Microbial Biotechnology (2008) **1**(1), 79–86 doi:10.1111/j.1751-7915.2007.00012.x

Massively parallel pathogen identification using high-density microarrays

Nicolas Berthet,[1] Philip Dickinson,[2] Ingrid Filliol,[1] Anita K. Reinhardt,[1] Christophe Batejat,[3] Tatiana Vallaeys,[4] Katherine A. Kong,[2] Christopher Davies,[2] Walter Lee,[2] Shenglan Zhang,[2] Yaron Turpaz,[2] Beate Heym,[5] Gilberte Coralie,[3] Laurent Dacheux,[6] Ana Maria Burguière,[3] Hervé Bourhy,[6] Iain G. Old,[7] Jean-Claude Manuguerra,[3] Stewart T. Cole[8] and Giulia C. Kennedy[2]*

[1]*Genotyping of Pathogens and Public Health Technological Platform,* [3]*Laboratory for Urgent Response to Biological Threats,* [4]*Biological Software and Databases,* [6]*Lyssavirus Dynamics and Host Adaptation Unit,* [7]*European Office and* [8]*Bacterial Molecular Genetics Unit, Institut Pasteur, Paris, France.*

[2]*Affymetrix, 3420 Central Expressway, Santa Clara, CA, USA.*

[5]*Ambroise-Paré Hospital, Boulogne-Billancourt, France.*

Summary

Identification of microbial pathogens in clinical specimens is still performed by phenotypic methods that are often slow and cumbersome, despite the availability of more comprehensive genotyping technologies. We present an approach based on whole-genome amplification and resequencing microarrays for unbiased pathogen detection. This 10 h process identifies a broad spectrum of bacterial and viral species and predicts antibiotic resistance and pathogenicity and virulence profiles. We successfully identify a variety of bacteria and viruses, both in isolation and in complex mixtures, and the high specificity of the microarray distinguishes between different pathogens that cause diseases with overlapping symptoms. The resequencing approach also allows identification of organisms whose sequences are not tiled on the array, greatly expanding the repertoire of identifiable organisms and their variants. We identify organisms by hybridization of their DNA in as little as 1–4 h. Using this method, we identified Monkeypox virus and drug-resistant *Staphylococcus aureus* in a skin lesion taken from a child suspected of an orthopoxvirus infection, despite poor transport conditions of the sample, and a vast excess of human DNA. Our results suggest this technology could be applied in a clinical setting to test for numerous pathogens in a rapid, sensitive and unbiased manner.

Received 14 September, 2007; accepted 28 September, 2007. *For correspondence. E-mail giulia_kennedy@affymetrix.com; Tel. (+1) 408 731 5562; Fax (+1) 408 481 0435.

Introduction

Infectious diseases remain a leading cause of morbidity and mortality in humans. Rapid and accurate identification of the aetiological agent, be it an old or an emerging pathogen, can influence prognosis by enabling appropriate interventions, such as drug treatment, to be promptly initiated.

Current methods for pathogen identification possess a number of advantages and shortcomings with respect to clinical utility. For example, nucleic acid detection by quantitative PCR (qPCR) is extremely rapid and sensitive, but provides simultaneous results on a limited number of primer pair-determined genomic regions. In contrast, traditional culture-based, phenotypic methods are often limited by lengthy incubation periods, but allow for the identification of multiple pathogens and are currently considered the gold standard in clinical microbiology laboratories.

A disadvantage of most nucleic acid-based methods is the need for a hypothesis to guide testing as they require DNA amplification with a set of sequence-specific primers designed to amplify only the desired nucleic acids in the specimen. Thus, in cases where there is no clear suspect, testing with serial methods such as qPCR will become onerous; in addition, if the hypothesis is wrong, a negative result will be uninformative. At the other end of the spectrum, methods for completely unbiased identification of organisms, such as *de novo* genome sequencing, are slow, costly, resource intensive and therefore not amenable to clinical settings.

We sought to devise a protocol that maximizes the advantages of all these approaches, while minimizing the less desirable characteristics. We describe here a generic, highly parallel method to rapidly identify multiple pathogens and their genetic elements in an unbiased manner and without the need for a specific hypothesis.

