

Chapter 15

Resequencing Arrays for Diagnostics of Respiratory Pathogens

Baochuan Lin and Anthony P. Malanoski

Abstract

Microarray technology has revolutionized the detection and analysis of microbial pathogens. The success of this technology is evident from the various microarrays that have been developed for this purpose, variation in the density of probes, and the time ranges required for assay completion. Among these, high-density re-sequencing microarrays have demonstrated great potential for detecting bacterial, viral pathogens, and virulence markers. Resequencing microarrays use closely overlapping probe sets to determine a target organism's nucleotide sequence. Hybridization to a series of perfect matched probes provides confirmatory presence/absence information, while hybridization to mismatched probes reveals strain-specific single nucleotide polymorphism (SNP) data. This approach provides sequence information of the diagnostic regions of detected organisms that is considerably more informative over that provided from other microarray techniques.

Key words: Resequencing microarray, pathogen detection, probe sets, respiratory pathogens, single nucleotide polymorphism, febrile respiratory illness.

1. Introduction

High-density resequencing microarrays were developed to detect single nucleotide polymorphisms (SNP) and so produce detailed genetic sequence reads. A resequencing microarray comprises “probe sets” – high-density arrangements of short highly specific oligonucleotide probes (25 and 29 used currently) where each base in a reference sequence is queried by four probes. One probe is an exact match of the reference sequence and the other three represent the same section of reference sequence with the central base position replaced by one of the possible SNP variants. In practical use, the number of probes is doubled so that for each

base both the forward (sense) and reverse (antisense) directions are contained in a probe set. It is possible to completely “resequence,” resolve every base in the sequence of an unknown sample, the reference sequence itself or any other sequence that differs from the reference by one mismatch or fewer per 25 base pair (bp) (1). At higher mismatch rates, a large number of bases can still be identified, but an increasing number will fail as the differences increase. This array-based format, combined with specific PCR, has proven ideal for SNP genotyping and phylogenetic analysis (2–6). Initial work demonstrated the advantages of using a resequencing array with many short reference sequences to detect multiple bacterial and viral pathogens (7–11). Taking full advantage of the sequential base resolution capability of resequencing microarrays, similarity searches of DNA databases have been incorporated into the analysis allowing for fine detailed discrimination of closely related pathogens and tracking mutations within the targeted pathogen even with only partial base call resolution in a reference sequence (8,10–12).

The effective use of resequencing microarrays for respiratory pathogen detection or any large collection of organisms relies on the integration of several components. The overall design for the resequencing microarray and selection of primers for amplification must occur first. This consists of several tasks: First, selection of organisms and desired level of discrimination for each organism and whether specific nucleic acid markers must be tested for; second, determination from known sequence data of sequence regions to choose reference sequences from; third, selection of reference sequences and check for possible conflicts; fourth, primer selection. The order of several of these steps can be interchanged and refinements consist of repeating several of these steps after making changes. Once fabricated, an amplification method is required in order to achieve the sensitivity required for diagnosis/surveillance applications, so that any of the target pathogens can be detected directly from collected samples. Finally, because so many potential organism detection events are to be dealt with, a standardized algorithm is applied to determine if pathogens are detected and report the maximum level of detail possible using the resolved base sequence information from the multiple-pathogen resequencing microarrays.

2. Materials

2.1. Controls

1. IQ-Ex (0.2 pg/ μ L) control template, control forward 1.0 kb PCR primers (20 μ M), control reverse primers (20 μ M), and oligonucleotide control reagent (3.2 nM). These are part of

GeneChip[®] Resequencing Assay Kit (Affymetrix Inc., Santa Clara, CA). These reagents can be stored at -20°C , for up to 6 months. Set up 100 μL PCR for 1.0 kb IQ-EX containing 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM MgCl_2 , 200 μM dNTPs, 1 U of Platinum *Taq* DNA polymerase (Invitrogen Life Technologies, Carlsbad, CA), 3 μL each of control forward 1.0 kb PCR primers (20 μM) and control reverse primers (20 μM), and 5 μL IQ-Ex (0.2 $\text{pg}/\mu\text{L}$) control template. The amplification reaction is carried out with initial denaturation at 95°C for 10 min, followed by 30 cycles of: 94°C for 30 s, 68°C for 30 s, 72°C for 60 s, and a final extension at 72°C for 5 min (**Note 1**).

2. Internal controls for reverse transcription (RT) and PCR. Plasmid vector containing DNA fragment from non-related organism, such as *Arabidopsis thaliana* plant genes, will be suitable for this purpose. Two *Arabidopsis thaliana* plant genes, corresponding to NAC1 and triosphosphate isomerase (TIM), were chosen as internal controls for reverse transcription (RT) and PCR.
3. MEGAscript[™] High Yield Transcription Kit (Ambion, Austin, TX). This kit is stored at -20°C , for up to 3 months (**Note 2**).

2.2. Sample Extraction

1. MasterPure[™] DNA and RNA purification kit (Epicentre Technologies, Madison, WI). Proteinase K is stored at -20°C frost-free freezer. The 2X T&C lysis solution and MPC protein precipitation reagents are stored at room temperature (**Note 3**).

2.3. RT-PCR

1. PCR primers are dissolved in TE buffer to make 100 μM stock and stored at -80°C for up to 1 year. 1–10 μM working solutions are prepared by dilution in nuclease-free H_2O and stored at -20°C for up to 6 months.
2. RT master mix: mix 4 μL 5X first strand RT buffer and 2 μL 0.1 M DTT, 1 μL RNaseOUT[™] (40 U/ μL), and 1 μL SuperScript[™] III reverse transcriptase (200 U/ μL ; Invitrogen Life Technologies, Carlsbad, CA). Add 8 μL RT master mix to each RT reaction. The final RT reaction should contain 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl_2 , 500 μM dNTPs, 40 U of RNaseOUT[™], 10 mM DTT, 2 μM primer LN (5'-CGA TAC GAC GGG CGT ACT AGC GNN NNN NNN N-3'), 200 U of Superscript[™] III reverse transcriptase.
3. Platinum Taq (Invitrogen) or GoTaq[®] Flexi DNA polymerase (Promega Corporation, Madison, WI) are stored at -20°C for up to 6 months (**Note 4**).
4. 10 mM dNTPs (Invitrogen) is stored in 100 μL aliquots at -20°C for up to 6 months.
5. 50X dNTPs (ACGU, Sigma-Aldrich, St. Louis, MO) is stored in 100 μL aliquots at -20°C for up to 6 months (**Note 5**).

6. Heat-labile uracil-DNA glycosylase (1 U/ μ L, USB Corporation, Cleveland, OH) is stored at -20°C for up to 6 months (**Note 5**).
7. PCR master mix: mix 5 μ L 10X PCR buffer (200 mM Tris-HCl, pH 8.4, 500 mM KCl), with 2 μ L MgCl_2 (50 mM), 2 μ L 50x dNTPs (ACGU), 1 μ L Uracil-DNA glycosylase, heat-labile (USB; 1 U/ μ L), 2 μ L primer L (100 μ M, 5'-CGA TAC GAC GGG CGT ACT AGC G-3'), 2 μ L of primers mix A (1 μ M, **Table 15.1**) or 2.5 μ L of primer mix B (1 μ M, **Table 15.2**), 2 μ L Platinum *Taq* DNA (5 U/ μ L) polymerase. Add 40 μ L PCR master mix to 10 μ L of RT product.

