Toward Universal *Flavivirus* Identification by Mass Cataloging

George W. Jackson,^{*,†} Roger J. McNichols,[†]
George E. Fox,^{*,‡} and Richard C. Willson^{*,‡}

From the Departments of Chemical and Biomolecular
Engineering* and Biology and Biochemistry,[‡] University of
Houston; and BioTex, Inc.,[†] Houston, Texas

Development of rapid and specific molecular diagnostics for flaviviruses remains an important global health challenge. Herein a platform technology using mass spectrometry that can be used for universal identification and genotyping of these viruses is described. The feasibility of the approach is demonstrated by using it to correctly identify and serotype two strains of dengue virus. Predictive calculations show that the approach can be expected to be equally efficacious for the identification and epidemiological tracking of other flaviviruses including West Nile, Japanese encephalitis, and Yellow Fever. In the case of dengue at least, the method can also distinguish major subgroupings within each serotype. All process steps are amenable to high-throughput, automated implementation. The assay protocol is also compatible with miniature mass spectrometers currently in development, thereby allowing the assay to be brought to remote locations for rapid response to and tracking of outbreaks. (J Mol Diagn 2008, 10:135–141; DOI: 10.2353/jmoldx.2008.070130)

The genus *Flavivirus* includes more than 70 single-stranded RNA viruses responsible for severe encephalitic, hemorrhagic, hepatic, and febrile illnesses in humans and other vertebrates.¹ Among the pathogens in this genus are yellow fever, tick-borne encephalitis, Japanese encephalitis, St. Louis encephalitis, West Nile, and dengue viruses. Together, they cause considerable morbidity and mortality worldwide. Of particular interest are the four serotypes of the mosquito-borne dengue virus (DEN-1, -2, -3, and -4) that can cause dengue hemorrhagic fever and dengue shock syndrome. Currently there is no protective vaccine or specific treatment available, and it is estimated that over 100 million cases of dengue occur annually worldwide.^{2,3} Typing of dengue is of critical importance to distinguish endemic strains from new outbreak strains so that new outbreaks can be readily contained. Typing is also important in treatment because infection by a new serotype in a patient previously infected by one of the other three is associated with

a greatly increased risk of developing dengue hemorrhagic fever and/or dengue shock syndrome.^{4,5}

The current reference method for identifying and typing the various flaviviruses is isolation of the virus in cell culture followed by immunofluorescence typing.⁶ This time-consuming procedure requires significant expertise and cell culture facilities that may not be readily available. Alternative serological tests are relatively easy to use and can accommodate a large number of samples, both necessities when confronting an epidemic. These benefits, however, come at a cost; tests such as hemagglutination inhibition, IgG-enzyme-linked immunosorbent assay and IgM antibody capture enzyme-linked immunosorbent assay cannot easily distinguish dengue at the serotype level and are likely to misidentify other flaviviruses as dengue.^{2,7,8} In addition, in areas where a virus is endemic, an outbreak strain will need to be distinguished from any endemic strain that might belong to the same serotype.

To address these challenges, nucleic acid-based assays are of considerable interest. The power of nucleic acid information is well established and numerous assays that use sequence information have been developed or are under development for various bacterial and viral threats. These assays are typically based on hybridization protocols or multiplex polymerase chain reaction (PCR) and tend to be unique to each infectious agent,^{8–12} or in the case of PCR assays of broader organism coverage, secondary analysis by sequencing or at least gel electrophoresis is required.^{1,13,14} A general platform technology capable of identifying multiple agents more rapidly than complete sequencing and without prior knowledge of which agent may be present in the sample would be extremely useful. The core goal is to enable a single laboratory using a single cluster of standard instruments with a variety of threat-specific protocols to do essentially all of the analysis associated with an ongoing event or the follow-up forensics. Mass

Supported in part by grants E-1451 (to G.E.F.) and E-1264 (to R.C.W.) from the Robert A. Welch Foundation; a grant from the Institute of Space Systems Operations (to G.E.F.); and Small Business Innovation and Research grants NNM06AA44C from NASA and 2006–33610-16775 from the U.S. Department of Agriculture (to G.W.J.).

Accepted for publication November 15, 2007.

G.W.J., G.E.F., and R.C.W. are coinventors with Dr. Zhengdong Zhang (Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT) on a pending U.S. Patent, "Microbial Identification Based on the Overall Composition of Characteristic Oligonucleotides," which has been exclusively licensed to BioTex, Inc., Houston, TX.

