
<td>NY 1991</td>
<td>New York City</td>
<td>JM</td>
<td>4/91</td>
<td>Uncomplicated</td>
<td>14 y</td>
<td>TS</td>
</tr>
<tr>
<td>TX 1992</td>
<td>Corpus Christi, TX</td>
<td>JS</td>
<td>3/92</td>
<td>Pneumonia</td>
<td>5 mo</td>
<td>TS</td>
</tr>
</tbody>
</table>

NOTE. NP = nasopharyngeal swab, NW = nasal wash, TS = throat swab, y = years, mo = months.

* Specimens or isolates were provided by respective state health departments (SHD) or following individuals and institutions: P. Wyde (PW, Baylor College of Medicine, Houston), M. Smaron (MS, University of Chicago, Chicago), E. Lennette (EL, Virolab, Inc., Berkeley, CA), A. Deforest (AD, St. Christopher's Hospital, Philadelphia), J. McPhee (JM, Presbyterian Hospital, New York), J. Smith (JS, Driscoll Children's Hospital, Corpus Christi, TX).

1 Referred to as "SO" in [7] (hemagglutinin [H] sequence) and [8] (nucleoprotein [N] sequence).
2 Referred to as "SD" in [7] (hemagglutinin [H] sequence) and [8] (nucleoprotein [N] sequence).
3 Referred to as "Chi" in [7] (H sequence) and [8] (N sequence).
4 Resulted in fatality.
5 Specimens or isolates were provided by respective state health departments (SHD) or following individuals and institutions: P. Wyde (PW, Baylor College of Medicine, Houston), M. Smaron (MS, University of Chicago, Chicago), E. Lennette (EL, Virolab, Inc., Berkeley, CA), A. Deforest (AD, St. Christopher's Hospital, Philadelphia), J. McPhee (JM, Presbyterian Hospital, New York), J. Smith (JS, Driscoll Children's Hospital, Corpus Christi, TX).

Following locations and years: a strain from Bilthoven, Netherlands, in 1991 (provided by A. D. M. E. Osterhaus, Erasmus University, Rotterdam), an isolate from the 1993 epidemic in the Republic of Palau (provided by D. Guris, National Immunization Program, CDC, Atlanta), a 1991 strain from Gambia (provided by H. Whittle, Medical Research Laboratories, Gambia), viruses from Spain in 1992 and 1993 (Madrid/92A, Madrid/93A), a virus from England isolated in 1993, and viruses from two west African countries, Gabon (R96) and Cameroon (Y14), isolated in 1984 and 1983, respectively (provided by F. Wild, Institut Pasteur de Lyon, Lyon, France). Additional reference sequences, which were previously published, include the hemagglutinin sequence from the 1988 Japanese isolate AK-1 [11] referred to here as Japan/88, the sequences from the low-passage stock of the 1954 Edmonston strain [7, 8], and the N sequences of the African viruses (Gabon and Cameroon) [8].

The propagation of measles virus and RNA extraction methods using vero E6 cells have been described [7]. An Epstein-Barr virus–transformed marmoset B lymphoblastoid cell line (B95a cells) [12] was also used to isolate measles virus from clinical specimens. A modified guanidine hydrochloride procedure [13] was used to extract the RNA from the viruses grown on B95a cells.

cDNA cloning and sequencing. The reverse transcription (RT) and polymerase chain reaction (PCR) conditions and primers have been described [7, 8]. Additional primers with alternate restriction sites for amplification of H were constructed. These were as follows: 5'-GCATCGAAGCTTTGACGATCCATCAAGCTG-3' (HindIII) and 5'-GCATCGCTGCAGGATCTGGGTGACATCATG-3' (PstI). Sequences were derived by direct sequencing of the amplified cDNA or by sequencing clones of the amplified DNA. A second round of RT-PCR with sequencing of the PCR products verified nucleotide changes encountered on sequencing of the cloned genes. Sequences from the isolate from England were obtained by direct PCR on clinical material. For simplicity, all sequences were converted to (+) sense DNA.