2.4. Fragmentation and Labeling of Amplified Product

1. QIAquick PCR purification kit (Qiagen, Valencia, CA), store at room temperature (**Note 6**).
2. Fragmentation solution: For each reaction, mix 4.3 μ L 10X GeneChip[®] fragmentation buffer (Affymetrix), 3.2 μ L nuclease-free H_2O , 0.1 μ L fragmentation reagent (3 U/ μ L; Affymetrix) on ice. The solution can be stored at 4°C for up to 24 h before using (**Note 7**). Individual components can be stored at -20°C for up to 6 months.
3. Labeling solution: For each reaction, mix 12 μ L 5X terminal deoxynucleotidyl transferase buffer (Affymetrix), 2 μ L GeneChip DNA labeling reagent (5 mM), and 3.4 μ L terminal deoxynucleotidyl transferase (30 U/ μ L) on ice. The solution can be stored at 4°C for up to 24 h before using (**Note 7**). Individual components can be stored at -20°C for up to 6 months.

2.5. Hybridization, Washing, and Staining

1. Anti-streptavidin antibody (goat), biotinylated (Vector Laboratories, Burlingame, CA) is dissolved at 0.5 mg/mL in water and stored in 50 μ L aliquots at 4°C for up to 6 months.
2. Pre-hybridization buffer: 10 mM Tris-HCl, pH 7.8 and 0.01% Tween 20. Store the solution at room temperature.
3. Hybridization buffer: Mix 138.6 μ L TMAC (5 M), 2.3 μ L each of 1 M Tris-HCl, pH 7.8, 1% Tween 20, 50 mg/ml BSA, 10 mg/mL Herring sperm DNA (Promega), and 1.8 μ L of oligonucleotide control reagent (Affymetrix), 5.4 μ L fragmented and labeled IQ-EX PCR product, and 13 μ L water. The hybridization master mix can be stored at -20°C for up to 8 weeks.
4. 12X MES stock buffer (1.22 M MES, 0.89 M $[\text{Na}^+]$, dissolve 70.4 g of MES hydrate and 193.3 g of MES sodium salt in 800 mL of molecular biology grade water. Mix and adjust the volume to 1,000 mL, the pH of the solution should be between 6.5 and 6.7. Filter the solution through a 0.2 μ m

Table 15.1
List of PCR primers in primer mix A used for multiplex PCR

Primer name	Sequence (5'→3')	Organism/gene	Amplicon size
FluAHA1-F2	CGA TAC GAC GGG CGT ACT AGC GGC CAA CAA CTC AAC CGA CAC	Influenza A	810 bp
FluAHA1-R2	CGA TAC GAC GGG CGT ACT AGC GAC ACT TCG CAT CAC ATT CAT CC	<i>hemagglutinin</i>	
FluAHA3-F6	CGA TAC GAC GGG CGT ACT AGC GAC TTC CCG GAA ATG ACA ACA	Influenza A	873 bp
FluAHA3-R7	CGA TAC GAC GGG CGT ACT AGC GGG TTT GTC ATT GGG AAT GCT	<i>hemagglutinin</i>	
FluAHA5-F2	CGA TAC GAC GGG CGT ACT AGC GGC CAT TCC ACA ACA TAC ACC	Influenza A	736 bp
FluAHA5-R2	CGA TAC GAC GGG CGT ACT AGC GAG CTA CCA TGA TTG CCA GTG	<i>hemagglutinin</i>	
FluANA1-F5	CGA TAC GAC GGG CGT ACT AGC GAC GTT GTT GCT GGA AAG GAC	Influenza A	1,000 bp
FluANA1-R6	CGA TAC GAC GGG CGT ACT AGC GAA ACT TCC GCT GTA CCC TGA	<i>neuraminidase</i>	
FluANA2-F6	CGA TAC GAC GGG CGT ACT AGC GGG AAA TAT GCC CCA AAC TAG C	Influenza A	1,029 bp
FluANA2-R6	CGA TAC GAC GGG CGT ACT AGC GAT GCA GCT TTT GCC TTC AAC	<i>neuraminidase</i>	
FluAMA-F4	CGA TAC GAC GGG CGT ACT AGC GTT CTA ACC GAG GTC GAA ACG	Influenza A	891 bp
FluAMA-R5	CGA TAC GAC GGG CGT ACT AGC GCT CTG GCA CTC CTT CCG TAG	<i>matrix</i>	
FluBHA-F5	CGA TAC GAC GGG CGT ACT AGC GGG GAG GTC AAT GTG ACT GGT	Influenza B	898 bp
FluBHA-R5	CGA TAC GAC GGG CGT ACT AGC GGG GCA ATT TCC TAT GGC TTT	<i>hemagglutinin</i>	
FluBNA-F4	CGA TAC GAC GGG CGT ACT AGC GGT GAA CCG TTC TGC AAC AAA	Influenza B	899 bp
FluBNA-R3	CGA TAC GAC GGG CGT ACT AGC GCC AAT CTT GGA TGC CAT TCT	<i>neuraminidase</i>	

(continued)

Table 15.1 (continued)

Primer name	Sequence (5' → 3')	Organism/gene	Amplicon size
FluBMA-F2	CGA TAC GAC GGG CGT ACT AGC GCA TTG ACA GAA GAT GGA GAA GG	Influenza B	411 bp
FluBMA-R2	CGA TAC GAC GGG CGT ACT AGC GAA GCA CAG AGC GTT CCT AG	<i>matrix</i>	
Ad5 hexon-F2	CGA TAC GAC GGG CGT ACT AGC GCT GTG GAC CGT GAG GAT ACT	Adenovirus 5 hexon	1,768 bp
Ad5 hexon-R2	CGA TAC GAC GGG CGT ACT AGC GTT GGC GGG TAT AGG GTA GAG C		
Ad5 fiber-F2	CGA TAC GAC GGG CGT ACT AGC GTT ATT CAG CAG CAC CTC CTT G	Adenovirus 5	2,046 bp
Ad5 fiber-R2	CGA TAC GAC GGG CGT ACT AGC GGG TGG CAG GTT GAA TAC TAG	<i>fiber</i>	
Ad5 E1A-F3	CGA TAC GAC GGG CGT ACT AGC GGG CTG ATA ATC TTC CAC CTC C	Adenovirus 5	808 bp
Ad5 E1A-R3	CGA TAC GAC GGG CGT ACT AGC GCT CTC ACG GCA ACT GGT TTA A	<i>E1A</i>	
Ad4 hexon-F3	CGA TAC GAC GGG CGT ACT AGC GGA CAG GAC GCT TCG GAG TAC	Adenovirus 4 hexon	1,334 bp
Ad4 hexon-R3	CGA TAC GAC GGG CGT ACT AGC GGG CAA CAT TGG CAT AGA GGA AG		
Ad4 fiber-F2	CGA TAC GAC GGG CGT ACT AGC GGG TGG AGT GAT GGC TTC G	Adenovirus 4	1,245 bp
Ad4 fiber-R2	CGA TAC GAC GGG CGT ACT AGC GAG TGC CAT CTA TGC TAT CTC C	<i>fiber</i>	
Ad4F E1A-F1	CGA TAC GAC GGG CGT ACT AGC GGC CGT GGA GTA AAT GGC TAA	Adenovirus 4	1,506 bp
Ad4F E1A-R1	CGA TAC GAC GGG CGT ACT AGC GAG TCT TCC AAG ACC GTC CAA	<i>E1A</i>	
Ad7 hexon-F2	CGA TAC GAC GGG CGT ACT AGC GAT GTG ACC ACC GAC CGT AG	Adenovirus 7 hexon	2,417 bp
Ad7 hexon-R2	CGA TAC GAC GGG CGT ACT AGC GGT TGC TGG AGA ACG GTA TG		