Address reprint requests to Bill Jackson, Ph.D., 8058 El Rio St., BioTex, Inc., Houston, TX 77054. E-mail: bill@biotexmedical.com.

spectrometry, as described herein, can provide such a universal analysis platform.

The goal of the work described here was to assess the feasibility of developing mass spectrometric protocols for the broad identification and serotyping of flaviviruses based on rapidly observable molecular information, without the use of sequencing. Once successfully implemented for a standard laboratory environment, it may well be possible to adapt these protocols for use with portable field-usable mass spectrometers such as those currently being developed.¹⁵ In addition, analogous assays using very similar protocols can be developed for many other high-priority pathogens such as the National Institutes of Health National Institute of Allergy and Infectious Diseases category A, B, and C bacterial and viral threats (of which the flaviviruses are a highest level, priority A pathogen).

Overview of Assay Approach

Detailed descriptions of our approach to microbial genotyping by mass cataloging using mass spectrometry have recently been published.^{16–18} While these studies focused on broad or “universal” genotyping of bacteria, the methods are extendible to viral identification as described herein. In brief, the approach relies on identification of widely-conserved genomic regions for universal PCR priming (or reverse transcription-PCR). In the case of bacteria, primers targeting 16S ribosomal RNA (16S rRNA) are used, thereby leveraging the enormous number of available 16S sequences. After PCR (currently requiring ~35 minutes) using primers appended with an appropriate RNA polymerase promoter sequence, DNA is rapidly (~20 minutes) transcribed to RNA and digested in a highly specific, predictable manner using ribonuclease (RNase) T₁ to create a mixture of “base-specific cleavage” products.^{16–20} A matrix-assisted laser desorption/ionization/time of flight (MALDI-TOF) mass spectrum of the digest is rapidly acquired, and the resulting catalog of masses is compared to a database of predicted digestion patterns for thousands or more organisms. Finally, a computationally efficient spectral scoring function based on inner products, which is referred to as spectral “coincidence” (Equations 1 and 2) is used to quantitatively score the agreement of the observed spectrum with predicted spectra for every organism in the database(s). For this purpose, we define a scalar product (often referred to as a “dot-product”) of two mass spectra as

$$1 \quad \langle M, M' \rangle = M \cdot M' \equiv \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \delta(m_i - m'_j)$$

where m_i are the masses of each of the N_1 individual cleavage products in the spectrum for species 1 and m'_j are the masses of each of the N_2 cleavage products for species 2, and δ is the discrete (Kronecker) delta function defined as

$$2 \quad \delta(k) = \begin{cases} 1 & k = 0 \\ 0 & \text{otherwise} \end{cases}$$

While similar in concept to restriction fragment length polymorphism in which double-stranded DNA is digested

with a DNA restriction enzyme, cleavage of single-stranded RNA with base-specific endoribonucleases generates fragments of appropriate mass for unambiguous determination of composition in modern mass spectrometers, and the analysis is much more rapid and reproducible than electrophoresis.

Flavivirus Assay Development

The objective of this work was to assess the feasibility of adapting the mass cataloging approach from bacteria to *Flavivirus* identification. To this end, “almost-universal” primer pairs for flaviviruses in general and all four serotypes of dengue in particular were derived from the literature.^{1,9,13,21} The coverage of each primer pair was evaluated against large sequence data sets. The sequences that would be expected to be amplified were digested with RNase T₁ *in silico*, and the expected mass catalogs were calculated. The coincidence function described above was used to calculate distance matrices. Subsequently, cluster analysis was used to determine which viruses and serotypes would be expected to be resolvable. Based on the successful clustering of the various flaviviruses and serotypes of dengue in these predictive calculations, the approach was experimentally verified on representative strains of DEN1 and DEN4.

