Follow-Up on Diagnostic Proficiency of Laboratories Equipped To Perform Orthopoxvirus Detection and Quantification by PCR: the Second International External Quality Assurance Study†

Matthias Niedrig,1 Hermann Meyer,2 Marcus Panning,3 and Christian Drosten3*

Robert Koch Institute, Berlin,1 Bundeswehr Institute of Microbiology, Munich,2 and Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany

Received 10 October 2005/Returned for modification 5 December 2005/Accepted 2 February 2006

Two years after the first external quality assurance study on bioterrorism-relevant viruses, we have conducted a follow-up study on orthopoxvirus detection by PCR. Thirty-three laboratories (27 European, 4 Austral-Asian, and 2 American) participated. Samples contained 0 to 40,000,000 DNA copies of lyophilized monkeypox, cowpox, and vaccinia virus per ml. Laboratories achieved a >80% detection chance above 56,234 copies per ml. Global sensitivity was not significantly improved over that of the first study. Twenty-seven and 9 participants, respectively, were able to genotype and quantify virus. Four of 27 genotyping results were incorrect. Quantification accuracy was significantly better for vaccinia virus than for the other viruses. False-positive results occurred in 22 (11.8%) of all 186 tests on negative samples, but 18 of these were contributed by only five laboratories. Fifty-five percent of laboratories could appropriately detect PCR inhibition. The use of either real-time PCR or commercial diagnostic kits had significant positive influence on laboratory performance.

Variola major virus (VAR) is the most relevant virus in the context of possible biological crime and terrorism. Although it has been eradicated from nature by vaccination, the virus is still stored in two selected World Health Organization repositories (2, 3). Its deliberate release could have catastrophic consequences in today's unvaccinated population (9). Other orthopoxviruses, like monkeypox (MPX), cowpox (CPX), and vaccinia virus (VAC), can imitate aspects of the clinical picture of VAR. Molecular virus detection tools for orthopoxviruses in general, and for VAR in particular, are essential for both biosecurity and public health bodies. The performance of such tools, however, is very heterogeneous in different laboratories; clinical evaluation is impossible in the absence of cases, and relevant strains for assay validation are not widely available.

External quality assurance (EQA) has improved the general diagnostic performance of laboratories for a number of viruses (7, 8, 13, 14, 17, 18). This is not only because EQA identifies weaknesses but also because it provides virus material which is necessary for assay optimization. Because no EQA program was available for rare and emerging viruses, we have recently organized the first proficiency studies in this field, focusing on EBola virus, Marburg virus, Lassa virus, and orthopoxviruses in one study and on severe acute respiratory syndrome coronavirus in another (5, 10).

To find out whether our prior studies had an effect on diagnostic quality, we have now conducted a new such study after 2 years. Focusing on orthopoxviruses this time, we have extended the scope of tested performance criteria. Sensitivity, rates of false-negative results, and the ability of laboratories to detect the presence of PCR inhibitors were queried in order to provide a comprehensive picture on how well laboratories are prepared for a real diagnostic scenario.

MATERIALS AND METHODS

Selection of participants. To be eligible for participation, institutions had to be operating on the national or supraregional level, and they had to be officially responsible for diagnosing poxvirus infections. Calls for participation were distributed through the European Network for the Diagnosis of Imported Viral Diseases (ENIVD) as well as through public health organizations (World Health Organization and Pan-American Health Organization). The study was announced as an EQA study on diagnostic proficiency, including certification and publication of results in a comparative and anonymous manner. All in all, 33 laboratories from 18 countries (27 European, 4 Austral-Asian, 2 American) enrolled in the study. Among these were 16 public health institutions, 10 universities with public health duties, two veterinary institutions, and five military facilities. One commercial laboratory also participated in addition, but it was not included in the evaluation because it did not fulfill public health or biosecurity duties. A full list of participants is given in the Acknowledgments section. A graphical overview is provided in Fig. S1 in the supplemental material.