Recently, DNA microarrays have been used for the simultaneous detection and identification of numerous pathogens using locus-specific amplification methods

(Bryant *et al.*, 2004). Several microarray-based assays have been developed and employed to track or identify multiple pathogenic microorganisms, such as bacteria (Gingeras *et al.*, 1998; Troesch *et al.*, 1999; Wilson *et al.*, 2002; Vora *et al.*, 2004; Davignon *et al.*, 2005), including potential biowarfare agents such as *Bacillus anthracis* (Zwick *et al.*, 2005), respiratory pathogens in clinical samples (Lin *et al.*, 2006) and miscellaneous viruses (Wang *et al.*, 2003; Wong *et al.*, 2004; Palacios *et al.*, 2007). However, none of these methods provides comprehensive information about the pathogen at the single nucleotide level. Consequently, we devised a diagnostic protocol that employs microarray-based resequencing technology (described in Fodor *et al.*, 1993; 1991; Pease *et al.*, 1994; Chee *et al.*, 1996; Hacia, 1999; Cutler *et al.*, 2001; Warrington *et al.*, 2002) to meet the following requirements: it should be unbiased, requiring no prior information, universal, highly specific and able to discriminate between pathogens that impart similar or overlapping symptoms. In addition, the approach should be comprehensive, sensitive, not influenced by human DNA, and in this case detecting agents on the National Institute of Allergy and Infectious Disease (NIAID) Category A, B and C Priority pathogen list. Rapid turnaround is also important and ideally results should be generated the same day.

For each base in the reference sequence, a resequencing microarray incorporates eight features consisting of 25-base oligonucleotides centred at the base: four equal to the reference with A, C, G or T at centre and four equal to the reverse complement with A, C, G or T at centre. Analysing the intensity from these multiple features permits detection of the reference sequence as well as characterization of relative Single-nuclectide-polymorphisms (SNPs). Our results indicate that resequencing microarrays are capable of accurate identification and characterization of pathogens, without the need for locus-specific PCR amplification. This rapid, comprehensive assay provides clinically relevant information which can be used to accelerate treatment decisions, and ultimately improve the medical care of patients with bacterial or viral infections.

Results

Detection in complex mixtures

We studied three separate complex mixtures of pathogens which confer overlapping clinical symptoms and which might pose a challenge for diagnosis. The first mixture included DNA from pathogens associated with respiratory syndromes, the second DNA from pathogens associated with meningitis and encephalitis, and the third DNA from organisms associated with septicaemia. All three mixtures contained a background of human DNA to mimic a clinical specimen: DNA prepared from nasal washes for the respiratory mixture, cerebrospinal fluid (CSF) for the meningitis–encephalitis mixture or a human cell line (HeLa) for the septicaemia mixture. Each mixture was tested on the array and the resulting sequences were analysed. The three experiments detect and sequence 48, 11 and 25 genetic fragments respectively (Table 1). The sequence for each conserved element underwent BLAST analysis in which the highest scoring matches determine the identity. The antibiotic resistance (ABR) elements predict likely resistance to their associated antibiotics. Finally, the pathogenicity genes represent the various toxins and virulence factors that the pathogen can generate. Results shown in Table 1 indicate that despite the presence of background nucleic acids of human origin, all expected organisms were detected. We also tested DNA extracted from normal nasal wash, CSF and sputum samples to determine the hybridization pattern of background organisms. We detected *Neisseria* 16S, *Pseudomonas* 16S and human 18S and mitochondrial DNA in the nasal wash. No specific organisms were found in the CSF sample. We identified *Haemophilus* 16S, *Bacillus* 16S and human 18S and mitochondrial DNA in sputum, which is a very complex sample with a large bacterial load and complexity due to normal flora.

Antibiotic resistance

We tested the antibiotic resistance profile of a strain of *Enterococcus faecium* using the multi-pathogen microarray and compared it with results obtained using an antibiogram, a culture-based phenotypic reference method. The antibiotic resistance profile determined by the antibiogram was confirmed by results from the multi-pathogen microarray (Table 2). Furthermore, inclusion of genes on the array, such as *tetM* or *vanA*, allowed identification of the different resistance mechanisms utilized by the strain.

In addition to detecting antibiotic resistance caused by the presence of specific ABR genes, the resequencing design of the pathogen microarray can also identify drug resistance caused by non-synonymous mutations that confer antibiotic resistance. For example, pefloxacine resistance in *Staphylococcus aureus* can be conferred by a mutation in the *parC* gene. We detected a single-base-pair mutation in a portion of the *parC* gene in *Staphylococcus haemolyticus* which results in an amino acid change and subsequent resistance to pefloxacine (Fig. 1).

In another set of experiments, six bacterial strains isolated from different clinical samples were tested in a double-blind experiment. All these strains were correctly identified using the 16S rRNA sequence and other house-

Table 1. Pathogen detection in complex mixtures, in the presence of clinically relevant background DNA.