Materials and Methods

Informatics

Scaramozzino et al,¹ Drosten et al,¹³ and others have identified primer sets useful for both broad and species-specific amplification of flaviviral sequences.^{1,9,13,21} These primer sets were tested and extended as follows. For broad *Flavivirus* amplification a large set of *Flavivirus* complete genomes, partial sequences, and sequences for the fifth nonstructural protein NS5 were collected into a database. A ClustalW multiple sequence alignment²² was performed, and the degenerate positions were compared with the forward (MAMD) and reverse (cFD2) primers reported by Scaramozzino et al¹ targeted at the NS5 coding region in most flaviviruses. Based on our alignments, variants of these degenerate forward and reverse primers were chosen. Specifically, the literal search strings or *in silico* “primers” (using standard IUPAC base degeneracies; N = A, C, G, or T; D = A, G, or T; H = A, T, or C; M = A or C; R = A or G; S = G or C; Y = C or T) were ScaraFWD: 5'-ACATSATGGGNAARMGDRAR-3' and ScaraREV: 5'-GAYACMGCHGGHTGGGA-3'. The latter reverse primer sequence must be reverse complemented for actual PCR. As of July 2007, both primer sequences were simultaneously present in 287 of 383 or 75% of complete *Flavivirus* genomes downloaded from GenBank (Entrez nucleotide). In practice, these primers could be used under conditions of “permissive” or “mismatch-tolerant” PCR (or RT-PCR), further extending their coverage. Similarly, the following dengue-specific primers described by Drosten et al¹³ and targeted at the 3'

noncoding region of the dengue genome were used to create a database of expected base-specific cleavage masses from the parent sequence database: DrostFWD: 5'-AGACCAGAGATCCTGCTGT-3' and DrostREV: 5'-RCGCCRSAARATGGAWTG-3'.

The sense strand sequences of the resulting virtual amplicons were then fragmented by cleavage after each guanosine residue to simulate an RNase T₁ digest of RNA transcribed from the cDNA product. As in previous work in which phylogenetic trees were developed based on mass spectrometric observables,¹⁷ a matrix of distances (1 – coincidence values) between the virtual mass spectra of all of the amplified flaviviruses was constructed. The pairwise distances were then entered into the program MEGA3²³ version 3.1 to generate trees using the neighbor-joining algorithm.

Experimental Methods

Plasmids containing the entire cDNA genomes of DEN-1 (WP) and DEN-4 (341750) were graciously provided by Dr. Barry Falgout and Dr. Robin Levis of the Food and Drug Administration. Based on the alignments above, variants of the degenerate primers described by Scaramozzino et al¹ were obtained from Sigma-Genosys (The Woodlands, TX) to produce an amplicon of ~260 bp (DEN-1, less primer appendages; see below). The primers used were ScaraFWD: 5'-taatacgactcactataaggA-CATSATGGGNAARMGDRAR-3' (where lowercase indicates T7 RNA polymerase promoter) and ScaraREV: 5'-ctatatataTCCCADCCDGCKGTRTC-3' (where italicized lowercase indicates a sequence used to generate a calibration mass; this unique mass created by multiple amino-allyl U's incorporated in the transcript also serves as a confirmation that the RNA transcription was full length). Similarly, the dengue-specific primers DrostFWD and DrostREV were used to obtain a 63-bp amplicon from the DEN4 genetic material. These forward and reverse primers were also 5'-appended with the T7 promoter and calibration sequences, respectively. Approximately 2 ng of dengue type 1 cDNA template was used in an optimized universal PCR reaction (~105 minutes for 30 cycles). Following PCR, excess primer was digested for 5 minutes with DNA exonuclease I, and the exonuclease was then deactivated at 80°C for 5 minutes. The remaining double-stranded amplicon was then used as a template for *in vitro* transcription using a T7-Flash kit (Epicenter, Madison, WI) for 20 minutes. Importantly, amino-allyl UTP was used as a 100% substitution for natural UTP in the reaction to give improved distinction of U/C-containing mass fragments. Finally, the transcripts were completely digested by RNase T₁ at 37°C for 5 minutes and then placed on ice for MALDI preparation (~5 minutes/sample by hand including ZipTip desalting).

Mass spectra of the digests were acquired in linear, negative ion mode on a Voyager DE-STR MALDI-TOF (Applied Biosystems). All spectra were processed in an identical fashion as described in detail elsewhere¹⁶ including noise filtering, mass calibration, centroiding, and thresholding. The resulting short peak lists were then subjected to automated comparison to large, predicted

mass-fragment databases for either the Scaramozzino or Drosten interprimer region (depending on the primers used). For example, our database created to contain all GenBank-available *Flavivirus* genomes and NS5 sequences contained 443 records (strains) and 4099 predicted base-specific fragment masses. Overall scores for viral typing were obtained by spectral scoring to compare observed fragment masses to predicted masses.