(continued)

Table 15.1 (continued)

Primer name	Sequence (5' → 3')	Organism/gene	Amplicon size
Ad7 fiber-F1	CGA TAC GAC GGG CGT ACT AGC GTC TAC CCC TAT GAA GAT GAA AGC	Adenovirus 7 <i>fiber</i>	688 bp
Ad7 fiber-R1	CGA TAC GAC GGG CGT ACT AGC GGG ATA GGC AGT TGT GCT GGG CAT		
Ad7 E1A-F2	CGA TAC GAC GGG CGT ACT AGC GTG AGT GCC AGC GAG AAG AG	Adenovirus 7 <i>E1A</i>	786 bp
Ad7 E1A-R2	CGA TAC GAC GGG CGT ACT AGC GCA GGA GGT GAG GTA GTT GAA TC		
A tha TIM-F2	CGA TAC GAC GGG CGT ACT AGC GTC AAA TCC TCG TTG ACA GAC	A. thaitiana <i>TIM</i>	503 bp
A tha TIM-R2	CGA TAC GAC GGG CGT ACT AGC GTG CAC TGT TGC CTC CAT TGA		

Table 15.2
List of PCR primers in primer mix B used for multiplex PCR

Primer name	Sequence (5' → 3')	Organism/gene	Amplicon size
PIV I HN-F2	CGA TAC GAC GGG CGT ACT AGC GAC AGG AAT TGG CTC AGA TAT G	Parainfluenza 1 <i>hemagglutinin-neuraminidase</i>	382 bp
PIV I HN-R2	CGA TAC GAC GGG CGT ACT AGC GAC ATG ATC TCC TGT TGT CGT		
PIV III HN-F2	CGA TAC GAC GGG CGT ACT AGC GTC GAG GTT GCC AGG ATA TAG G	Parainfluenza 3 <i>hemagglutinin-neuraminidase</i>	477 bp
PIV III HN-R2	CGA TAC GAC GGG CGT ACT AGC GGG ACT ATG AGA TGC CTG ATT GC		
PIV III 5'ND-F2	CGA TAC GAC GGG CGT ACT AGC GCA ACT ATT AGC AGT CAC ACT CG	Parainfluenza 1 <i>5' noncoding region</i>	180 bp
PIV III 5'ND-R2	CGA TAC GAC GGG CGT ACT AGC GAA GTT GGC ATT GTG TTC AGT G		
HRhino 5'ND-F2	CGA TAC GAC GGG CGT ACT AGC GTC ATC CAG ACT GTC AAA GG	Rhinovirus 89 <i>5' noncoding region</i>	423 bp
HRhino 5'ND-R2	CGA TAC GAC GGG CGT ACT AGC GAA ACA GGA AAC ACG GAC ACC		
RSV Lpol-F2	CGA TAC GAC GGG CGT ACT AGC GCT CTA TCA TCA CAG ATC TCA GC	RSV*-A <i>L-polymerase</i>	388 bp
RSV Lpol-R2	CGA TAC GAC GGG CGT ACT AGC GCA TGA GTC TGA CTG GTT TGC		

(continued)

Table 15.2 (continued)

Primer name	Sequence (5' → 3')	Organism/gene	Amplicon size
RSVA MNN-F2	CGA TAC GAC GGG CGT ACT AGC GAC AAA GAT GGC TCT TAG CAA AG	RSV* -A <i>major nucleocapsid</i>	196 bp
RSVA MNN-R2	CGA TAC GAC GGG CGT ACT AGC GAC CCA GTG AAT TTA TGA TTA GC		
RSVB MNN-F2	CGA TAC GAC GGG CGT ACT AGC GAA AAC CAA CCC AAC CAA ACC	RSV* -B <i>major nucleocapsid</i>	248 bp
RSVB MNN-R2	CGA TAC GAC GGG CGT ACT AGC GGC ACA TCA TAA TTG GGA GTG TC		
WNV C-F 2	CGA TAC GAC GGG CGT ACT AGC GGC TCT CTT GGC GTT CTT CAG	West Nile virus C and prM	407 bp
WNV C-R2	CGA TAC GAC GGG CGT ACT AGC GTC ATT ACC AGC CGA CAG CAC		
WNV E-F2	CGA TAC GAC GGG CGT ACT AGC GCC GTC AGC GAT CTC TCC AC	West Nile Virus E	107 bp
WNV E-R2	CGA TAC GAC GGG CGT ACT AGC GCC TGT CCA CCA CTC CTT GTC		
WNV NS1-F2	CGA TAC GAC GGG CGT ACT AGC GGC TTG AAA GGG CAG TTC TGG	West Nile Virus NS1	150 bp
WNV NS1-R2	CGA TAC GAC GGG CGT ACT AGC GCA GGT CTC CGA TTG TGA TTG C		

(continued)

Table 15.2 (continued)

Primer name	Sequence (5' → 3')	Organism/gene	Amplicon size
coron229E MG-F2	CGA TAC GAC GGG CGT ACT AGC GCT CTG GTG TGT GGT GCT TAT A	Coronavirus 229E <i>membrane glycoprotein</i>	718 bp
coron229E MG-R2	CGA TAC GAC GGG CGT ACT AGC GCT CGG CAC GGC AAC TGT C		
coronOC43 MG-F2	CGA TAC GAC GGG CGT ACT AGC GAT GTG GAT GAC GTT TAG GTA	Coronavirus OC43 <i>Membrane glycoprotein</i>	676 bp
coronOC43 MG-R2	CGA TAC GAC GGG CGT ACT AGC GGG TTG ATG GCA GTC GGT AA		
S pne lytA-F2	CGA TAC GAC GGG CGT ACT AGC GAA GAA GAG TTC ATG ACG GAC	<i>S. pneumoniae</i> <i>Autolysin</i>	148 bp
S pne lytA-R2	CGA TAC GAC GGG CGT ACT AGC GTG GTT GTT TGG TTG GTT ATT CG		
S pne ply-F2	CGA TAC GAC GGG CGT ACT AGC GCC GAT GAC TTA TAG TAT TGA	<i>S. pneumoniae</i> <i>pneumolysin</i>	129 bp
S pne ply-R2	CGA TAC GAC GGG CGT ACT AGC GAT AAT CTT GAT GCC ACT TAG C		
M pne CytPI-F2	CGA TAC GAC GGG CGT ACT AGC GGT TCT TCA GGC TCA GGT CAA TC	<i>M. pneumoniae</i> <i>Cytadhesin PI protein</i>	390 bp
M pne CytPI-R2	CGA TAC GAC GGG CGT ACT AGC GAC AGC GGT ATG TAC TGG TCA T		

(continued)

Table 15.2 (continued)