Experiment	Organism	Type	Genetic elements detected
Respiratory	*Staphylococcus aureus*	Conserved	*16S rRNA, gyrA, gyrB, rpoB*
		ABR	*gyrA-abr, gyrB-abr, parC-abr, parE-abr, rpoB-abr, aac6-aph2, ant6-Ia, ant9-Ia, aph3-IIIa, blaZ, dfrB, ermA, mecA, tetM, vatB, vgaB, vgaC*
	Francisella tularensis	Conserved	16S rRNA
		ABR	*gyrA-abr, gyrB-abr, parC-abr, parE-abr*
	Yersinia pestis	Conserved	*16S rRNA, gyrA, gyrB, rpoB*
		ABR	*gyrA-abr, gyrB-abr, parC-abr, parE-abr*
		Pathogenicity and virulence	*yopT, pla, psaA, yopM, cafA, yplA, yplB, invA, irp2, yadA*
	Influenza B virus	Conserved	RNA polymerase
	SARS virus	Conserved	RNA polymerase
	Rattus norvegicus	Conserved	18S rRNA
	Homo sapiens	Conserved	Mitochondrial fragment
Encephalitis	*Neisseria meningitidis*	Conserved	16S rRNA
		ABR	*gyrA-abr, gyrB-abr, parC-abr, parE-abr, erm, cat, aph*
	Rabies virus	Conserved	RNA polymerase
	Vesicular stomatitis virus	Conserved	RNA polymerase
	Homo sapiens	Conserved	18S rRNA
Septicaemia	*Staphylococcus aureus*	Conserved	*16S rRNA, gyrA, gyrB, rpoB*
		ABR	*gyrA-abr, gyrB-abr, parC-abr, parE-abr, rpoB-abr, aac6-aph2, ant6-Ia, ant9-Ia, aph3-IIIa, blaZ, dfrB, ermA, mecA, tetM, vatB, vgaB, vgaC*
	Dengue virus type II	Conserved	Polyprotein
	Rift Valley fever virus	Conserved	RNA polymerase
			18S rRNA
	Homo sapiens	Conserved	Mitochondrial fragment

Respiratory, encephalitis and septicaemia syndrome experiments contain nasal washes, cerebrospinal fluid and HeLa DNA as background. ABR, antibiotic resistance; elements suffixed with '-abr' are regions of housekeeping genes in which non-synonymous point mutations may confer antibiotic resistance. SARS, severe acute respiratory syndrome. The -abr suffix indicates a region of the gene monitored for antibiotic resistance-conferring mutations.

keeping genes such as *rpoB* or *gyrB*. Globally, the correspondence between antibiotic resistance genes and point mutations detected by the microarray and resistance phenotypes determined by an antibiogram was excellent. Indeed, as shown in Table 3, resistance phenotypes resulting from acquired genes or point mutations were identified successfully with the microarray, including the point mutation in the *parC* gene conferring fluoroquinolone resistance to the *S. haemolyticus* strain. The non-detection of resistance phenotypes was due either to the absence of the gene sequence tiled on the array (such as the non-detection of rifampicin resistance for the *S. haemolyticus* strain), or to resistance mechanisms likely caused by overexpression of an efflux pump. The microarray approach allows us to identify antibiotic resistance genes, such as *ant9-I* (conferring spectinomycin resis-

Table 2. Comparison of the antibiotic resistance profile for astrain of *Enterococcus faecium* determined using the pathogen microarray and the antibiogram.

Class	Antibiotic	Antibiogram	Genes	Microarray
Aminoglycoside	Gentamicin	Susceptible	*aac(6′)-aph(2″)-Ia*	Not detected
			aac(6′)-aph(2″)-Ib	Not detected
			aac(6′)-aph(2″)-Ic	Not detected
			aac(6′)-aph(2″)-Id	Not detected
			aac(6′)-aph(2″)-Ie	Not detected
	Kanamycin	Resistant	*aac(6′)-aph(3′)-IIIa*	Detected
	Streptomycin	Resistant	*ant(6)-Ia*	Detected
	Tobramycin	Not done	*aac(6′)-Ii*	Detected
Glycopeptide	Vancomycin	Resistant	*vanA*	Detected
	Teicoplanin	Resistant	*vanA* to *vanG*	*vanA* Detected
Tetracyclin	Tetracyclin	Resistant	*tet(K)*	Not detected
			tet(L)	Not detected
			tet(M)	Detected
			tet(O)	Not detected
			tet(S)	Not detected

The Class and Antibiotic columns describe the antibiotic compounds. The Antibiogram column shows the effect of the antibiotic measured using a classical antibiogram. The Genes column indicates genes associated with resistance to the antibiotic. Finally, the Microarray column shows whether the microarray detected the gene.

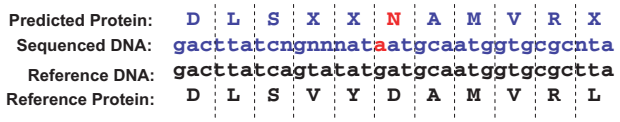

Fig. 1. Antibiotic resistance conferred by a point mutation. The centre amino acid for the measured sequence is different from the reference sequence. This non-synonymous mutation is known to confer resistance to pefloxacine in *S. aureus*.

tance), which are not tested routinely in an antibiogram. The results we obtained confirm the ability of the microarray to detect acquired genes accurately as well as point mutations when the corresponding sequence was tiled on the microarray. A limitation of this approach, however, is that in cases where multiple bacteria are present, it may not be possible to associate the resistance genes conclusively with their respective organism.