Results

Prediction of Assay Specificity by Mass-Based Neighbor Joining

Figure 1 shows the result of cluster analysis by neighbor joining of 287 strains of flaviviruses using the predicted base-specific cleavage patterns on the homologous Scaramozzino amplicon for each virus. As can be seen, despite some information loss in moving from analysis of primary sequences to compositions (masses), the viruses cluster quite well with no strains falling into the wrong serotype groups. In our bacterial studies we have shown that, because such trees are based on realistic observables, such clustering is generally predictive of the degree to which organisms (or in this case, viruses) can be experimentally resolved. The Scaramozzino primers were intended to identify a homologous genomic region from all members of the genus *Flavivirus*. We have also applied the same methods for mass fragment-based cluster analysis using the dengue-specific primers described by Drosten et al¹³ for amplifying the 3' noncoding region from all four serotypes of dengue. When targeted at our database of 383 complete *Flavivirus* genomes, only

Figure 1. Cluster analysis based on mass spectral patterns of flaviviruses amplified by universal PCR of the NS5 coding region and fragmented by RNase T₁.

Figure 2. Dengue serotype clustering based on mass fragments between primers described by Drosten et al.¹⁵

dengue strains were “amplified,” verifying the dengue specificity of the Drosten primers. Figure 2 shows the results of clustering based on mass fragment patterns and distances derived from the Drosten universal dengue amplicon. To conserve space, a large branch length was removed between serotype 4 and serotypes 1, 2, and 3, and a segment of the tree was shifted to the left.

As with the analysis of the NS5 region, significant separation was achieved indicating the feasibility of serotyping by this method. The only difficulty encountered was that DEN2 and DEN3 strains could not be resolved by the mass fragments located between the Drosten primers. As shown in Figure 1, however, DEN2 and DEN3 can be readily resolved by base-specific cleavage masses derived from the NS5 region (Scaramozzino primers). The result is also consistent with previous work

showing that dengue typing and molecular relatedness can be empirically achieved by digestion of dengue genomic RNA with RNase T₁. In this case molecular information was obtained from two-dimensional electrophoresis of the RNA fragments.²⁴

Experimental Dengue Serotyping by Mass Spectrometry

Using cDNA template genomes for DEN1 and the Scar-aFWD and Scar-aREV primers described above (see Experimental Methods) we amplified the corresponding region of the DEN1 genome. Base-specific fragmentation spectra were acquired and scored by coincidence analysis. Figure 3 shows a typical spectrum acquired from a DEN1 sample using the Scaramozzino primers and corresponds to the genomic region analyzed by the cluster analysis of Figure 1. Note that formally, the abscissa of Figure 3 is the mass-to-charge ratio, m/z ; in MALDI, however, most ions are singly charged. Thus the x axis can be read practically as “mass.” For the purposes of universal *Flavivirus* identification, an expanded database containing additional sequence information for the NS5 genes as well as complete *Flavivirus* genomes was constructed. Taking the spectral acquisition from Figure 1 and comparing it to the database of predicted spectra for 338 flaviviruses allowed a rapid unambiguous typing of the dengue sample. Table 1 gives the identification rank by coincidence analysis in the context of 338 other *Flavivirus* strains that would have also been amplified by the universal (Scaramozzino) primers.

Identification rank is based on normalizing the final calculated mass spectral coincidence by the highest value. Notable in the table is that coincidence analysis with proper documentation gives a quantitative means to compare strains based on date, locale, and so forth. The highest ranked 29 strains were all DEN1. The strains ranked between 30 and 90 alternated between DEN2 and DEN3 except for the conspicuous appearance of the Cambodian type 1 strain that ranked 37th. The “co-mingling” of the lower ranking types 2 and 3 can be explained at least

Figure 3. Representative mass spectrum for dengue virus type-1 corresponding to the fragments derived using universal flavivirus primers analyzed in the cluster analysis of Figure 1.