Primer name	Sequence (5' → 3')	Organism/gene	Amplicon size
N men ctrA-F2	CGA TAC GAC GGG CGT ACT AGC GTG GGA ATA GTG TGC GTA TGC	<i>N. meningitidis</i> capsular transport protein	195 bp
N men ctrA-R2	CGA TAC GAC GGG CGT ACT AGC GAC ATC ACC GCG ACG CAG CAA		
N men crgA-F2	CGA TAC GAC GGG CGT ACT AGC GGA TTC CGC GAT GCC GAT G	<i>N. meningitidis</i> regulatory protein, <i>crgA</i>	318 bp
N men crgA-R2	CGA TAC GAC GGG CGT ACT AGC GCG CCC ATG TAT TTA GAG AAC CG		
B per P1XP-F2	CGA TAC GAC GGG CGT ACT AGC GCC GGC GTC GTG CGC GAA A	<i>B. pertussis</i> <i>pertussis toxin promoter region</i>	361 bp
B per P1XP-R2	CGA TAC GAC GGG CGT ACT AGC GCA GCC ACG TCA GCC AGC C		
B per ptxS1-F3	CGA TAC GAC GGG CGT ACT AGC GGA GCG AAT ATC TGG CAC ACC	<i>B. pertussis</i> <i>pertussis toxin S1 subunit</i>	337 bp
B per ptxS1-R3	CGA TAC GAC GGG CGT ACT AGC GGG GCC AGG TCT AGA ACG AAT		
C pne VD4-F2	CGA TAC GAC GGG CGT ACT AGC GTG GAG TAC AAT GGT CTC GAG C	<i>C. pneumoniae</i> major outer membrane protein VD4	161 bp
C pne VD4-R2	CGA TAC GAC GGG CGT ACT AGC GTT TGC ATG AAG TCT GAG AAC GA		

(continued)

Table 15.2 (continued)

Primer name	Sequence (5' → 3')	Organism/gene	Amplicon size
C pne rpoB-F2	CGA TAC GAC GGG CGT ACT AGC GAC GGC ATT ACA ACG GCT AG	<i>C. pneumoniae</i> DNA directed RNA polymerase	406 bp
C pne rpoB-R2	CGA TAC GAC GGG CGT ACT AGC GCA TCT TCT GGT AAT CCC TGT TC		
C pne VD2-F2	CGA TAC GAC GGG CGT ACT AGC GAC AGC GTT CAA TCT CGT TGG	<i>C. pneumoniae</i> major outer membrane protein VD2	249 bp
C pne VD2-R2	CGA TAC GAC GGG CGT ACT AGC GAG AGA ATT GCG ATA CGT TAC AG		
S pyo speB-F2	CGA TAC GAC GGG CGT ACT AGC GCC TTA CAA CCT ATT GAC ACC TG	<i>S. pyogenes</i> pyrogenic exotoxin B	371 bp
S pyo speB-R2	CGA TAC GAC GGG CGT ACT AGC GAC ACG AGA GCT ACC TGC AGA		
S pyo mef-F2	CGA TAC GAC GGG CGT ACT AGC GTT TAT ACA ATA TGG GCA GGG	<i>S. pyogenes</i> macrolide-efflux determinant (<i>mefA</i> , <i>mefE</i>)	381 bp
S pyo mef-R2	CGA TAC GAC GGG CGT ACT AGC GTC GTA AGC TGT TCT TCT GGT AC		
S pyo ermB-F2	CGA TAC GAC GGG CGT ACT AGC GTC ATT GCT TGA TGA AAC TGA T	<i>S. pyogenes</i> erythromycin resistance methylase (<i>ermB</i>)	244 bp
S pyo ermB-R2	CGA TAC GAC GGG CGT ACT AGC GTT GGA TAT TCA CCG AAC ACT AG		

(continued)

Table 15.2 (continued)

Primer name	Sequence (5' → 3')	Organism/gene	Amplicon size
S pyo ermTR-F2	CGA TAC GAC GGG CGT ACT AGC GCT TGT GGA AAT GAG TCA ACG G	<i>S. pyogenes</i> <i>erm(TR)</i>	233 bp
S pyo ermTR-R2	CGA TAC GAC GGG CGT ACT AGC GAG GTA GCT ATA TTT CGC TTG AC		
B ant rpoB-F2	CGA TAC GAC GGG CGT ACT AGC GGA GCG TCT ACG TCC TGG TGA	<i>B. anthracis</i> <i>RNA polymerase beta-subunit</i>	291 bp
B ant rpoB-R2	CGA TAC GAC GGG CGT ACT AGC GCA TTG GTT TCG CTG TTT TGA		
B ant pag-F2	CGA TAC GAC GGG CGT ACT AGC GTG GAA GAG TGA GGG TGG ATA C	<i>B. anthracis</i> <i>protective antigen</i>	486 bp
B ant pag-R2	CGA TAC GAC GGG CGT ACT AGC GAA TAA TCC CTC TGT TGA CGA A		
B ant capB-F2	CGA TAC GAC GGG CGT ACT AGC GAG GAG CAA TGA GAA TTA CAC G	<i>B. anthracis</i> <i>poly(D-glutamic acid) capsule</i>	311 bp
B ant capB-R2	CGA TAC GAC GGG CGT ACT AGC GCT AAG TTC CAA TAC TCT TGC		
VMVHA-F3	CGA TAC GAC GGG CGT ACT AGC GGC CCG TAC TTA TGT ATG TGC ATT	Variola Major Virus <i>hemagglutinin</i>	439 bp
VMVHA-R3	CGA TAC GAC GGG CGT ACT AGC GCA TCA TTG GCG GTT GAT TTA		

(continued)

Table 15.2 (continued)

Primer name	Sequence (5' → 3')	Organism/gene	Amplicon size
VMVcrmB-F3	CGA TAC GAC GGG CGT ACT AGC GGG GAA CAT ACG CTT CCA GAT	Variola Major Virus <i>cytokine response modifier B</i>	257 bp
VMVcrmB-R3	CGA TAC GAC GGG CGT ACT AGC GTT CCA CAT TTT GTT TGG GAA A		
EVLGV-F3	CGA TAC GAC GGG CGT ACT AGC GCC TTA TCC GAC TCG CAA TGT	Ebola Virus <i>L₁ gene</i>	366 bp
EVLGV-R3	CGA TAC GAC GGG CGT ACT AGC GCA GTG TGA GGT TAT GTG GTG GA		
LVGPC-F4	CGA TAC GAC GGG CGT ACT AGC GTT GGT TGC GCA ATT CAA GT	Lassa Virus <i>GPC₁ gene</i>	302 bp
LVGPC-R4	CGA TAC GAC GGG CGT ACT AGC GTG TTG TTC TTT GTG CAG GAG A		
F tul 13Kd-F2	CGA TAC GAC GGG CGT ACT AGC GTC GTA ATG TTA GCT GTA TCA TC	F. tularensis <i>13-kDa lipoprotein</i>	513 bp
F tul 13Kd-R2	CGA TAC GAC GGG CGT ACT AGC GTA CAT TAG CTG TCC ACT TAC CG		
F tul FopA-F2	CGA TAC GAC GGG CGT ACT AGC GGT GGG TGG TGG TCT TAA GTT T	F. tularensis <i>FopA</i>	192 bp
F tul FopA-R2	CGA TAC GAC GGG CGT ACT AGC GCT GGA TAT TAC CAG TGT CAT T		

(continued)

Table 15.2 (continued)