Computer analyses of nucleotide sequences. Sequence data were analyzed using version 8.0 of the sequence analysis software package of the University of Wisconsin Genetics Computer Group [14] and the Phylip (Phylogeny Inference Package, version 3.5) software package [15]. All phenograms were drawn as unrooted trees.

Results

Genetic analysis of the US wild types: 1989–1992. To assess the degree of heterogeneity among measles viruses isolated during the US resurgence, the H and N genes from a collection of viruses isolated during that period were sequenced. The sequences were compared with each other and with the sequences from a low-passage stock of the Edmonston virus (wt-Edmonston) [7, 8]. The isolates were obtained from patients with diverse clinical conditions, ranging from uncomplicated measles to fatal pneumonia and AIDS (table 1). The H genes from 12 isolates collected between 1989 and 1992 were analyzed. The N gene was sequenced from 3 of these 12 isolates.

The 12 H genes from wild type viruses isolated from 1989 to 1992 were nearly identical, containing <0.4% nucleotide divergence among the strains. There were 53 nucleotide changes common to all 12 of the viruses, 17 of which predicted amino acid substitutions relative to the H sequence of wt-Edmonston. The 4 viruses from the Chicago (IL 1–4) outbreak in 1989, at the beginning of the resurgence, contained only one
Table 2. Measles isolates analyzed from the United States in 1994.

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Location</th>
<th>Provider*</th>
<th>No. of cases†</th>
<th>Status†</th>
</tr>
</thead>
<tbody>
<tr>
<td>NJ 1994</td>
<td>Rutgers University, NJ</td>
<td>RU</td>
<td>94</td>
<td>Import</td>
</tr>
<tr>
<td>WA 1994</td>
<td>Seattle</td>
<td>SHD</td>
<td>2</td>
<td>Import</td>
</tr>
<tr>
<td>TN 1994</td>
<td>Memphis</td>
<td>SHD</td>
<td>28</td>
<td>Unknown</td>
</tr>
<tr>
<td>Guam 1994</td>
<td>US territory, Pacific</td>
<td>DG</td>
<td>228</td>
<td>Unknown</td>
</tr>
<tr>
<td>NE 1994</td>
<td>Omaha</td>
<td>KS</td>
<td>1</td>
<td>Import</td>
</tr>
<tr>
<td>IL 1994</td>
<td>Principia College, IL</td>
<td>SHD</td>
<td>54</td>
<td>Unknown</td>
</tr>
<tr>
<td>VT 1994</td>
<td>Windsor County, VT</td>
<td>SHD</td>
<td>1</td>
<td>Import</td>
</tr>
<tr>
<td>NY 1994</td>
<td>Jamaica, NY</td>
<td>KB</td>
<td>1</td>
<td>Import</td>
</tr>
<tr>
<td>CO 1994</td>
<td>Grand Junction, CO</td>
<td>SHD</td>
<td>62</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

* Specimens or isolates were provided by respective state health departments (SHD) or by following individuals and institutions: RU (Rutgers University, New Brunswick, NJ), D. Guris (CDC, Atlanta), K. Shuck (St. Joseph Hospital, Omaha), K. Bromberg (Kings County Hospital, Brooklyn, NY).

† Epidemiologically linked.

1 Status of index case based on conventional epidemiology at time of receipt of specimen or isolate.

unique noncoding change that distinguished these epidemiologically linked viruses from the other viruses examined from the resurgence. Among all of the viruses, there was variable substitution at 2 nucleotide positions, 1461 and 1656, even within the 4 viruses from the single outbreak in Illinois. This may be due to different cell culture passage levels, because in 9 vaccine strains previously characterized [10], these 2 positions were observed to be nonvariable. TX-1 1989 and TX-2 1989 contained 1 additional nucleotide change in H at position 1233 (histidine to asparagine), and TX 1992 also contained 1 additional nucleotide change at position 216, with a predicted change from tyrosine to histidine. Aside from these 3 strains from Texas, which contained 1 additional coding change, there were no other variations in the predicted amino acid composition of the H proteins among the 12 strains.