Identification of pathogens from a wound sample

A wound sample was analysed with our microarrays. All eukaryotic and bacterial DNA were extracted and amplified via two successive Ultra Fast whole-genome amplification (WGA) reactions. The *S. aureus* DNA responsible for this wound infection represented 1.35% of the total DNA extracted. BLAST analysis of sequences obtained from the hybridized array revealed the presence of *S. aureus* and one or more members of the bacterial class *Betaproteobacteria*, which are known commensal organisms in wound samples. BLAST analysis also identified four antibiotic resistance genes: *blaZ*, *ermC*, *mph* and *tetK*.

Table 3. Comparison of the antibiotic resistance profiles determined for six clinical samples using an antibiogram and the pathogen microarray for six clinical strains.

Identified pathogen	Antibiotic resistance determined by antibiogram	Antibiotic resistance genes and mutations detected by microarray
Staphylococcus haemolyticus	Penicillin	*blaZ*
	Methicillin	*mecA*
	Gentamycin/tobramycin	*aac6 aph2*
	Erythromycin	*ermC*
	Lincosamide/pristinamycin	*vgaB, vatB, vgaC*
	Trimethoprime	*dfrA*
	Fluoroquinolones	*parC* mutation
	Rifampicin	(*Staphylococcus haemolyticus rpoB* not on chip)
Enterococcus faecalis	Lincosamide	*lnuB1*
	Aminoglycosides	*aac6-aph2, aph3-IIIa, ant6-Ia*
	Erythromycin	*ermB, msrC*
	Pristinamycin	*lsaC*
	Tetracycline	*tetM*
Escherichia coli	Cotrimoxazole	*dhfrI*
	Tetracycline	*tetB*
	Chloramphenicol	*catI*
	Penicillin	(β-Lactamases conferring resistance to penicillin for *E. coli* not on chip)
	Fluoroquinolones	(No point mutation in quinolone resistance-determining regions found – resistance likely due to overexpressed efflux pump)
	(Streptomycin and spectinomycin were not tested)	*ant3-Ia*
Citrobacter freundii	Chromosomal cephalosporinase	*ampC*
	Pencillinase	*tem*
	Chloramphenicol	*catI*
	Cotrimoxazole and Fluoroquinolones	*aadA-1a*
Staphylococcus aureus	Penicillin	*blaZ*
	Methicillin	*mecA*
	Kanamycin and tobramycin	*ant4-Ia*
	Erythromycin/lincosamide	*ermA*
	(Streptomycin and spectinomycin were not tested)	*ant9-Ia*
	Susceptible to sulfa-trimethoprim-cotrimoxazole	No mutation in *dfrB*
	Fluoroquinolones	*parC* mutation
Campylobacter jejuni	Tetracycline	*tet O*
	Kanamycin	*aph3-IIIa*
	Amoxicillin	(β-Lactamases conferring resistance to amoxicillin for *E. coli* not on chip)
	Fluoroquinolones	(No point mutation in quinolone resistance-determining regions found – resistance likely due to overexpressed efflux pump)

The Identified pathogen column represents the species determined by the chip, which was correct in all cases. The other columns compare the antibiotic resistance measured by an antibiogram to antibiotic resistance genes and mutations discovered using the microarray.

Antibiogram results confirmed that the *S. aureus* strain was resistant to penicillin, consistent with the presence of *blaZ*. The detected *tetK*, *ermC* and *mph* genes were likely derived from the commensal organisms rather than the *S. aureus*.

Distinguishing between closely related organisms

Most molecular techniques are designed to provide a positive or negative result for a specific pathogen. The test may not be designed to detect – and distinguish – a closely related organism. The resequencing-based approach of the pathogen array tiles probes with all four nucleotides at each position, making it possible to detect point mutations in the sample relative to the reference sequence. This feature greatly expands the repertoire of identifiable organisms far beyond those that are actually included on the array.

For example, the Smallpox virus is a member of the *Poxviridae* family, which comprises, among others, the closely related Monkeypox, Camelpox and Vaccinia viruses. Smallpox, a once devastating disease, shares several key symptoms with Monkeypox, e.g. the formation of skin pustules. While Smallpox has been declared eradicated by the World Health Organization, the Smallpox virus could be used in a bioterrorism threat. In this case, it would be important to rapidly distinguish this organism from its close relative, Monkeypox virus, which causes a far less serious disease. To test the ability of the array to distinguish closely related pathogens in a clinical sample, we selected material from an African child suffering from a pustular infection of unknown aetiology, but where Monkeypox virus was suspected (Damon *et al*., 2006).