Primer name	Sequence (5' → 3')	Organism/gene	Amplicon size
Y pes cve-F2	CGA TAC GAC GGG CGT ACT AGC GAC TGA TAA AGG GGA GTG GAT A	<i>Y. pestis</i> <i>cve2155</i> sequence	388 bp
Y pes cve-R2	CGA TAC GAC GGG CGT ACT AGC GCT CGC CTT GCT CTT TGA GC		
Y pes cafl-F2	CGA TAC GAC GGG CGT ACT AGC GGG ACA CAA GCC CTC TCT ACG	<i>Y. pestis</i> <i>Cafl</i>	564 bp
Y pes cafl-R2	CGA TAC GAC GGG CGT ACT AGC GTA GAT ACG GTT ACG GTT ACA G		
AthaNAC1-F5	CGA TAC GAC GGG CGT ACT AGC GCA TGG GAA GCT GTT TTG ATG	<i>A. thaliana</i> NAC1	497 bp
AthaNAC1-R5	CGA TAC GAC GGG CGT ACT AGC GCC CGA AGA ATT GTT CCA ATC		

*RSV: Respiratory syncytial virus

filter, and store at 2–8°C shielded from light. The solution should be clear, discard solution if it turns yellow.

5. Wash A (non-stringent wash buffer): 6X SSPE, 0.01% Tween-20. Filter the solution through a 0.2 µm filter and store at room temperature for up to 6 months.
6. Wash B (stringent wash buffer): 0.6X SSPE, 0.01% Tween-20. Filter the solution through a 0.2 µm filter, and store at room temperature for up to 6 months.
7. 1X array holding buffer (100 mM MES, 1 M [Na⁺], 0.01% Tween-20). Mix 8.3 mL of 12X MES stock buffer with 18.5 mL of 5 M NaCl, 0.1 mL of 10% Tween 20, and 73.1 mL of water. Store the buffer at 2–8°C and shield from light.
8. SAPE stain solution: 6X SSPE, 0.01% Tween-20, 1X Denhardt's solution (Sigma-Aldrich), 10 µg/mL of Streptavidin Phycoerythrin (SAPE, Molecular Probes). The solution can be stored at 4°C shielded from light for up to 1 week.
9. Antibody stain solution: 6X SSPE, 0.01% Tween-20, 1X Denhardt's solution (Sigma-Aldrich), 5 µg/mL of biotinylated anti-streptavidin antibody (goat, Vector Labs). The solution can be stored at 4°C for up to 1 week.

3. Methods

The success of using resequencing microarray for multi-organism detection, as in broad spectrum detection of various respiratory pathogens, relies on resolving two issues before the assay can be applied to samples. First the chip must be designed by selecting appropriate reference sequences to answer the questions that will be asked. The second consideration is multiplex primer selection since the assay outlined, rapid analysis of samples with large amounts of background nucleic acid material, requires the use of specific or semi-specific primers.

3.1. Chip Design

Selection of partial genomic sequences from pathogens (reference or target sequences) for placement on a resequencing microarray to provide direct sequence-based identification of multiple pathogens depends on what specific knowledge is required for the various pathogens. For example, the respiratory pathogen microarray v.1 (RPM v.1) chip design includes 57 target genes, partial sequences from the genes containing diagnostic regions of each pathogen (i.e., *E1A*, *hexon*, and *fiber* for human adenoviruses (HAdVs); *hemagglutinin*, *neuraminidase* and the *matrix* genes for influenza A viruses). The targets for both HAdVs and

influenza are both long enough that RPM v.1 not only allows identification but also produce strain-specific sequence data at the same time. The remaining respiratory pathogens only required detection, so fewer and shorter partial sequences were selected allowing resequencing of 29.7 kb of sequences to provide at least species level identification of 26 distinct organisms (10) (**Note 8**). For the RPM v.1 chip, selection of partial sequences to generate probe sets on the microarray was based on the same rules used in selecting probes for long-oligonucleotide spotted microarrays even though such rules do not account for the strengths and weaknesses of a resequencing microarray. Overall, the detection and discrimination performance of such sequences was good. In fact, probes that for a spotted array would only discern to a particular level such as serotype will at least give the same level of discrimination on a resequencing array and often provide more detailed discrimination such as strain differentiation. Using these selection rules can however lead to wasted space on the resequencing microarray because in some cases where two probes were required on a spotted microarray, the information from only one is sufficient to provide equivalent or greater detection and discrimination on a resequencing microarray. Design methods have since been refined to reduce redundancy and better incorporate the advantages of resequencing arrays into probe selection. Once selected, the sequence file was sent to Affymetrix for fabrication (**Note 9**).

3.2. Multiplex Primer Design

The gene-specific primer pairs for all targets on the RPM v.1 chip (8) were designed according to the following criteria to meet minimum amplification efficiency requirement.

1. From our work we have established a gross predictor that hybridization will occur for an organism on the array and that at least 70% of the bases match between the sequence used on the microarray and the organism when aligned (BLAST) (12). A list of sequences that may potentially hybridize to the reference sequence is constructed using a BLAST query. Primers are selected from consensus sequences of well-conserved regions flanking the reference sequence from the list. All potential primers that are 18–24 bases in length with ~50% GC content, with no repetitive sequences and have annealing temperature range from 55 to 60°C without potential for self annealing and hairpin formation are considered. This list is further filtered to ensure uniqueness with respect to the other pathogens and human genome by using a full search of the GenBank database with the BLAST program. This insured that the potential primers for an organism have a number of mismatches with these two groups of sequences and would not mis-prime on a sequence region not of interest in the assay (**Note 10**).

2. Once selected, all primers in the same primer cocktails are checked for potential hybridization to other primers to reduce the potential of primer-dimer formation. The primers that form conceivable primer dimers, 8 or more contiguous base matches between the primers, are replaced with new ones until all potential primer dimers are removed. Also, we adapt a method developed by Shuber et al. and Brownie et al. (13,14) to further suppress primer-dimer formation by adding a linker sequence of 22 bp (primer L) to the 5'-end of primers used (**Note 11**).
3. To minimize the possibility of intra-primer interactions, the number of primers in a mix is kept to no more than 100. For RPM v.1, the primers were divided into two independent reactions to satisfy this requirement. Fine-tuning adjustments to both mixtures (swapping primers that amplified poorly for new ones) were carried out to ensure all target genes from the 26 targeted pathogens (West Nile Virus is included on the array but not in this amplification scheme) would amplify sufficiently to generate detectable hybridization. Primer sequences are listed in **Tables 15.1** and **15.2 (8)**.

3.3. Sample Preparation

1. Mix 150 μL of the fluid samples (nasal washes or throat swabs in storage media) with 150 μL of 2X T&C lysis solution premixed with 1 μl of 50 $\mu\text{g}/\mu\text{L}$ proteinase K thoroughly by vortexing. The sample mixture is incubated at 65°C for 15 min with vortex mixing every 5 min. After incubation, place the sample on ice for 3–5 min.
2. Add 150 μL of MPC protein precipitation reagent to the sample mixture and vortex vigorously for 10 s. At this point, sample mixture should appear cloudy. Pellet the debris by centrifugation for 10 min at 13,000 rpm at room temperature using microcentrifuge. If the pellet is clear, small or loose, add an additional 25 μL of MPC protein precipitation reagent, mix, and spin again.
3. Transfer the supernatant to a clean 1.5 mL tubes and discard the pellet (**Note 12**). Add 500 μL of isopropanol, then invert tube several times to mix thoroughly. Pellet the DNA by centrifugation at 4°C for 10 min at 13,000 rpm using microcentrifuge. Pour off the isopropanol, be careful not to lose the DNA pellet. Rinse twice with 75% ethanol, centrifuge briefly if the pellet is dislodged. Remove all the residual ethanol with a pipette and air dry the pellet for 5–10 min. Resuspend the total nucleic acids in 25 μl of nuclease-free water and store at –20°C until further use (**Note 13**).