Table 1. Mass Spectral Typing of Dengue-1 against All Other Flaviviruses by C, the Normalized Coincidence

ID rank	C	Virus name
1	1.0000	Dengue type 1 recombinant clone rDEN1mutF complete genome
1	1.0000	Dengue type 1 recombinant clone rDEN1delta30 complete
1	1.0000	Dengue type 1 recombinant clone rDEN1 complete genome
4	0.9650	Dengue type 1 genomic RNA complete genome strain 98901530
25	0.5938	Dengue type 1 strain D1.Myanmar.059/01
29	0.5714	Dengue virus type 1 genomic RNA complete genome
30*	0.5313	Dengue type 3 strain BR74886/02, complete genome
37†	0.5196	Dengue virus type 1 from Cambodia complete genome
64	0.4454	Dengue type 2 strain New Guinea
90	0.3711	Dengue type 3 strain Singapore 8120/95, complete genome
109	0.3571	Tick-borne encephalitis strain MDJ-01 nonfunctional
109	0.3571	Edge Hill strain V366 NS5 protein gene partial cds.
117	0.3090	Usutu strain Vienna 2001 from Austria complete genome
129	0.2912	Yellow fever strain Ivory Coast 1999 complete genome
158‡	0.2184	Dengue type 2 strain Ven2 polyprotein gene complete cds.
161	0.1545	West Nile strain 02002684 NS5 gene partial cds.
161	0.1545	Kunjin gene for polyprotein (C prM E NS1 NS2A . . .)
171	0.1485	Japanese encephalitis strain CC27-S8 complete genome
185	0.1456	Powassan 1427-62 NS5 gene complete cds.
186	0.0789	West Nile strain NY99-flamingo382-99 complete genome
338	0.0742	Japanese encephalitis FU strain complete genome

*Highest ranking non-DEN1 strain.

†Lowest ranking DEN1 strain.

‡Lowest ranking dengue strain of any serotype.
cds., coding region.

semiquantitatively by reference to Figure 1, where it is noted that the DEN2 and DEN3 clusters are roughly equidistant from DEN1. Again by reference to Figure 1, West Nile virus and Japanese encephalitis are predicted to have spectra that are the least similar to DEN1, and these are in fact ranked low in Table 1. As a benchmark to Table 1, we extracted the Scaramozzino amplicon from the known sequence of the experimentally analyzed DEN1 cDNA genome and performed BLAST searches of the entire NCBI nucleotide database to determine which organisms/viruses would be identified by conventional sequencing of this genomic subregion. Abbreviated results are shown in Table 2.

We also successfully typed DEN4 in the laboratory by our rapid mass-cataloging approach. In an analysis completely analogous to that in Table 1, Table 3 shows successful experimental typing of DEN4 using the Drosten interprimer region. Using the Drosten genomic region in conjunction with base-specific cleavage, 16 DEN4 strains all tied with a relative coincidence of 1.0. Serotypes 1, 2, and 3 all tied with much lower scores of 0.30.

Table 2. Benchmark Typing of Dengue by Conventional Sequencing and BLAST

ID rank	Virus name	(Bits)	Value
1	Dengue virus type 1 recombinant clone	513	8E-143
1	Dengue virus type 1 recombinant clone	513	8E-143
1	Dengue virus type 1 recombinant clone	513	8E-143
1	Dengue virus type 1 clone 45AZ5	513	8E-143
67	Dengue virus type 2 strain BR64022	180	1E-42
98	Dengue virus type 2 strain 100467	151	1E-33

Discussion

Bacterial identification using mass spectrometry can rely on the well-established utility of 16S ribosomal RNA. Because the mass-cataloging approach loses information that would be obtained by full sequencing, it was critical to the present effort to determine the extent to which the regions encompassed between the useful general primers are sufficiently discriminatory in distinguishing the target viral strains. Using a data set of *Flavivirus* genetic information representative of the majority of previously characterized strains, it was possible to predict that certain genomic regions would lend themselves to unambiguous identification and serotyping by the mass-cataloging approach. While we have used the term “serotype” in the context of the dengue virus due to its familiarity, this term is directly relevant to the extent that true genotype correlates with the observed immune interactions used for conventional dengue typing. Molecular methods such as the one described should allow much higher resolution identification and epidemiological tracking of various viral strains. The experimental results obtained here establish mass cataloging as a promising approach that may indeed be useful in the reliable identification of dengue as well other flaviviruses such as West Nile and St. Louis encephalitis. Furthermore, this identification is feasible within the context of all members of the family *Flaviviridae* predicted to be amplified by the primers used.