DNA was extracted from a skin sample and amplified using WGA with generic random hexamers. The resulting target was then hybridized to the multi-pathogen microarray and nucleotide sequences were determined from the signals as described (Cutler *et al*., 2001), and compared with sequences in GenBank. The results from a portion of the RNA polymerase gene are shown in Fig. 2. Sequence analysis of the clinical sample and comparison with GenBank sequences of various orthopoxviruses revealed several SNPs which clearly distinguish the agent detected in the clinical sample from the Smallpox virus and confirm its identity as Monkeypox virus (Fig. 2).

As expected, analysis of other sequences obtained from the array revealed that the samples also contain human DNA, as evidenced by the 18S ribosomal RNA gene and human mitochondrial sequences. Sequence analysis also revealed the presence of *S. aureus*, a normal skin commensal organism (see Appendix S2 in *Supplementary material* for further details). Detailed sequence analysis of the pathogenicity profile of the *S. aureus* sequences determined by the multi-pathogen microarray revealed seven genes potentially conferring resistance to antibiotics of the β-lactam (*blaZ*), macrolide (*ermC*), aminoglycoside [*aph(6)-Id* and *aph(3′)-Ib*], chloramphenicol (*catA-7*) and tetracycline (*tetK* and *tetM*) families, but no quinolone or methicillin resistance genes. The simultaneous detection of Monkeypox virus and multidrug-resistant *S. aureus* in the clinical sample was not hampered by a vast excess of human DNA. As *S. aureus* was the only bacteria detected in the sample, the antibiotic resistance genes identified were presumed to derive from *S. aureus*.

We confirmed the Monkeypox results with qPCR, a method that is often used by reference laboratories to detect pathogens in clinical specimens. Using specific primers, a fragment of the haemagglutinin gene was amplified by real-time PCR, and the presence of SNPs that distinguish particular orthopoxviruses was inferred from the melting curve peaks. Using known quantities of plasmid DNA to generate standard curves, the viral load in the scab was determined to be 4.32×10^4 genome copies per microlitre. From this result, it was estimated that only 0.025% of the total DNA extracted from the clinical sample was of viral origin.

Our generic approach allowed identification of the suspected agent in a clinical setting, and uncovered a drug-resistant strain of *S. aureus*. Following a strictly hypothesis-based approach, the Monkeypox virus would have been detected; however, the *S. aureus* would have likely gone unnoticed.

Sensitivity

We performed titration experiments with known quantities of *S. aureus*, both with and without a vast excess of human genomic DNA, to determine the limits of detection using the current assay and base-calling algorithms. Results are shown in Appendix S1 in *Supplementary material*. In the presence of more than a 100-fold mass excess of human DNA, we can detect as little as 3000 genome copies of *S. aureus*. In the absence of human DNA background, we detect as little as a single genome copy of *S. aureus*.