For simplicity, the sequences from the 12 US viruses were combined into a single consensus H sequence for comparison with other strains and in the dendrograms. The individual sequences from the 12 wild type viruses vary from the consensus sequence by ≤4 nucleotides. Although the N gene was sequenced from only 3 of the 12 US isolates, the divergence within the N gene was comparable (<0.5%) to that observed with H. These sequences were also combined into a single consensus sequence for the purposes outlined above.

Genetic analysis of US wild types: 1994. In contrast, measles viruses from outbreaks investigated in 1994 contained sequences that were genetically diverse and were representative of several genetic lineages, designated A-E in the dendrograms (figures 1, 2). For epidemiologic purposes, we have chosen to discriminate between lineages A and B, although an ancestral node for both of the clusters is shared. It was curious that the H sequence from Japan/88 was genetically more similar to the US consensus H sequence (lineage A) than to the more recent isolates epidemiologically linked to Japanese nationals, NE 1994 and Palau/93 in lineage B. The viruses from 2 known imported measles cases, NE 1994 and WA 1994, were both closely related to Palau/93 (lineage B), which was traced to a Japanese tourist. The virus isolated from an outbreak in Colorado in late 1994 was also genetically related to the Palau strain, although more distantly than NE 1994 and WA 1994.

The H sequence from the isolate from an outbreak at Rutgers University (NJ 1994) was only 1 nucleotide different from that obtained from VT 1994 (lineage C). Viruses independently isolated in England (England/93) and Spain (Madrid/93A) were found to be nearly identical to each other and to NJ 1994 and VT 1994 (figures 1, 2).

TN 1994 and IL 1994 were closely related to Netherlands/91 and Madrid/92A in lineage D. Other viruses that have been identified as belonging to this lineage include the JM strain, which was isolated in the United States in 1977, and recent wild type viruses isolated in Germany [9]. In addition, isolates from the United Kingdom in 1994 (unpublished observations) were determined to belong in this lineage.

NY 1994 was related to the strains in lineage E isolated in Africa. Although there was an 11-year span between the isolation of Cameroon/83 and the NY 1994 isolate, the two sequences are within 1.5% nucleotide divergence in the H coding gene and 1.9% divergence in the N gene. This was far more similarity than was observed between the sequences from vi-

Figure 1. Dendrogram based on nucleotide (NT) sequences of H coding genes of wild type measles strains. US 1989-1992 consensus H sequence was constructed using 12 isolates listed in table 1. Scale for this unrooted tree is based on no. of NT changes. A-E, 5 genetic lineages.
The genetic analysis of measles isolates involved the construction of a dendrogram based on nucleotide (NT) sequences of the N coding gene of wild type measles strains. US 1989-1992 consensus N sequence was constructed using 3 of 12 isolates listed in table 1: IL-1 1989, CA 1989, and TX 1992. N sequence for Japan/88 was not available, and N sequence for VT 1994 was not done. Scale for this unrooted tree is based on no. of NT changes. A–E, 5 genetic lineages.

**Figure 2.** Dendrogram based on nucleotide (NT) sequences of N coding genes of wild type measles strains. US 1989–1992 consensus N sequence was constructed using 3 of 12 isolates listed in table 1: IL-1 1989, CA 1989, and TX 1992. N sequence for Japan/88 was not available, and N sequence for VT 1994 was not done. Scale for this unrooted tree is based on no. of NT changes. A–E, 5 genetic lineages.

To more clearly describe the genetic relationships between the viruses and incorporate them into a meaningful context for epidemiologic interpretation, we chose to designate viruses as a measles genotype when the H sequences are within 0.5% nucleotide divergence (<10 nucleotides apart). Pairwise comparisons of H nucleotide differences between the strains are shown in figure 3, with bold blocks indicative of relationships between the paired viruses that share a common genotype.