It should be emphasized that only a single cleavage reaction and only the sense strand were used in each case, and results obtained with different cleavage enzymes or different strands could provide independent information. If desired, therefore, even greater subtyping

Table 3. Mass Spectral Typing of Dengue-4 with Dengue-Specific Primers (from Drosten et al.¹⁵) by C, the Normalized Coincidence

ID rank	C	Virus name
1	1.0000	Dengue vector p4(Delta30) complete seq.
1	1.0000	Dengue type 4 vector p4 complete seq.
1	1.0000	Dengue type 4 strain ThD4_0485_01 complete genome
1	1.0000	Dengue type 4 strain Singapore 8976/95 complete genome
1	1.0000	Dengue type 4 strain 814669 complete genome
1	1.0000	Dengue type 4 recombinant clone rDEN4del30 complete
1	1.0000	Dengue type 4 recombinant clone rDEN4 complete seq.
1	1.0000	Dengue type 4 recombinant clone 2Adel30 complete seq.
1	1.0000	Dengue type 4 recombinant clone 2A complete genome
1	1.0000	Dengue type 4 polyprotein precursor gene complete coding region
11	0.3000	Dengue type 3 strain Singapore 8120/95, complete genome
11	0.3000	Dengue type 3 strain BR74886/02, complete genome
11	0.3000	Dengue type 3 strain 80-2 complete genome
11	0.3000	Dengue type 3 isolate PF94/136116, complete genome
11	0.3000	Dengue type 2 vector p4(delta30)-4995 complete seq.
11	0.3000	Dengue type 2 vector p4(delta30)-4995 complete seq.
11	0.3000	Dengue type 2 vector p2(delta30) complete seq.
11	0.3000	Dengue type 2 strain ThNH81/93 complete genome
11	0.3000	Dengue type 2 strain New Guinea
11	0.3000	Dengue type 2 strain Cuba89/97 complete genome
11	0.3000	Dengue type 2 China isolate 04 complete genome
11	0.3000	Dengue type 2 (S1 vaccine strain) complete genome
11	0.3000	Dengue type 1 strain Fj231/04 complete genome

seq., sequence.

resolution should be readily attainable by analysis of the antisense strand and/or by using alternate base-specific cleavage reactions.²⁵ While the endoribonuclease RNase T₁ is highly specific for cleavage after G residues, incomplete digestion products will have masses other than those predicted by complete sequence cleavage after every G. The intermediate product of endoribonuclease treatment, a 2'-3' cyclic phosphate intermediate, will have a mass 17 da less than that of the final product. Natural isotopic distribution also has a "smearing" effect on the acquired spectra, and for the purposes of our predictive calculations and spectral scoring, average masses are always used. For an in-depth discussion and simulations regarding these issues, see our previous related work.¹⁸ Additionally, it is known that at very high enzyme concentrations, nonspecific "over-cutting" by RNase T₁ can occur.²⁶ Other unexpected ion signals may arise from cleavage due to contamination by other ribonucleases and chemical fragmentation during the MALDI event itself. We have shown repeatedly that accurate microbial identification is possible despite the observance of spurious or unexpected masses (G.W. Jackson, unpublished observations). One reason for this is that many of the spurious masses are not found in the possible mass-basis space for any of the organisms in the databases. Another is that the enormous richness of base-specific fragment patterns (total number of possible fingerprints) is far greater than the number of species or strains under consideration. Finally, in practice, relatively heavy thresholding (5–10% relative ion abundance) is used to ignore much of the unexpected background with minimal loss of expected ions.

To compare masses from real spectra to large predicted databases, we replace the δ function in Equation 2 with the following:

$$2b \delta_{tol}(k) = \begin{cases} 1 & |k| \leq tol \\ 0 & otherwise \end{cases}$$

In practice the tolerance parameter, *tol*, is generally set to 1 da, although we have shown that the tolerance window can be widened substantially up to ~3.0 da with little or no effect on identification specificity (as greatly aided by amino-allyl U substitution to separate nearest neighboring compositions). Lowering the tolerance parameter below ~0.5 da can begin to drastically reduce overall coincidence scores as it is difficult to obtain this level of mass measurement accuracy across the entire mass range of interest with just a single point calibration. Nevertheless, as shown here, it is unnecessary to measure each oligoribonucleotide fragment mass with such high accuracy as might be afforded by higher resolution instruments; the success of the method ultimately relies on informatics to identify genetic regions of sufficient variability on a compositional basis and a high-throughput mass spectrometry technique of appropriate accuracy and resolution.