Test specimens. To obtain test specimens, genotyped MPX, CPX, and VAC stocks were grown on Vero cells. Freeze-thawed preparations were heated to 56°C for 1 h, followed by gamma irradiation with 30 kGy to inactivate viruses. Residual infectivity was excluded by inoculation in Vero cell cultures (three passages). Inactivated virus stock solutions were aliquoted, lyophilized, and redissolved, and virus DNA was quantified by two different real-time PCR assays (11, 12). Quantification with other methods was not attempted after inactivation, but virion integrity was confirmed morphologically by electron microscopy. The final panel of test samples was generated by diluting inactivated virus stock solutions in human fresh-frozen plasma testing negative for human immunodeficiency virus type 1, hepatitis B virus, hepatitis C virus, and orthopoxviruses by PCR or reverse transcription-PCR. Aliquots of 100 μl each were lyophilized and shipped at ambient temperature to the participating laboratories. Each participant received a coded panel of seven virus-positive samples containing between 4,000 and 40,000,000 DNA copies per ml after resuspending in 100 μl of water. Six negative plasma samples were also included. A negative and a positive plasma sample were included which contained 10% (wt/vol) humic acid (Sigma, Munich, Germany), as well as one negative plasma sample containing 10,000 U/ml of heparin (Sigma, Munich, Germany). Both additives represented components...
which cause strong inhibitory effects in PCR (unpublished observations). The presence of a detectable inhibitory effect in both samples was confirmed in the reference laboratory using a commercial PCR kit (RealArt Orthopox PCR kit; Artus, Hamburg, Germany) after extracting nucleic acids with a commonly used method of sample preparation (QIAGEN silica-based columns).

Statistical analysis. All statistical analysis was carried out using the Statgraphics 5.1 software package (Manugistics, Dresden, Germany). Analysis of variance (ANOVA) analysis used type III sums of squares to obtain evaluations of each influence factor without the effect of the remaining factors.

RESULTS

Participants were asked to analyze the material with the molecular methods they routinely use in suspected human cases. Details about the utilized methods were requested, such as the sources of PCR primers and protocols, the type of extraction method used, and suppliers and types of commercial kits, if utilized.

Sensitivity. In our prior EQA study, the overall sensitivity of laboratories was characterized on the basis of the detection rate that was cumulatively achieved in all virus-containing samples. Seventy-eight percent of all tests in virus-containing samples and 44% of tests in samples containing less than 10,000 copies per ml were positive in the first study (10). In the current study, the corresponding numbers were 85% and 58%, respectively. However, as the composition of the test panel in the second study was not identical to that of the first one, these numbers were not directly comparable. To achieve a precise comparison, data from both studies were evaluated using a dose-response model (Probit analysis). As shown in Fig. 1, there was no obvious difference in overall sensitivities in both studies. In the earlier EQA, the average laboratory achieved detection with 50% probability when at least 1,380 copies per ml were present in a sample, compared to exactly 1,000 copies in the current study. Eighty percent detection chances were achieved at 97,723 and 56,234 copies per ml, respectively, in the old and the new study. When both data sets were compared in a Wilcoxon two-sample test, contrasting the input copy numbers required to achieve given detection chances, starting at 20% and ending at a 99.9% chance, the differences between the studies were not significant ($P = 0.694$). Using the same criteria, the 15 laboratories which had already participated in the earlier study did not achieve significantly higher sensitivities than the 18 first-time participants (Wilcoxon two-sample test; $P = 0.76$). Among the five laboratories which had achieved the worst results in the first study, three were now among the top performers, while two others still remained far below an acceptable level: no virus detection was achieved in any sample containing less than 11,000,000 virus genome copies per ml.

Genotyping. The capability of genotyping orthopoxviruses was not a necessary requirement for participation. However, because of the relevance of this task in clinical cases, all participants were asked to provide genotyping data if possible. Twenty-seven of 33 participants did so, using PCR product sequencing ($n = 10$), real-time PCR typing ($n = 18$), and/or restriction fragment length polymorphism (RFLP) analysis ($n = 4$) for typing.