Speed

In certain life-threatening conditions the time to diagnosis can be critical. While molecular techniques for pathogen identification suffer from lack of parallelism and the need to test for a suspected pathogen, they are nevertheless very rapid and can produce results in a matter of a few hours. Although microarray experiments typically require 48 h to generate results, we set out to determine a set of conditions that would allow us to combine the high parallelism of microarray analysis with the speed of PCR-


```
Variola          TCTAAGGTTACGTATAGTCTATATGATCAAAAAGAGATTAATGCTACAGATATTATCATT 60
Camelpox         TCTAAGGTTACGTATAGTCTATATGATCAAAAAGAGATTAATGCTACAGATATTATCATT 60
Vaccinia         TCTAAGGTTACGTATAGTCTATATGATCAAAAAGAGATTAATGCTACAGATATTATCATT 60
Cowpox           TCTAAGGTTACGTATAGTCTATATGATCAAAAAGAGATTAATGCTACAGATATTATCATT 60
Monkeypox        TCTAAGGTTACGTATAGTCTATACGATCAAAAAGAGATTAATGCTACAGATATTATCATT 60
CIBU05-335       TCTAAGGTTACGTATAGTCNATACGATCAAAAAGAGATTAATGCTACAGATATTATCATT 60
                 ****************** *** ************************************

Variola          AGTCATGTTAAAAATGACGACGATATCGGTACCGTTAAAGATGGTAGACTAGGTGCTATG 120
Camelpox         AGTCATGTTAAAAATGACGACGATATCGGTACCGTTAAAGATGGTAGACTAGGTGCTATG 120
Vaccinia         AGTCATGTTAAAAATGACGACGATATCGGTACCGTTAAAGATGGTAGACTAGGTGCTATG 120
Cowpox           AGTCATGTTAAAAATGACGACGATATCGGTACCGTTAAAGATGGTAGACTAGGTGCTATG 120
Monkeypox        AGTCATATTAAAAATGACGACGATATCGGTACCGTTAAAGATGGTAGACTAGGTGCTATG 120
CIBU05-335       AGTCATATTAAAAATGACGACGATATCGGTACCGTTAAAGATGGTAGACTAGGTGCTATG 120
                 ****** ************************************************** 

Variola          GATGGGGCATTATGTAAGACTTGTGGGAAAACGGAATTGGAATGTTTCGGTCACTGGGGT 180
Camelpox         GATGGGGCATTATGTAAGACTTGTGGGAAAACGGAATTGGAATGTTTCGGTCACTGGGGT 180
Vaccinia         GATGGGGCATTATGTAAGACTTGTGGGAAAACGGAATTGGAATGTTTCGGTCACTGGGGT 180
Cowpox           GATGGGGCATTATGTAAGACTTGTGGGAAAACGGAATTGGAATGTTTCGGTCACTGGGGT 180
Monkeypox        GATGGGGCATTATGTAAGACTTGTGGGAAAACGGAATTGGAATGTTTCGGTCACTGGGGT 180
CIBU05-335       GATGGGGCATTATGTAAGACTTGTGGGAAAACGGAATTGGAATGTTTCGGTCACTGGGGT 180
                 ************************************************************

Variola          AAAGTAAGTATCTATAAAACTCATATAGTTAAGCCTGAATTTATTTCAGAAATTATTCGT 240
Camelpox         AAAGTAAGTATCTATAAAACTCATATAGTTAAGCCTGAATTTATTTCAGAAATTATTCGT 240
Vaccinia         AAAGTAAGTATTTATAAAACTCATATAGTTAAGCCTGAATTTATTTCAGAAATTATTCGT 240
Cowpox           AAAGTAAGTATTTATAAAACTCATATAGTTAAGCCTGAATTTATTTCAGAAATTATTCGT 240
Monkeypox        AAAGTAAGTATTTATAAAACTCATATAGTTAAGCCTGAATTTATTTCAGAAATTATTCGT 240
CIBU05-335       AAAGTAAGTATTTATAAAACTCATATAGTTAAGCCTGAATTTATTTCAGAAATTATTCGT 240
                 *********** ************************************************

Variola          TTACTGAATCATATATGTATTCACTGCGGATTATTG 276
Camelpox         TTACTGAATCATATATGTATTCACTGCGGATTATTG 276
Vaccinia         TTACTGAATCATATATGTATTCACTGCGGATTATTG 276
Cowpox           TTACTGAATCATATATGTATTCACTGCGGATTATTG 276
Monkeypox        TTACTGAATCATATATGTATTCATTGCGGATTATTG 276
CIBU05-335       TTACTGAATCATATNTGTATTCATTGCGGATTATTG 276
                 ************** ******** ************
```

Fig. 2. Identification of Monkeypox virus. Comparison of reference sequences for the RNA polymerase in Variola, Camelpox, Vaccinia, Cowpox and Monkeypox with the measured sequence. The polymorphisms marked with yellow permit identification of the measured sample as Monkeypox.

based techniques. The inverse relationship between concentration and microarray hybridization time has been well established for PCR products, but has not been reported for complex genomic samples. Based on results from human genomic DNA hybridizations to high-density microarrays (Kennedy *et al.*, 2003), we reasoned that hybridization times for whole-genome mixtures of pathogens should be amenable to short hybridization times as long as appropriate amounts of target are hybridized.

We therefore tested the effect of hybridization time on samples containing 10 μg of *S. aureus*. We found that hybridization times as short as 15 min can provide sufficient signal to identify the genus and species of this organism, and hybridization times of as little as 1 h can unambiguously identify the antibiotic resistance spectrum (see Appendix S1 in *Supplementary material* for details). Organisms with higher G+C content, such as *Vibrio vulnificus*, require longer hybridization times, due to higher background, but this effect can be improved partially by increasing the amount of target hybridized to the array (data not shown).

Discussion

Molecular diagnostic approaches afford a number of advantages over the traditional, phenotypic methods but

have only recently begun to find applications in a clinical setting. Quantitative PCR requires the design and optimization of sequence-specific primer pairs for amplification and detection. Knowledge of the suspected pathogen is therefore required to guide testing of a clinical sample. Information obtained from qPCR is restricted to the presence or absence of the pathogen tested, and no additional information is obtained.