Recently we have also implemented thermal cycling conditions for faster PCR allowing greatly shortened overall reaction times (~35 minutes). Alternatively, capillary PCR devices (for example the RapidCycler, Idaho Technology) could be used for faster cycling, potentially reducing PCR times to 15 minutes. At present, the spectral scoring routine requires approximately 4 seconds per acquired spectrum. Thus, it is presently feasible to analyze a single clinical sample in approximately 1.5 hours or less with results similar to those shown here, and with additional informatics work extremely specific genotyping of any given *Flavivirus* should be possible. Moreover, all the steps in the protocol are amenable to high-throughput automated im-

plementation; we estimate that 384 samples could be processed in under 5 hours at a rate corresponding to greater than one identification per minute. Finally, with the development of miniaturized mass spectrometers,^{27–30} it will soon be possible to take such culture-independent molecular assays to remote locations to improve greatly the response time for epidemiological and clinical intervention.

In addition to speed, the primary advantage of the approach described here is that members of large groups of viral or bacterial pathogens can be identified using an identical protocol. Thus, it should be quite possible to develop analogous assays for other broad groups of viruses. All that is required are at least two conserved regions for targeting by broad-range PCR and an interprimer region that is discriminating enough to achieve the desired assay specificity. As motivated here, that specificity can be largely predicted *in silico*. Researchers interested in developing such assays for particular viral groupings will typically be familiar with such genetic regions as these same regions are likely to be most useful in distinguishing strains and deducing evolutionary relationships.

Acknowledgments

We thank Dr. Barry Falgout and Dr. Robin Levis of the Food and Drug Administration for the gift of the plasmids encoding the entire genomes of DEN-1 (WP) and DENV-4 (341750) in DNA form.