A sample containing 28,000 copies of MPX was typed as CPX by three laboratories. One of these laboratories and one additional laboratory mistook a sample containing 11,000,000 copies of CPX for VAC. To test for the ability of laboratories to discriminate orthopoxviruses from other poxviruses, one sample containing a yatapox virus (tanapox) and one sample containing a parapoxvirus (orf virus) were also included in our sample panels. Of 27 laboratories, four mistook tanapox for VAC or an unspecified orthopoxvirus, and three mistook orf virus for either MPX, CPX, or an unspecified orthopoxvirus. All wrong genotyping results in total were contributed by seven laboratories, using either real-time PCR ($n = 2$), sequencing ($n = 1$), or RFLP analysis ($n = 4$) for typing.

Quantification. The relevance of poxvirus quantification in patients has not been evaluated clinically. However, such data have proven useful in the management of other rare virus infections in the past (16). Of all 33 participants in our study, 9 provided information on the virus concentration in test samples. Figure 2 summarizes a comparison of observed and expected quantification results. At the 95% confidence level, laboratories achieved quantification with a deviation of $\pm 1.4 \log_{10}$ from the expected value. Quantification of VAC was the most accurate, followed by MPX and CPX (significant differences in mean deviations from expected values; F test; $P = 0.037$). This observation is possibly associated with the fact that many laboratories used VAC for establishing their assays.

False positives. The study panels contained six negative plasma samples in total. All in all, 186 tests were performed on negative samples in the whole study. Of these, 22 tests (11.8%) were false positive. The 15 laboratories which participated for
from samples to a certain degree by efficient sample preparation. It can be eliminated by anticoagulant for blood samples. It can be eliminated by clinical diagnoses have to be reached by PCR (6). To test the inhibition control techniques applied by laboratories, also taking the effect of sample preparation methods into consideration. Twenty-six of 33 laboratories used silica-based affinity chromatography methods from major commercial manufacturers which were specified for use on viruses (20 laboratories used QIAGEN methods, 6 laboratories used those of Roche; no discrimination was made between manual and automated methods). Six laboratories used silica-based in-house protocols or commercial methods from other manufacturers which are not widely applied in the field. Only one laboratory used organic extraction with phenol-chloroform. ANOVA analysis on the three main categories of methods (QIAGEN, Roche, and in house) revealed a significantly smaller rate of inhibited samples in laboratories using in-house methods (8.3% samples) than in those using either QIAGEN or Roche methods (41% or 50%, respectively; the difference was not significant). Because in our hands commercial extraction methods are much more efficient at removing inhibitors than in-house methods, we concluded that those laboratories in the first group which did not detect an inhibitor probably missed it due to insufficient inhibition control technique. For the commercial assays, the presence of an inhibitor after extraction had been confirmed in the reference laboratory. It was thus assumed that with either of these methods an inhibitor should have been detected in both samples. Given this, 55% (18 laboratories) of participants could detect the strong inhibitory effect of humic acid, and only 12% (4 laboratories) were able to detect the weak effect of heparin in addition.

Influence factors. To determine which technical factors had general influence on laboratory performance, all applied variants of sample preparation and PCR methods were categorized (1, 4). The other sample contained humic acid, a substance which according to our experiences is notoriously difficult to remove by various methods of preparation.

Whether an inhibitor is detected depends on two factors: on the one hand, efficient sample preparation may remove the inhibitor; on the other hand, control techniques may fail to display an inhibitor due to breaches in technology, e.g., a too-high concentration of control DNA (15). This study intended to test the inhibition control techniques applied by laboratories, also taking the effect of sample preparation methods into consideration.

The correlation coefficient was 0.87 ($R^2 = 0.764$). (B) Deviations of determined DNA concentrations from the values expected according to prequantification data for three different orthopoxviruses.

the first time in our study contributed 20 of 22 positive results, which is a significantly higher rate than that in second-time participants (analysis of means; $P < 0.00001$). Five single laboratories alone were responsible for 18 false positives. Eliminating those from the evaluation, false-positive rates of first- and second-time participants were not significantly different anymore ($t$ test; $P = 0.59$). There was thus a small group of laboratories which was not proficient in the prevention of contamination, which is a general and basic skill in molecular diagnostics. The majority of participants had very good command of it.