3.4. Internal Controls

Two *Arabidopsis thaliana* plant genes, corresponding to NAC1 and TIM, were chosen as internal controls for reverse transcription

(RT) and PCR as they are unlikely to occur naturally in clinical samples (**Note 14**). Two plasmids, pSP64poly(A)-NAC1 and pSP64poly(A)-TIM, containing ~500 bp of the two genes were kindly provided by Dr. Norman H. Lee at The Institute for Genome Research (Rockville, MD) (15).

1. NAC1 fragment is amplified by PCR with SP6 (5'-GAT TTA GGT GAC ACT ATA-3' and M13R (5'-CAG GAA ACA GCT ATG AC-3') primers, and the PCR products are purified using QIAquick PCR Purification Kit (Qiagen).
2. To generate RNA from pSP64poly(A)-TIM, the plasmids were linearized with *EcoRI* and in vitro transcribed from the SP6 promoter using the MEGAscript™ High Yield Transcription Kit (Ambion). 60 fg each of NAC1 and TIM are used as internal controls for checking the amplification efficiency and the presence of inhibitors in the specimens.

3.5. Multiplex RT-PCR Amplification

1. Reverse transcription (RT) reactions are performed in 20 µL volumes. Mix 5–8 µL of extracted DNA with 1 µL 10 mM dNTPs (**Note 15**), 2 µL 40 µM primer LN, 60 fg/µL each of NAC1 and TIM, and bring the volume up to 12 µL. The reaction is incubated at 65°C for 5 min, then chilled on ice for at least 1 min. Add 8 µL RT master mix and incubate at 25°C for 10 min, 50°C for 50 min, then denature the enzyme at 85°C for 5 min.
2. The RT reaction products are split up into two 10 µL volumes and used in two different multiplex PCR. Primer mix A contains 19 primer pairs and amplifies 18 gene targets from three different influenza A viruses, 1 influenza B virus, 3 serotypes of HAdVs, and one internal control (TIM; **Table 15.1**). Primer mix B contains 38 primer pairs and amplifies the remaining 37 gene targets and the other internal control (NAC1; **Table 15.2**) (8). The amplification reaction is carried out with an initial incubation at 25°C for 10 min., preliminary denaturation at 94°C for 3 min., followed by 5 cycles of: 94°C for 30 s, 50°C for 90 s, 72°C for 120 s, then 35 cycles of: 94°C for 30 s, 64°C for 120 s, and a final extension at 72°C for 5 min (**Note 16**). The amplified products from both PCR are combined into a single volume and subjected to purification and processing prior to hybridizing to the RPM v.1 chips.

3.6. Microarray Hybridization and Processing

1. Combine PCR product into one tube, then add 5 volumes of PB buffer to one volume of the PCR samples (500 µL PB for 100 µL PCR product). Mix the reaction mixture by pipetting up and down. The color of the mixture should not change at this point. If the color turns orange or purple, add 10 µL of 3 M sodium acetate, pH 5.0 and mix again.

2. Pipette 700 μL of the mixture into QIAquick spin column in a 2 mL collection tube and centrifuge for 30–60 s at 13,000 rpm using microcentrifuge at room temperature. Discard flow-through and repeat the process if the mixture volume is larger than 700 μL .
3. Wash the spin column with 750 μL PE buffer with the indicated amount of ethanol added, and centrifuge for 30–60 s at 13,000 rpm using microcentrifuge at room temperature. Discard the flow-through, then place column back into the collection tubes. Centrifuge for 60 s at 13,000 rpm using microcentrifuge at room temperature to remove residual ethanol.
4. Elute the DNA by placing the spin column in a new 1.5 mL tube, and add 50 μL EB buffer (10 mM Tris–HCl, pH 8.5) to the center of the spin column and centrifuge for 60 s at 13,000 rpm using microcentrifuge at room temperature (**Note 17**).
5. Purify the IQ-EX PCR products as described in step 1. Determine the concentration of the IQ-EX using UV spectrophotometry.
6. Add 7.6 μL of the fragmentation solution to 35 μL of eluted DNA and 3 μg of IQ-EX PCR product in a final volume of 35 μL . Incubate the reaction mixture at 37°C for 5 min (**Note 18**), then denature the enzyme activity at 95°C for 15 min. Store at 4°C after incubation. At this point, you can store the sample at 4°C for up to 1 week before labeling.
7. Add 17.4 μL of the labeling solution to 35 μL of fragmented DNA and IQ-EX PCR product from step 3. Incubate the reaction mixture at 37°C for 30 min (**Note 18**), and then denature the enzyme activity at 95°C for 15 min. Store at 4°C. Use fragmented and labeled IQ-EX PCR product to prepare the hybridization buffer.
8. Add 160 μL of hybridization buffer to 60 μL of fragmented and labeled PCR products. At this point, you can store the sample at –20°C for up to 1 month before hybridization.
9. Add 200 μL pre-hybridization buffer to each chip, and incubate the chip in the hybridization oven at 49°C at 60 rpm for 15 min. In the meantime, denature the samples from step 5 at 95°C for 5 min, and then equilibrate the tubes at 49°C for 5 min (**Note 19**).
10. Remove the arrays from the hybridization oven, and remove and discard the pre-hybridization buffer. Add 200 μL of denatured samples. At this point, you should see a small bubble inside the chip which serves as a mixing mechanism for the microarray, so ensure such a bubble is present. Incubate the chip at the hybridization oven at 49°C at 60 rpm for

4–16 h (**Note 20**). Remove the hybridization mixture from the array, and fill with 250 μ L of array holding buffer. At this point, you can store the array at 4°C for up to 3 h before washing and staining (**Note 21**).

11. Prime the GeneChip[®] fluidic stations (Affymetrix) with Wash A and B. Register a new experiment in GeneChip Operating Software Service (GCOS). Load the array, SAPE and antibody stain solution, into the designed fluidic module. Start the washing and staining protocol “DNAArray_WS5_450.” Remove the array when the protocol is complete; make sure at this time that there is not a bubble. If there are visible bubbles, manually fill the array with array holding buffer using a pipette. Apply two tough spots to each of the two septa on the back of the array. The array can be stored at 4°C for up to 24 h before scanning (**Note 22**). Flush the fluidic stations with DI water and shut down.
12. Turn on the GeneChip Scanner 3000 at least 10 min before use. If the array was stored at 4°C, allow to warm to room temperature before scanning. Insert the array into scanner. Use GCOS to start scanning the array by selecting the corresponding experiment. GCOS will process the image file of the scanned microarray and create cell intensity data. An example of the results is shown in **Fig. 15.1**.

