



Defence Research and  
Development Canada

Recherche et développement  
pour la défense Canada



# **An Affymetrix Microarray Design for Microbial Genotyping**

B.N. Ford, D. Bader, and Y. Shei  
DRDC Suffield

C. Ruttan and D. Mah  
Canada West Biosciences

**Defence R&D Canada**

Technical Memorandum

DRDC Suffield TM 2009-183

October 2009

**Canada**



# **An Affymetrix Microarray Design for Microbial Genotyping**

B. N. Ford, D. Bader, Y. Shei, Defence R&D Canada – Suffield  
C. Ruttan, D. Mah, Canada West Biosciences

**Defence R&D Canada – Suffield**

Technical Memorandum

DRDC Suffield TM 2009-183

October 2009

Principal Author

*Original signed by Barry N. Ford*

---

Barry N. Ford

Defence Scientist

Approved by

*Original signed by S.G. Bjarnason*

---

S.G. Bjarnason

Section Head, Casualty Management Section

Approved for release by

*Original signed by Paul D'Agostino*

---

Paul D'Agostino

Chair, Document Review Panel

- © Her Majesty the Queen in Right of Canada, as represented by the Minister of National Defence, 2009
- © Sa Majesté la Reine (en droit du Canada), telle que représentée par le ministre de la Défense nationale, 2009

## Abstract

---

There is an ongoing requirement for development of high-density or multiplex assays for detection or identification of microbes. There is also a need to develop assays or toolsets that can detect or identify microbial threats without prior knowledge of the target microbe(s) in a given sample. Indeed, some samples that contain no culturable material (e.g. viable but non-culturable cells) will nonetheless contain detectable DNA fragments which might be of value with respect to forensics or attribution of source. For many pathogenic microbes, various specific tests already exist, but there are few general methods wherein a single adaptable tool can be applied to multiple species or to previously uncharacterized organisms. The high-density DNA microarray has the potential to address many of these requirements and thus complements existing identification tools. The microarray platform has for example, the ability to detect microbial DNA that is not a perfect match to known genomic DNA sequences, thus making it possible to detect microbial variants that might otherwise be missed. In this report, the design and preliminary testing of a high-density DNA microarray for the purpose of microbial identification and detection is described.

## Résumé

---

Les méthodes à haute densité et les techniques multiplex pour la détection et l'identification des microorganismes sont des outils toujours en demande. Il faut aussi des méthodes pour détecter et identifier les dangers microbiens savoir quels microorganismes peuvent être présents dans les échantillons. En fait, certains échantillons ne contenant que des espèces non cultivables (c.-à-d. qui sont viables mais qui ne peuvent être mises en culture) peuvent contenir des fragments d'ADN détectables qui pourraient être utiles à des fins criminalistiques ou pour la détermination de l'origine du matériel. Diverses méthodes spécifiques existent déjà pour de nombreux microorganismes pathogènes, mais il y a peu de méthodes générales avec lesquelles une même technique adaptable peut être appliquée à de multiples espèces ou à des microorganismes qui n'ont pas été préalablement caractérisés. La puce à ADN à haute densité pourrait satisfaire à un bon nombre de ces critères et être un complément aux outils d'identification dont nous disposons actuellement. Elle peut servir, par exemple, à détecter l'ADN microbien qui ne correspond pas entièrement aux séquences d'ADN génomique connues, ce qui permet de détecter des variants microbiens qui, autrement, auraient pu passer inaperçus. Dans le rapport présenté ici, nous décrivons la conception et les essais préliminaires d'une puce à ADN à haute densité mise au point pour la détection et l'identification des microorganismes.

## Executive summary

---

### An Affymetrix Microarray Design for Microbial Genotyping

Barry Ford; Doug Bader; Yimin Shei; Cindy Ruttan; David Mah; DRDC Suffield TM 2009-183; Defence R&D Canada – Suffield; October 2009.

**Background:** DNA and RNA (nucleic acids) are the genetic material of bacterial and viral species. The composition (e.g. DNA sequence) of the genetic material can be used to determine unambiguously, the identity of the species down to the level of individual unique strains. There are in existence assays useful for the detection and identification of single target nucleic acid sequences in microbial samples. In surveillance or diagnosis, it may be necessary to screen for many biological agents without knowing which in particular is of interest. While it is possible to run multiple single assays in parallel, the ability to execute multiple assays simultaneously within a single assay run is limited. Thus testing for multiple species of micro-organism currently requires multiple assay runs.