**PCR inhibition**. Clinical samples may contain substances which interfere with PCR. Appropriate spiking of samples with internal or external inhibition controls is mandatory when critical diagnoses have to be reached by PCR (6). To test the ability of laboratories to detect PCR inhibitors, two negative plasma samples were spiked with substances of known inhibitory influence on PCR. One sample contained heparin, a common anticoagulant for blood samples. It can be eliminated from samples to a certain degree by efficient sample preparation.

### TABLE 1. Factors influencing the performance of laboratories

<table>
<thead>
<tr>
<th>Test used</th>
<th>No. of laboratories</th>
<th>Positive influence on sensitivity ($P$ value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major commercial DNA extraction kit$^a$</td>
<td>21</td>
<td>0.86</td>
</tr>
<tr>
<td>Automated DNA extraction platform$^b$</td>
<td>6</td>
<td>0.71</td>
</tr>
<tr>
<td>Nested PCR</td>
<td>4</td>
<td>0.28</td>
</tr>
<tr>
<td>Real-time PCR (in house)</td>
<td>15</td>
<td>0.007</td>
</tr>
<tr>
<td>Commercial PCR kit$^d$</td>
<td>8</td>
<td>0.002</td>
</tr>
</tbody>
</table>

$^a$ Includes viral RNA mini kit, DNA mini kit, DNA blood mini kit, and DNA stool mini kit (all Qiayen) as well as High Pure Viral Nucleic Acid kit (Roche).

$^b$ Includes BioRobot M48, EZ1, and MDx (all Qiayen) as well as MagnaPure (Roche).

$^c$ Analysis of variance by factor, eliminating the influence of other factors. Participants were asked to provide details on reagents and instruments used for sample preparation and amplification. Each participant was assigned a true (1) or not true (0) value for each of the five listed criteria. Using the observed sensitivities (as a fraction of 1) for each participant as the dependent variable, a multifactor analysis of variance using type III sums of squares was used to determine the significance of each of the five factors separately, eliminating the influence of the others.

$^d$ All eight participants used the RealArt Orthopox LC PCR kit (Artus).
rized according to five major criteria, as listed in Table 1. For each laboratory, the influences of these factors on the detection rates in positive samples were analyzed by multifactor ANOVA. Samples which contained more than 10^5 copies per ml were excluded, because they were unlikely to reflect general lack of sensitivity. Those five laboratories which reported false-positive results in more than half of all negative samples were also excluded: their results in low-positive samples may have been due to contamination rather than due to true virus detection. As shown in Table 1, the use of real-time PCR protocols or commercial test kits had a significant positive influence on sensitivity. This correlated well with observations made in earlier studies (5, 10). The use of commercial sample preparation methods from major suppliers did not provide significant influence, and neither did the use of automated platforms for sample preparation.

Of those five laboratories which had to be excluded due to high contamination rates, all used conventional (n = 4) or nested (n = 1) PCR methods. Three of them used noncommercial methods of sample preparation, such as phenol-chloroform or manual silica-based extraction.

**DISCUSSION**

This second international EQA study on PCR-based orthopoxvirus detection provides a follow-up on the laboratory proficiency in bioterrorism-related diagnostics. More participants have taken part in the second than in the first study, reflecting the growing number of laboratories which are taking over duties in this field.

Though the study was widely announced, mainly European laboratories participated, responding to a call for participation from the European organization ENIVD. However, since the diversity in Europe regarding regulations, funding, and training is comparably high, we believe that the European setting will be similar to many other regions of the world.