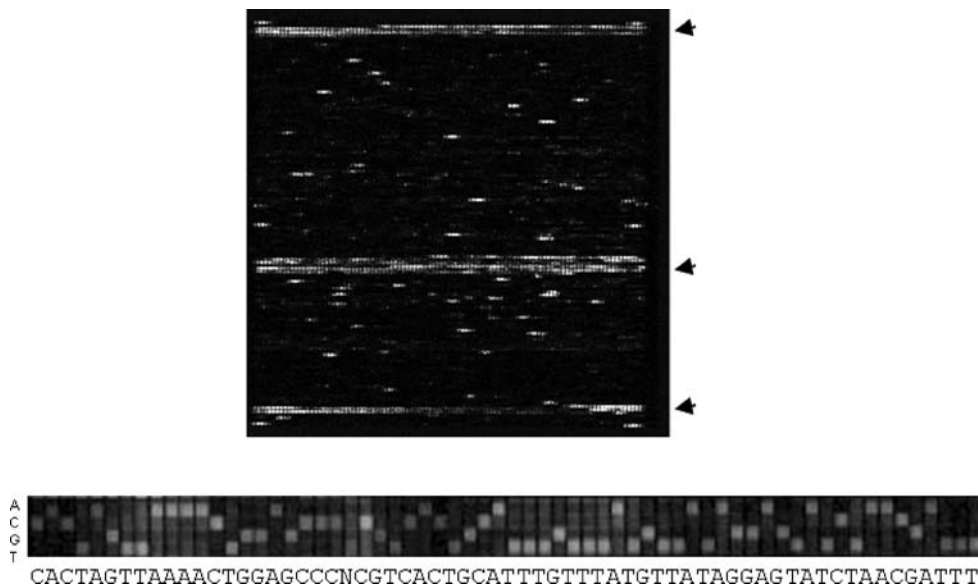


Fig. 15.1. *Top panel:* The hybridization image of the HAdV-7 strain NRRC 1315. *Lower panel:* magnification of a portion of Ad7EIA probe sets showing an example of the primary sequence data generated by the hybridization of amplified targets to RPM V.1. The primary sequence generated can be read from left to right. The *top arrow* indicates Affymetrix hybridization control-IQ-EX; the *middle arrow* indicates the regions that HAdV-7 hybridize; the *lower arrow* indicates MAC1 and TIM, internal amplification control.

13. Use GeneChip[®] Sequence Analysis Software (GSEQ) to analysis the cell intensity data and generate the base call (**Note 23**). Export the sequence information to FASTA file.

3.7. Pathogen Identification Algorithm

1. Sending entire FASTA files to be searched by BLAST is wasteful of time and potentially misleading (**Note 24**). The set of resolved bases resulting from hybridization is instead subjected to a filtering process. Each reference sequence is examined by itself and split into possible subsequences (SubSeqs) suitable for BLAST search. SubSeqs are found by finding seed locations within the sequence that have at least 18 of 20 bases resolved. One of these locations is increased in size while the total called base percentage stays above 40% unless a contiguous stretch of at least 18 or 19 bases of N calls is encountered. The section is marked as its own SubSeq if its length is at least 30 bases and the remaining seed locations are examined in a similar manner.
2. BLAST is used to perform a similarity search of the NCBI nr database using the SubSeqs as the queries. The BLAST program used is the NCBI Blastall `-p blastn` with a defined set of parameters. Masking of low complexity regions is performed for the seeding phase; however, such regions are included in the actual scoring. The default gap penalty and nucleotide match score are used. The nucleotide mismatch penalty, `-q`, parameter is set to `-1` rather than the default. The results of any BLAST query with an expected value <0.0001 are returned in tabular format from the blastall program. If any SubSeq has a value of 10^{-6} or less then it is considered positive for identification of whatever organism is reported in the return.
3. A SubSeq might return many records from the database with the same score. The identified organism is whatever taxonomic classification encompasses all tied best scoring returns when there are more than one (**Note 25**). If only a single return has the best score then that is considered the closest specific strain to the organism in the sample.
4. Different SubSeq for the same reference sequence are required to result in a single pathogen identification for that reference sequence. If one SubSeq has a significantly better score and more detailed identification than others that is taken as the identification of the reference sequence. If all SubSeqs have similar scores, then the taxonomic classification that is consistent for all of them is considered the best identification that can be made.
5. A final examination can be made for the results from reference sequences that targeted the same organism to insure that they are reporting consistent results. It is not required that all reference sequences identify an organism nor is strictly

required that they make the same exact identification. This process can be very laborious and time consuming considering the large number of reference sequences and automation of this process is possible (**Note 26**).

4. Notes



1. For amplification of IQ-EX, other Taq DNA polymerase can also be used. GeneChip[®] Resequencing Assay Kit also contains control forward 7.5 kb PCR primers (20 μ M) which can replace control forward 1.0 kb PCR primers. However, the 7.5 kb IQ-EX amplification requires using Taq DNA polymerase designed for long-range PCR amplification.
2. Other transcription kits can also be used to generate RNA controls.
3. The MasterPure[™] DNA and RNA purification kit gave us the highest yield of nucleic acids, but other nucleic acid extraction kits can also be used for this purpose.
4. For multiplex PCR, it is highly recommended that you test different Taq DNA polymerase. We tested several different Taq DNA polymerases from various companies; our experience suggests that Platinum Taq or GoTaq[®] Flexi DNA polymerase give the highest amplification yield.
5. Good laboratory practice, e.g., always use filtered pipette tips, clear gloves, and UV irradiate PCR hood after each PCR set up, is necessary to prevent contamination issue in the PCR. Additionally, 50x dNTPs (ACGU) and uracil-DNA glycosylase are used to prevent trace amount of carry-over contamination from the previous PCR.
6. Other PCR purification kit can also be used for cleaning up PCR products.
7. The fragmentation reagent is an equivalent of DNase I. Terminal deoxynucleotidyl transferase and biotinylated ddNTP from other sources can also be used for labeling reaction.
8. The resequencing microarray platform is generally more flexible at target selection. However, the design process can still be very complicated. Fortunately, the RPM v.1 chips are available through Tessarae Inc. (Potomac Falls, VA) with the permission from US Naval Research Laboratory. Newer version of RPM (RPM v.3.x) can be also obtained through Tessarae, Inc. (Potomac Falls, VA) with complete protocol. So initial use of the assay does not require design of a microarray.

9. The details of how to make Affymetrix CustomSeq resequencing microarray is described in GeneChip[®] CustomSeq Custom Resequencing Array Design Guide (http://www.affymetrix.com/support/technical/other/customseq_design_manual.pdf). Briefly, the first step is to identify and generate the base sequences of interest in FASTA format. Once the content of the array has been selected, the total number of bases is calculated and an array format (different formats have different sequence capacities) is chosen. The sequence submission documents can then be prepared from this information (design request form, sequence and instruction files and purchase order). During the fabrication process at Affymetrix, their chip design group will perform a design clarification process which will check and suggest removing ambiguous, repetitive and homologous sequences. Upon completion of this step, the masks required to produce chips are fabricated and checked for quality. Finally, the arrays will be manufactured. The clarification process can take as little as 2 weeks if little feedback is required but can run longer and the production of chips depends upon scheduling constraints to meet delivery to all their customers.
10. Primers selected should have at least three base mismatches with human genome sequences to avoid non-specific amplification.
11. Short linker primers can also be used, but the linker primers must be unique with higher melting temperature and unrelated to the target pathogens and background genome sequences that the samples may contain.
12. Try to avoid lipid and small white powdery protein substances in the supernatant.
13. If you air dry the pellet too long, you will need to add nuclease-free water to the pellet and store at 4°C overnight to ensure the complete suspension of the nucleic acids.
14. Internal control is not absolutely necessary for the reaction, but it is good to ensure there is no false negative result due to the RT or PCR step. Other genes besides NAC1 and TIM can also be used for internal controls provided the custom design chip has reference sequences to hybridize with the control genes.
15. Do not use 50x dNTPs (ACGU, Sigma) at this step.
16. Two-step PCR are used here to shorten the cycling time, alternatively, a three-step PCR can be used as long as the annealing temperature is raised enough to switch to linker only priming.
17. Smaller amount of elution buffer can be used; EB buffer can be used to bring up the volume for the next step.