References

- Scaramozzino N, Crance JM, Jouan A, Debriel Da, Stoll F, Garin D: Comparison of flavivirus universal primer pairs and development of a rapid, highly sensitive heminested reverse transcription-PCR assay for detection of flaviviruses targeted to a conserved region of the NS5 gene sequences. *J Clin Microbiol* 2001, 39:1922–1927
- Kao CL KC, Chao DY, Wu HL, Chang GJJ: Laboratory diagnosis of dengue virus infection: current and future perspectives in clinical diagnosis and public health. *J Microbiol Immunol Infect* 2005, 38:5–16
- Monath TP: Dengue: the risk to developed and developing countries. *Proc Natl Acad Sci USA* 1994, 91:2395–2400
- Shu PY, Chen LK, Chang SF, Su CL, Chien LJ, Chin C, Lin TH, Huang JH: Dengue virus serotyping based on envelope and membrane and nonstructural protein NS1 serotype-specific capture immunoglobulin M enzyme-linked immunosorbent assays. *J Clin Microbiol* 2004, 42:2489–2494
- Halstead SB: Pathogenesis of dengue: challenge to molecular biology. *Science* 1988, 239:476–481
- Teles FR, Prazeres DM, Lima-Filho JL: Trends in dengue diagnosis. *Rev Med Virol* 2005, 15:287–302
- Putonti C, Chumakov S, Mitra R, Fox GE, Willson RC, Fofanov Y: Human-blind probes and primers for dengue virus identification: exhaustive analysis of subsequences present in the human and 83 dengue genome sequences. *FEBS J* 2006, 273:398–408
- De Paula SO, da Fonseca BAL: Dengue: a review of the laboratory tests a clinician must know to achieve a correct diagnosis. *Brazil J Infect Dis* 2004, 8:390–398
- Lancioti RS CC, Gubler DJ, Chang GJ, Vorndam AV: Rapid detection and typing of dengue viruses from clinical samples by using reverse transcriptase-polymerase chain reaction. *J Clin Microbiol* 1992, 30:545–551
- Harris ERT, Smith L, Selle J, Krammer LD, Valle S, Sandoval E, Balmaseda A: Typing of dengue viruses in clinical specimens and mosquitoes by single-tube multiplex reverse transcriptase PCR. *J Clin Microbiol* 1998, 36:2634–2639
- Wang WK ST, Tsai YC, Kao CL, Chang SM, King CC: Detection of dengue virus replication in peripheral blood mononuclear cells from dengue virus type 2-infected patients by a reverse transcription-real-time PCR assay. *J Clin Microbiol* 2002, 40:4472–4478
- De Paula SOD LC, Torres MP, Pereira MR, da Fonseca BAL: One-step RT-PCR protocols improve the rate of dengue diagnosis compared to two-step RT-PCR approaches. *J Clin Virol* 2004, 30:297–301
- Drosten CGS, Schilling S, Asper M, Panning M, Schmitz H, Gunther S: Rapid detection and quantification of RNA of Ebola and Marburg viruses, Lassa virus, Crimean-Congo hemorrhagic fever virus, Rift Valley fever virus, dengue virus, and yellow fever virus by real-time reverse transcription-PCR. *J Clin Microbiol* 2002, 40:2323–2330
- de Moraes Bronzoni R, Baleotti FG, Nogueira RMR, Nunes, M, Figueiredo: LTM duplex reverse transcription-PCR followed by nested PCR assays for detection and identification of Brazilian alphaviruses and flaviviruses. *J Clin Microbiol* 2005, 43:696–702
- Collins S: Automated MALDI mass spectrometry for rapid biological agent detection. In *Abstracts of Bio-Sensors for Homeland Security Conference*, October 24–28, 2005, College Station, TX
- Jackson GW, McNichols RM, Fox GE, Willson RC: Universal bacterial identification by mass spectrometry of 16S ribosomal RNA cleavage products. *Int J Mass Spectrom* 2007, 261:218–226
- Jackson GW, McNichols RJ, Fox GE, Willson RC: Bacterial genotyping by 16S ribosomal RNA mass cataloging. *BioMed Central Bioinformatics* 2006, 7(321)
- Zhang Z, Jackson GW, Fox GE, Willson RC: Microbial identification by mass cataloging. *BioMed Central Bioinformatics* 2006, 7(117)
- Lefmann M, Honisch C, Bocker S, Storm N, von Wintzingerode F, Schlotelburg C, Moter A, van den Boom D, Gobel UB: Novel mass spectrometry-based tool for genotypic identification of mycobacteria. *J Clin Microbiol* 2004, 42:339–346
- von Wintzingerode F, Bocker S, Schlotelburg C, Chiu NH, Storm N, Jurinke C, Cantor CR, Gobel UB, van den Boom D: Base-specific fragmentation of amplified 16S rRNA genes analyzed by mass spectrometry: a tool for rapid bacterial identification. *Proc Natl Acad Sci USA* 2002, 99:7039–7044
- Fulop L, Barrett ADT, Phillpotts R, Martin K, Leslie D, Titball RW: Rapid identification of flaviviruses based conserved NS5 gene sequences. *J Virol Methods* 1993, 44:179–188
- Thompson JD, Higgins DG, Gibson TJ: Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994, 22:4673–4680
- Kumar S, Tamura K, Nei M: MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings Bioinformatics* 2004, 5:150–163
- Trent DW, Grant JA, Monath TP, Manske CL, Corina M, Fox GE: Genetic variation and microevolution of dengue 2 virus in Southeast Asia. *Virology* 1989, 172:523–535
- Stanssens P, Zabeau M, Meersseman G, Remes G, Gansemans Y, Storm N, Hartmer R, Honisch C, Rodi CP, Bocker S, van den Boom D: High-throughput MALDI-TOF discovery of genomic sequence polymorphisms. *Genome Res* 2004, 14:126–133
- Woese C, Sogin M, Stahl D, Lewis BJ, Bonen L: A comparison of the 16S ribosomal RNAs from mesophilic and thermophilic bacilli: some modifications in the Sanger method for RNA sequencing. *J Mol Evol* 1976, 7:197–213
- English RD, Warscheid B, Fenselau C, Cotter RJ: Bacillus spore identification via proteolytic peptide mapping with a miniaturized MALDI TOF mass spectrometer. *Anal Chem* 2003, 75:6886–6893
- Gardner BD, Cotter RJ: A miniature MALDI-time-of-flight mass spectrometer using nonlinear ion optics for improved performance. In *Abstracts of Papers, 225th ACS National Meeting*, March 23–27, 2003, New Orleans, LA
- Ecelberger SA, Cornish TJ, Bryden WA: The improved teeny-TOF mass spectrometer for chemical and biological sensing. In *Third Harsh-Environment Mass Spectrometry Workshop and Second NASA/JPL Miniature Vacuum Pumps Workshop*, March 25–28, 2002
- Cotter RJ, Fancher C, Cornish TJ: Miniaturized time-of-flight mass spectrometer for peptide and oligonucleotide analysis. *J Mass Spectrom* 1999, 34:1368–1372