The major finding in our study is an only marginal improvement of performance: detection sensitivity in general remained at the same level, and second-time participants did not show better sensitivity than first-time participants. Other general diagnostic skills, such as the detection of inhibitory substances, have been systematically queried in this study for the first time. In view of experiences with molecular diagnostics of severe exotic infections (6), the rate of laboratories which were capable of detecting at least a strong PCR inhibitor was too low at 55%. Current methods of inhibition control should be adapted from published PCR-based assays for more common viruses. On the other hand, laboratories in general seemed to have good command in avoiding cross-contamination, which is probably the most important skill in this context. It is interesting that contamination was associated with only a few single laboratories; rates seemed to correlate positively with the application of non-real-time PCR methods and simple nucleic acid extraction procedures. Obviously, general improvements in PCR technology are taking effect, even in this very specialized field of molecular diagnostics. This is also documented by the fact that the majority of participants were able to genotype and quantify virus. Real-time PCR was the technical basis for quantification in the majority of laboratories (n = 23). Remarkably, quantification of VAC was more accurate than that of other viruses, probably because it is the only virus widely available for the establishing of assays. Genotyping was also done by real-time PCR in most laboratories, although results obtained from PCR product sequencing seemed more reliable. Interestingly, those laboratories that used classical RFLP genotyping generated the most typing errors.

The expertise for application of sophisticated molecular diagnostic technology is now available in most laboratories. However, the costs of such methods still exceed those of conventional techniques severalfold. In view of the described outcomes, investment into the improvement of methodology does seem well justified. On the other hand, most western countries have already provided generous funding for bioterrorism-related diagnostics. In this light, the observed outcome in second-time participants in our study appears disappointing, and improvements should be implemented very soon. Other issues in improving diagnostics, e.g., the availability of reference materials, cannot be accomplished as simply. Our study demonstrates the relevance of reference materials and provides a useful contribution in this area: all inactivated test samples are available for a nonprofit charge from ENIVD.

**ACKNOWLEDGMENTS**

This study was performed by the European Network for Diagnostics of Imported Viral Diseases (ENIVD), funded in part by the European Community DG SANCO under the program AIDS and other communicable diseases, grant no. SI2.299717(2000CVG4-26). Work of the Bernhard Nocht Institute was funded by the German Ministry of Health under grant no. 325-4539-85/3, the Bundesamt für Bevölkerungsschutz und Katastrophenhilfe (grant no. BBK F2-440-00-167/04), and the European Commission (grant no. SPSE-CT-2003-502567). The Bernhard Nocht Institute is Germany’s National Reference Centre for Tropical Infectious Diseases.

We are grateful to the following participants, listed by country, for making this study possible: INEI Anlis “Carlos G. Malbran,” Buenos Aires, Argentina; Victorian Infectious Diseases Reference Laboratory, Melbourne, Australia; Katholieke Universiteit Leuven, Belgium; Bureau of Microbiology Health Canada, Winnipeg, Canada; Statens Serum Institut, Copenhagen, Denmark; Haartman-Institute, Helsinki, Finland; UIBVE, Lyon, France; Unite de Virologie, CRSSA, La Tronche, France; Artus GmbH, Hamburg, Germany; Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany; Institut für Mikrobiologie der Bundeswehr, Munich, Germany; Institut für Lebensmittel, Arzneimittel und Tierseuchen, Berlin, Germany; Ludwigs-Maximilians-Universität, Munich, Germany; Medizinische Hochschule, Hannover, Germany; Niedersächsisches Landesgesundheitsamt, Hannover, Germany; Philippus Universität, Marburg, Germany; Robert Koch-Institut, Berlin, Germany; Wehrwissenschaftliches Institut für Schutztechnologien-ABC-Schutz, Munster, Germany; Aristotelian University, Thessaloniki, Greece; University College, Dublin, Ireland; Army Medical and Veterinary Research Center, Rome, Italy; Instituto Nazionale per le Malattie Infective, Rome, Italy; Erasmus University, Rotterdam, the Netherlands; RIVM, Bilthoven, The Netherlands; Military Institute of Hygiene and Epidemiology, Warsaw, Poland; Russian Academy of Sciences, Novosibirsk, Russia; VECOR, Koltsovo, Russia; University of Lubljana, Ljubljana, Slovenia; Instituto de Salud Carlos III, Madrid, Spain; Swedish Institute for Infectious Disease Control, Stockholm, Sweden; Spiez Laboratory, Spiez, Switzerland; Siriraj Hospital Mahidol University, Bangkok, Thailand; Health Protection Agency, Porton Down, United Kingdom; and Health Protection Agency, London, United Kingdom.

**REFERENCES**