There is also a need to develop toolsets that can assay microbial targets without extensive microbiological culture or analysis of the specific sample. For example, some samples that can't be cultured at all will nonetheless contain detectable DNA fragments which might be of value with respect to forensics or attribution of source. There are currently few general methods wherein a single adaptable tool can be applied to multiple species or to previously uncharacterized organisms.

The high-density DNA microarray has the potential to complement existing identification tools, especially for multiple species or strains, or samples which can't be cultured using conventional microbiology. Microarrays are a technology which permits the detection of many nucleic acid sequences in a single run, with identification of each detected sequence. The basic microarray is comprised of many individual DNA sequence targets on a tiny microscope slide. The Affymetrix platform represents the current state of the art in microarray density (more than 221,000 individual targets) and throughput. An advantage of the microarray platform is the ability to detect microbial DNA sequences that are not a perfect match to the DNA sequences on the microarray chip.

**Results:** In this report, the design and preliminary testing of a high-density Affymetrix DNA microarray for the purpose of microbial identification and detection is described. The microarray can discriminate between multiple species of interest using qualitative analysis.

**Significance:** The DNA microarray is a single adaptable high-density platform useful for detection, identification and discrimination of multiple threat agents simultaneously. It is a complementary diagnostic technology to existing low-density microbiological and assay systems.

**Future plans:** Detailed validation of the current microarray and comparison to other microarray systems is planned. Additional testing is required to fully assess the real-world value of the DNA microarray as a tool for diagnosis, detection, and identification of microbial samples. Mixed microbial DNA samples and samples with human DNA (a frequent element in diagnostic samples) will be evaluated.

# Sommaire

---

## An Affymetrix Microarray Design for Microbial Genotyping

**Barry Ford; Doug Bader; Yimin Shei; Cindy Ruttan; David Mah; DRDC Suffield TM 2009-183; R & D pour la défense Canada – Suffield; Octobre 2009.**

**Contexte :** L'ADN et l'ARN (acides nucléiques) constituent le matériel génétique des espèces bactériennes et virales. La composition (c.-à-d. la séquence des acides nucléiques) du matériel génétique peut servir à déterminer sans aucune ambiguïté l'identité d'une espèce jusqu'au niveau des souches individuelles uniques. Diverses méthodes servent à la détection et à l'identification de séquences d'acides nucléiques cibles dans les échantillons microbiens. Pour des besoins de surveillance ou de diagnostic, il peut être nécessaire de rechercher de nombreux agents biologiques sans savoir lesquels cibler plus particulièrement. Il est possible de faire plusieurs analyses parallèlement, mais peu de méthodes permettent de rechercher simultanément plusieurs microorganismes par une seule analyse. Ainsi, actuellement, lorsqu'il faut rechercher plusieurs microorganismes dans un échantillon, il faut le soumettre à plusieurs analyses.

Il faudrait aussi mettre au point des méthodes qui permettent de cibler des microorganismes sans qu'il soit nécessaire de faire des cultures ou des analyses avancées. Par exemple, certains échantillons ne se prêtant pas à la culture peuvent contenir des fragments d'ADN détectables qui pourraient être utiles à des fins criminalistiques ou pour la détermination de leur origine. Actuellement, il existe peu de méthodes générales avec lesquelles une technique adaptable peut être appliquée à de multiples espèces ou à des microorganismes qui n'ont pas été préalablement caractérisés.

La puce à ADN à haute densité pourrait être un complément utile des outils d'identification actuels, surtout pour les espèces ou souches multiples ou encore pour les échantillons qui ne se prêtent pas aux méthodes culturales de la microbiologie classique. La puce à ADN est une technologie qui permet la détection et l'identification de nombreuses séquences d'acide nucléique en une seule analyse. Essentiellement, une puce à ADN est une petite lamelle de microscope sur laquelle ont été déposées un grand nombre de séquences individuelles d'ADN cible. La plate-forme Affymetrix est actuellement le dernier cri de la technologie des puces à ADN à haute densité (plus de 221 000 cibles individuelles) et à débit élevé. L'un des avantages de cette plate-forme vient de ce qu'elle permet de détecter des séquences d'ADN microbien qui ne correspondent pas entièrement aux séquences utilisées sur la puce.