In contrast, high-density oligonucleotide microarrays allow determination of nucleotide sequence at > 99.9% accuracy for up to 300 000 double-stranded base pairs of DNA sequence information in a single experiment (Cutler *et al*., 2001). The multi-pathogen microarray contains sequences representing 962 genes, including conserved genes, pathogenicity and virulence genes and antibiotic resistance genes. Furthermore, the ability to determine the nucleotide sequence at single-base-pair resolution allows the potential identification of pathogens with similar, but not identical, sequences to those represented on the microarray. For example, although the pathogen *Staphylococcus epidermidis* is not represented on the array, we identified it from a sample by resequencing a portion of the 16S rRNA gene for *S. aureus* (data not shown). As further exemplified by results from the scab samples, identification of the Monkeypox virus, which was not included on the array, was possible through SNP detection from the Smallpox virus, which was present. In addition to its potential use in identifying known pathogens in a clinical setting, the multi-pathogen array could theoretically be used to discover new organisms and their variants. Thus, a single application of sample to the multi-pathogen microarray potentially replaces thousands of individual PCR reactions. Similarly, the highly multiplexed microarray-based assay can replace hundreds of culture-based tests.

From a clinical microbiologist's perspective, the survival and virulence genes are as much a part of the identity of the pathogen as the genus and species. Hence, it is important to look beyond the conserved genes that determine the species and examine the antibiotic resistance profile to inform treatment decisions. Our results demonstrate the ability to detect a large number of genetic elements, even within a complex mixture, and in the presence of large amounts of human DNA. At present, the pathogen identification method described here is limited by sequences deposited in the NCBI non-redundant sequence database. With the increasing availability of sequence data for new organisms, these databases will improve and identification of more organisms in clinical samples will be facilitated.

We tested and successfully identified many pathogens on the array individually (see Appendix S1 in *Supplementary material*). However, an important feature of unbiased pathogen detection in a clinical setting is the ability to identify organisms in mixtures. We created mixtures of pathogens and show that the multi-pathogen array is capable of deconvoluting and correctly identifying these organisms. In addition to the mock clinical specimens, the scab sample illustrates the ability to detect bacteria, associated ABR genes and viral DNA, all in the same test. Their detection was not influenced by an excess of human DNA in the sample, demonstrating the discriminatory power of the tool and the strength of pathogen detection in a complex environment. Testing samples from different mock syndromes demonstrated the ability of this new diagnostic tool to distinguish between two viruses and three bacteria even when these organisms were phylogenetically close. Moreover, the ability to perform BLAST analysis with each sequence obtained allowed us to identify pathogens with confidence, to eliminate all risk of cross-hybridization and to detect pathogens that could not be detected by classical methods.

Compared with many protocols for DNA microarrays, which use overnight hybridization, qPCR allows extremely rapid pathogen identification – in only 2 h. We have optimized our multi-pathogen microarray assay (data not shown) to allow positive identification after only 2–4 h of hybridization time: a significant reduction in the overall assay time to under 10 h. It is therefore possible to carry out pathogen identification using this technique in the course of a single day or less. Because the microarray potentially replaces thousands of individual PCR reactions, this translates to a major reduction in cumulative assay time. Moreover, the speed of the microarray-based assay is a significant improvement over culture-based methods, which can take days to weeks to produce results.

The technology described here has potential future utility in a clinical setting. While the exploratory chips used in this study contain 2.56 million probes and were deliberately designed to over-represent genetic information, further iterations may involve selecting only those probes that convey the highest discriminatory value in identifying pathogens. We envision that future arrays will contain a smaller subset of probes and will therefore be smaller in size and consequently more economical. An additional requirement for clinical utility is that the protocols we describe become streamlined, standardized and automated to provide time to result within a few hours in a routine setting. While we currently perform much of the analysis by hand, software under development will simplify interpretation of the results.

Although the DNA microarray method does not currently permit identification as rapidly as qPCR for a single pathogen, it nonetheless offers the advantage of generating a more comprehensive detection profile of aetiological agents present in the specimen and of predicting antibiotic resistance thereby enabling a suitable treatment

to be implemented promptly. Should a serious outbreak occur, a rapid and effective response could be initiated based on data provided by the microarray, thereby facilitating appropriate patient treatment and limiting the propagation of the pathogen.

Experimental procedures

The description of the experimental procedures resides in Appendix S1 in *Supplementary material*.

Acknowledgments

This study was supported by Grant No. UC1 AI062613 (Kennedy) from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. We thank A. Baldwin, S. Brisse, E. Carniel, J.-Y. Coppée, J.-M. Fournier, P. Glaser, P. Grimont, M. Mock, M. Popoff, O. Chesneau, P. Courvalin, P. Despres, C. LeBouguenec, B. Regnault, M.-L. Quilici and their staff for helpful suggestions and for providing nucleic acids from various bacteria and viruses. Vesicular stomatitis virus was a gift from Danielle Blondel, Unité Mixte de Virologie Moléculaire et Structurale, UMR 2472, CNRS, 91198 Gif sur Yvette Cedex, France. We thank C. Drosten (Bernhard Nocht Institute, Hamburg, Germany) for kindly making primer sequences and plasmids available for this study. We thank Guoliang Xing, Gangwu Mei, Brant Wong and Michael Mittman for assistance in array design.