It is our intent to correlate these genetic relationships to transmission patterns and epidemiologic links. The criteria for defining a genotype was based on sequence data collected along with the corresponding known epidemiologic information. We would expect viruses with epidemiologic links to be within one genotype. There were several instances in our small sampling of measles in the United States during 1994 in which the conventional epidemiology was insufficient to identify sources for the cases (table 2). The sequence data support the likelihood that many of the cases for which the source was unknown resulted from imports. For example, the source of the outbreak in Tennessee was not traced, and without information to the contrary, the source for this outbreak was reported to be indigenous. Although the Tennessee outbreak remains classified as indigenous, by using molecular analysis, we can conclude that the genotype connected to that outbreak is unlike the viruses that circulated during the resurgence and instead is related to viruses that have been identified in Europe.

Because of the geographic proximity of Guam and Palau, with only conventional epidemiology, the outbreak on Guam was assumed to be linked to the outbreak that occurred in Palau from June to October 1993. That theory was ruled out following sequence analysis of the strains isolated from the two outbreaks. Similarly, the outbreak in New Jersey was presumed to be linked to the outbreak in Illinois at Principia College, due to the temporal proximity of the two outbreaks. It was suspected that students from Rutgers University in New Jersey spread the outbreak to Colorado during the college break. Prior to onset of measles, the index case for the outbreak at Principia College had been to Breckenridge, Colorado, where an outbreak of measles was occurring [16, 17]. However, the sequence information revealed that the viruses from New Jersey and Illinois were from different sources on the basis of the different genotypes identified.

VT 1994 was identified as an import because the family had arrived in the United States from England 12 days before rash appeared. On the basis of sequence similarity, it was first suspected that the Vermont case may have had an epidemiologic link to the Rutgers outbreak, which had been traced to a Spanish student. However, the observation that sequences from the 1993 isolates from Spain and England were strikingly similar suggests that a common genotype was circulating in both countries, and therefore the Vermont and New Jersey cases were not necessarily epidemiologically connected.

NE 1994, isolated from a Japanese student who had just traveled from Japan, was very similar to the Palau virus, corroborating the epidemiologic link to the index case in the Palau outbreak, a Japanese tourist. It is unknown why the virus isolated from the child who had traveled to Switzerland (WA 1994) had a sequence similar to those of the Palau and Nebraska viruses, but the source could have been on the airplane.

NY 1994 was isolated from a child who had just arrived from Kenya in eastern Africa. The H sequences from NY 1994 and Cameroon/83 were surprisingly similar. Although an African lineage is apparent from the dendrograms (E; figures 1, 2), the divergence between the various African genotypes in our sampling group was as high as 3.45% (64 nucleotides different; figure 3).

The last outbreak investigated during 1994 occurred in Grand Junction, Colorado. The index case was known to have traveled to another state before onset of symptoms, but no source had been identified. The molecular characterization of CO 1994 indicated that the virus was similar to strains with links to Japanese nationals in lineage B and clearly unrelated to any of the previous outbreaks investigated during 1994.

**Discussion**

Regardless of the temporal and geographic divergence of the isolates obtained from clinical cases of measles during the

resurgence, the sequences of the H or N genes of these viruses were nearly identical. This observation suggests that a single, predominant viral genotype was responsible for the bulk of the 50,000 measles cases that occurred during the resurgence. Presumably viruses from other genetic lineages were introduced into the United States during that period but at too low a frequency to be represented in our sampling of viruses. Our data suggest that other genotypes did not circulate widely and probably played a minor role during the resurgence.