18. Longer incubation time is recommended for both fragmentation (30 min) and labeling (2 h) by Affymetrix. However, our experience suggested that a shorter incubation time is sufficient.
19. It is recommended that the chip should be warm up to room temperature before performing pre-hybridization. Our experiences suggested that this is not a critical step.
20. It is recommended that the hybridization should be carried out for 16 h in order to reach equilibrium. Our experience indicates that 4 h hybridization can generate sufficient base call for pathogen identification.
21. Wash A can also be used to fill the chip at this stage. Chips can be stored for up to 24 h before washing.
22. Chips can still be scanned after storing for more than 24 h and less than 1 week, although the fluorescence signal will be weaker.
23. The base calling algorithm for resequencing microarray is changed to allow for the most “permissive” base calling to be executed compared to “conservative” settings that are defaults for GDAS. The parameters are listed below:

“Permissive” Base Calling Algorithm Settings –

 - a. Filter Conditions
 - i. No Signal threshold = 0.500 (default = 1.000000)
 - ii. Weak Signal Fold threshold = 20000.000 (default = 20.000000)
 - iii. Large SNR threshold = 20.000000 (default = 20.000000)
 - b. Algorithm Parameters
 - i. Strand Quality Threshold = 0.000 (default = 0.000000)
 - ii. Total Quality Threshold = 25.0000 (default = 75.000000)
 - iii. Maximum Fraction of Heterozygote Calls = 0.99000 (default = 0.900000)
 - iv. Model Type (0 = Heterozygote, 1 = Homozygote) = 1
 - v. Perfect Call Quality Threshold = 0.500 (default = 2.000000)
 - c. Final Reliability Rules
 - i. Min Fraction of Calls in Neighboring Probes = 1.0000 (disables filter)
 - ii. Min Fraction of Calls of Samples = 1.0000 (disables filter)
24. Resequencing arrays provide positional information which allows for the use of similarity searches; however, because of how the information is obtained they can potentially bias for or against variants with insertions or deletions depending on

the reference sequence selected. Splitting regions that are separated by large sections of N calls reduces this bias.

25. The organism identified may not necessarily be the primary organism the reference sequence is intended to identify but a near-neighbor species.
26. The algorithm as described was built into a new software program, Computer-Implemented Biological Sequence-based Identifier system, version 2 (CIBSI 2.0) to automate the pathogen identification process for the RPM v.1 array (12). CIBSI 2.0 besides determining what each reference sequence detects furthermore, whether the identifications from separate targets support a common organism identification and determine whether detected organisms belong to the target set that the assay was designed to detect or are related to close genetic near neighbors. Target pathogens are the organisms the assay was specifically designed to detect.

References

1. Hacia, J.G. (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nat Genet*, **21**, 42–47.
2. Gingeras, T.R., Ghandour, G., Wang, E., Berno, A., Small, P.M., Drobniewski, F., Alland, D., Desmond, E., Holodniy, M. and Drenkow, J. (1998) Simultaneous genotyping and species identification using hybridization pattern recognition analysis of generic *Mycobacterium* DNA arrays. *Genome Res*, **8**, 435–448.
3. Kozal, M.J., Shah, N., Shen, N., Yang, R., Fucini, R., Merigan, T.C., Richman, D.D., Morris, D., Hubbell, E., Chee, M. et al. (1996) Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. *Nat Med*, **2**, 753–759.
4. Wong, C.W., Albert, T.J., Vega, V.B., Norton, J.E., Cutler, D.J., Richmond, T.A., Stanton, L.W., Liu, E.T. and Miller, L.D. (2004) Tracking the evolution of the SARS coronavirus using high-throughput, high-density resequencing arrays. *Genome Res*, **14**, 398–405.
5. Wilson, K.H., Wilson, W.J., Radosevich, J.L., DeSantis, T.Z., Viswanathan, V.S., Kuczmariski, T.A. and Andersen, G.L. (2002) High-density microarray of small-subunit ribosomal DNA probes. *Appl Environ Microbiol*, **68**, 2535–2541.
6. Wilson, W.J., Strout, C.L., DeSantis, T.Z., Stilwell, J.L., Carrano, A.V. and Andersen, G.L. (2002) Sequence-specific identification of 18 pathogenic microorganisms using microarray technology. *Mol Cell Probes*, **16**, 119–127.
7. 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.
8. Lin, B., Blaney, K.M., Malanoski, A.P., Ligler, A.G., Schnur, J.M., Metzgar, D., Russell, K.L. and Stenger, D.A. (2007) Using a resequencing microarray as a multiple respiratory pathogen detection assay. *J Clin Microbiol*, **45**, 443–452.
9. Lin, B., Malanoski, A.P., Wang, Z., Blaney, K.M., Ligler, A.G., Rowley, R.K., Hanson, E.H., von Rosenvinge, E., Ligler, F.S., Kusterbeck, A.W. et al. (2007) Application of broad-spectrum, sequence-based pathogen identification in an urban population. *PLoS ONE*, **2**, e419.
10. Lin, B., Wang, Z., Vora, G.J., Thornton, J.A., Schnur, J.M., Thach, D.C., Blaney, K.M., Ligler, A.G., Malanoski, A.P., Santiago, J. et al. (2006) Broad-spectrum

- respiratory tract pathogen identification using resequencing DNA microarrays. *Genome Res*, **16**, 527–535.
11. Wang, Z., Daum, L.T., Vora, G.J., Metzgar, D., Walter, E.A., Canas, L.C., Malanoski, A.P., Lin, B. and Stenger, D.A. (2006) Identifying influenza viruses with resequencing microarrays. *Emerg Infect Dis*, **12**, 638–646.
 12. Malanoski, A.P., Lin, B., Wang, Z., Schnur, J.M. and Stenger, D.A. (2006) Automated identification of multiple micro-organisms from resequencing DNA microarrays. *Nucleic Acids Res*, **34**, 5300–5311.
 13. Brownie, J., Shawcross, S., Theaker, J., Whitcombe, D., Ferrie, R., Newton, C. and Little, S. (1997) The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Res*, **25**, 3235–3241.
 14. Shuber, A.P., Grondin, V.J. and Klinger, K.W. (1995) A simplified procedure for developing multiplex PCRs. *Genome Res*, **5**, 488–493.
 15. Wang, H.Y., Malek, R.L., Kwitek, A.E., Greene, A.S., Luu, T.V., Behbahani, B., Frank, B., Quackenbush, J. and Lee, N.H. (2003) Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays. *Genome Biol*, **4**, R5.