References

Bryant, P.A., Venter, D., Robins-Browne, R., and Curtis, N. (2004) Chips with everything: DNA microarrays in infectious diseases. *Lancet Infect Dis* **4:** 100–111.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., *et al*. (1996) Accessing genetic information with high-density DNA arrays. *Science* **274:** 610–614.

Cutler, D.J., Zwick, M.E., Carrasquillo, M.M., Yohn, C.T., Tobin, K.P., Kashuk, C., *et al*. (2001) High-throughput variation detection and genotyping using microarrays. *Genome Res* **11:** 1913–1925.

Damon, I.K., Roth, C.E., and Chowdhary, V. (2006) Discovery of Monkey pox in Sudan. *N Engl J Med* **355:** 962–963.

Davignon, L., Walter, E.A., Mueller, K.M., Barrozo, C.P., Stenger, D.A., and Lin, B. (2005) Use of resequencing oligonucleotide microarrays for identification of *Streptococcus pyogenes* and associated antibiotic resistance determinants. *J Clin Microbiol* **43:** 5690–5695.

Fodor, S.P., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., and Solas, D. (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science* **251:** 767–773.

Fodor, S.P., Rava, R.P., Huang, X.C., Pease, A.C., Holmes, C.P., and Adams, C.L. (1993) Multiplexed biochemical assays with biological chips. *Nature* **364:** 555–556.

Gingeras, T.R., Ghandour, G., Wang, E., Berno, A., Small, P.M., Drobniewski, F., *et al*. (1998) Simultaneous genotyping and species identification using hybridization pattern recognition analysis of generic *Mycobacterium* DNA arrays. *Genome Res* **8:** 435–448.

Hacia, J.G. (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nat Genet* **21:** 42–47.

Kennedy, G.C., Matsuzaki, H., Dong, S., Liu, W.M., Huang, J., Liu, G., *et al*. (2003) Large-scale genotyping of complex DNA. *Nat Biotechnol* **21:** 1233–1237.

Lin, B., Wang, Z., Vora, G.J., Thornton, J.A., Schnur, J.M., Thach, D.C., *et al*. (2006) Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. *Genome Res* **16:** 527–535.

Palacios, G., Quan, P.L., Jabado, O.J., Conlan, S., Hirschberg, D.L., Liu, Y., *et al*. (2007) Panmicrobial oligonucleotide array for diagnosis of infectious disease. *Emerg Infect Dis* **13:** 73–81.

Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., and Fodor, S.P. (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci USA* **91:** 5022–5026.

Troesch, A., Nguyen, H., Miyada, C.G., Desvarenne, S., Gingeras, T.R., Kaplan, P.M., *et al*. (1999) *Mycobacterium* species identification and rifampin resistance testing with high-density DNA probe arrays. *J Clin Microbiol* **37:** 49–55.

Vora, G.J., Meador, C.E., Stenger, D.A., and Andreadis, J.D. (2004) Nucleic acid amplification strategies for DNA microarray-based pathogen detection. *Appl Environ Microbiol* **70:** 3047–3054.

Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., *et al*. (2003) Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* **1:** E2.

Warrington, J.A., Shah, N.A., Chen, X., Janis, M., Liu, C., Kondapalli, S., *et al*. (2002) New developments in high-throughput resequencing and variation detection using high density microarrays. *Hum Mutat* **19:** 402–409.

Wilson, K.H., Wilson, W.J., Radosevich, J.L., DeSantis, T.Z., Viswanathan, V.S., Kuczmarski, T.A., and Andersen, G.L. (2002) High-density microarray of small-subunit ribosomal DNA probes. *Appl Environ Microbiol* **68:** 2535–2541.

Wong, C.W., Albert, T.J., Vega, V.B., Norton, J.E., Cutler, D.J., Richmond, T.A., *et al*. (2004) Tracking the evolution of the SARS coronavirus using high-throughput, high-density resequencing arrays. *Genome Res* **14:** 398–405.

Zwick, M.E., Mcafee, F., Cutler, D.J., Read, T.D., Ravel, J., Bowman, G.R., *et al*. (2005) Microarray-based resequencing of multiple *Bacillus anthracis* isolates. *Genome Biol* **6:** R10.

Supplementary material

The following supplementary material is available for this article online:

Appendix S1: Methods and basic results.

Appendix S2: Monkeypox clinical sample results.

This material is available as part of the online article from http://www.blackwell-synergy.com