In contrast, the H and N gene sequences from viruses isolated from outbreaks occurring in 1994 were heterogeneous, reflecting the diverse geographic origins of these viruses. However, none of the viruses isolated in the United States during 1994 was a member of the genotype associated with the resurgence. In addition to the molecular diversity of the 1994 isolates, the epidemiologic characteristics of these cases were different from those of the majority of cases occurring during the resurgent years. Between 1989 and 1992, there was a marked increase in the number of measles cases among children <1 year of age, and the largest outbreaks occurred among preschool-age children living in inner cities [2]. In 1994, there was a decline in the numbers of cases among children <5 years old, and the largest proportion of cases occurred in high school- and college-aged persons. The cases occurred sporadically during 1994 and were not located in major urban areas, whereas during the resurgence, cases were concentrated in large urban areas, such as Chicago and Houston [4].

During and following the 1989–1992 resurgence, renewed efforts to increase vaccination levels, especially among inner city populations, along with a second-dose schedule of measles vaccine for older children and college students may have been responsible for an all-time low of 312 reported measles cases in 1993. The paucity of cases in 1993 suggests that transmission of the indigenous virus that was responsible for the majority of cases during the resurgence had been interrupted. In fact, there was a 6-week period in the latter part of 1993 during which no measles cases were reported that were not directly linked with importations, including 3 consecutive weeks in 7–27 November, during which no measles cases were reported [3].

Before 1993, the lowest recorded incidence of measles had occurred in 1983, following a measles elimination campaign [18]. Although there were few wild type measles isolates available from 1983 for sequence analysis, the only genotype identified during that time, which may have been the indigenous lineage in the United States in 1983, has not been isolated again. Viruses isolated in the United Kingdom between 1983 and 1988 [8] were more closely related to the US viruses from the resurgent period than to the viruses isolated in the United States in 1983. Whether the genotype prevalent during the resurgence represents an endemic strain that evolved within the United States or was introduced into the United States before the 1989 resurgence is unknown. However, the sequence data suggest that this genotype was introduced by importation and became widely disseminated.
In 1994, none of the outbreaks investigated yielded a virus with the same genotype as that identified during the resurgence with the exception of the outbreak on Guam. Despite its absence from the United States, this genotype was introduced into Guam and a neighboring island (Chuuk) in Micronesia (unpublished data), resulting in >1000 cases and 16 measles-related deaths [19].

The information from sequence analysis of the 1994 isolates made a substantial contribution to the understanding of measles transmission. The sequence data confirmed epidemiologic links between outbreaks in some instances and, in others, supported or contradicted suspected links that were based on the temporal or geographic characteristics of the outbreaks. With time, regional or endemic genotypes may change because of the continual reintroduction of measles virus into areas in which indigenous virus has been eliminated. This has already resulted in a very mixed pattern of viral genotypes across large geographic areas as in the United States in 1994, Europe [9], and the United Kingdom (unpublished observations).

The United States has established 1996 as a target date for the elimination of indigenous transmission of measles, and continued molecular surveillance of wild type measles strains will be necessary to monitor progress toward this goal. Our data suggest that the 1996 goal can be achieved by maintaining the current two-dose vaccination strategy and increasing control of imported measles. The success of recent efforts to control measles in the United States is being aided by ongoing campaigns in Mexico, the English-speaking Caribbean, and Central and South America organized by the Pan American Health Organization (PAHO). Since initiation of PAHO’s National Day vaccination program in 1991, the number of importations from these countries into the United States has dropped substantially [1, 20]. If the interruption of indigenous measles transmission can be accomplished by maintaining high immunization levels, there is hope for eventual eradication of measles. Indeed, the only truly cost-effective strategy will be a worldwide effort toward eradication. Until then, the costs of vaccination programs to individual countries attempting elimination will be high [21]. For regional programs to be completely effective, there must be a concerted global effort to control measles.

Acknowledgments

We thank Ali Khan for his assistance with the outbreak in New Jersey, Yvonne Villamarzo for technical assistance, and Charles Vittek and Susan Redd for clarifications and helpful discussion. We dedicate this report to the memory of Gail Eugene King, whose insight led to greater collaboration and knowledge towards measles control and eventual elimination.

References