**Résultats :** Dans le rapport présenté ici, nous décrivons la conception et les essais préliminaires d'une puce à ADN à haute densité Affymetrix mise au point pour la détection et l'identification des microorganismes. Cette puce permet de différencier un grand nombre d'espèces d'intérêt par une analyse qualitative.

**Importance :** La puce à ADN est une plate-forme adaptable à haute densité qui peut servir pour la détection, l'identification et la différenciation simultanées de multiples agents dangereux. Cette technologie de diagnostic est un complément des systèmes de détection et d'analyse microbiologiques à faible densité.

**À venir :** Validation détaillée de la puce à ADN actuelle et comparaison avec d'autres systèmes à puce. D'autres essais seront nécessaires pour l'appréciation exacte de la valeur opérationnelle de la puce à ADN comme outil de diagnostic, de détection et d'identification des échantillons microbiens. Des échantillons d'ADN microbien mixtes et des échantillons contenant de l'ADN humain (dont la présence est fréquente dans les échantillons de diagnostic) seront évalués.

# Table of contents

---

|   |     |
|---|-----|
| Abstract .....  | i   |
| Résumé .....  | i   |
| Executive summary .....   | ii  |
| Sommaire .....  | iii |
| Table of contents .....   | v   |
| List of figures .....   | vi  |
| List of tables .....  | vi  |
| Acknowledgements .....  | vii |
| 1 Introduction.....   | 1   |
| 2 Methods .....   | 3   |
| 2.1 Microarray platform selection .....                                     | 3   |
| 2.2 Selection of targets .....  | 3   |
| 2.3 Target sequence extraction .....  | 4   |
| 2.4 Microarray in silico verification.....                                  | 5   |
| 2.5 Microarray fabrication.....   | 5   |
| Affymetrix photolithography.....  | 5   |
| 2.6 Microarray testing .....  | 7   |
| Microbial DNA samples.....  | 7   |
| Affymetrix DNA labeling.....  | 7   |
| 2.7 Data reduction and analysis.....  | 8   |
| 3 Results and Discussion .....  | 10  |
| 4 Conclusions.....  | 14  |
| References .....  | 16  |
| Annex A ..Species and strain-specific probes in the final array design..... | 19  |
| Annex B... Detailed Protocols for microarray labeling .....                 | 33  |
| List of symbols/abbreviations/acronyms/initialisms .....                    | 35  |

## List of figures

---

|   |    |
|---|----|
| Figure 1: The Affymetrix microarray system.....   | 2  |
| Figure 2: Photolithography of microarrays. (adapted from Affymetrix Inc., Santa Clara, CA)..... | 6  |
| Figure 3: End labelling sample DNA with terminal deoxynucleotide transferase (TdT).....         | 8  |
| Figure 4: Comparison of DNA from various E. coli replicates.....                                | 11 |
| Figure 5: Standard deviation versus mean signal value.....                                      | 12 |
| Figure 6: Plots of signal intensity versus position.....  | 13 |

## List of tables

---

|  |   |
|--|---|
| Table 1: DNA extracts used in initial testing of the microarray..... | 7 |
|--|---|

## **Acknowledgements**

---

The authors wish to thank Melissa Crichton and Kevin Ruttan for their technical assistance with the microarray work. The authors would also like to thank Chad Stratilo and Glen Fisher for preparations of DNA from microbial samples in the DRDC Suffield strain collection.

This page intentionally left blank.

# 1 Introduction

---

In the area of microbial genotyping there are multiple platforms that can identify one or a few microbial targets in a single assay iteration. For most pathogenic microbes, various specific methods exist, but there are few general methods wherein a single adaptable tool can be applied to multiple species or to previously uncharacterized organisms. There is a continuing need for the capability to detect or identify many possible microbial agents without having prior knowledge of the offending agent in a given sample [1,2]. Indeed, some samples that contain no live or culturable cells could contain detectable DNA fragments which may prove to be useful for clinical, forensic or attribution purposes [3]. One platform with potential to aid identification of multiple species or strains without culturing or specific prior sequence knowledge is the high-density microarray.

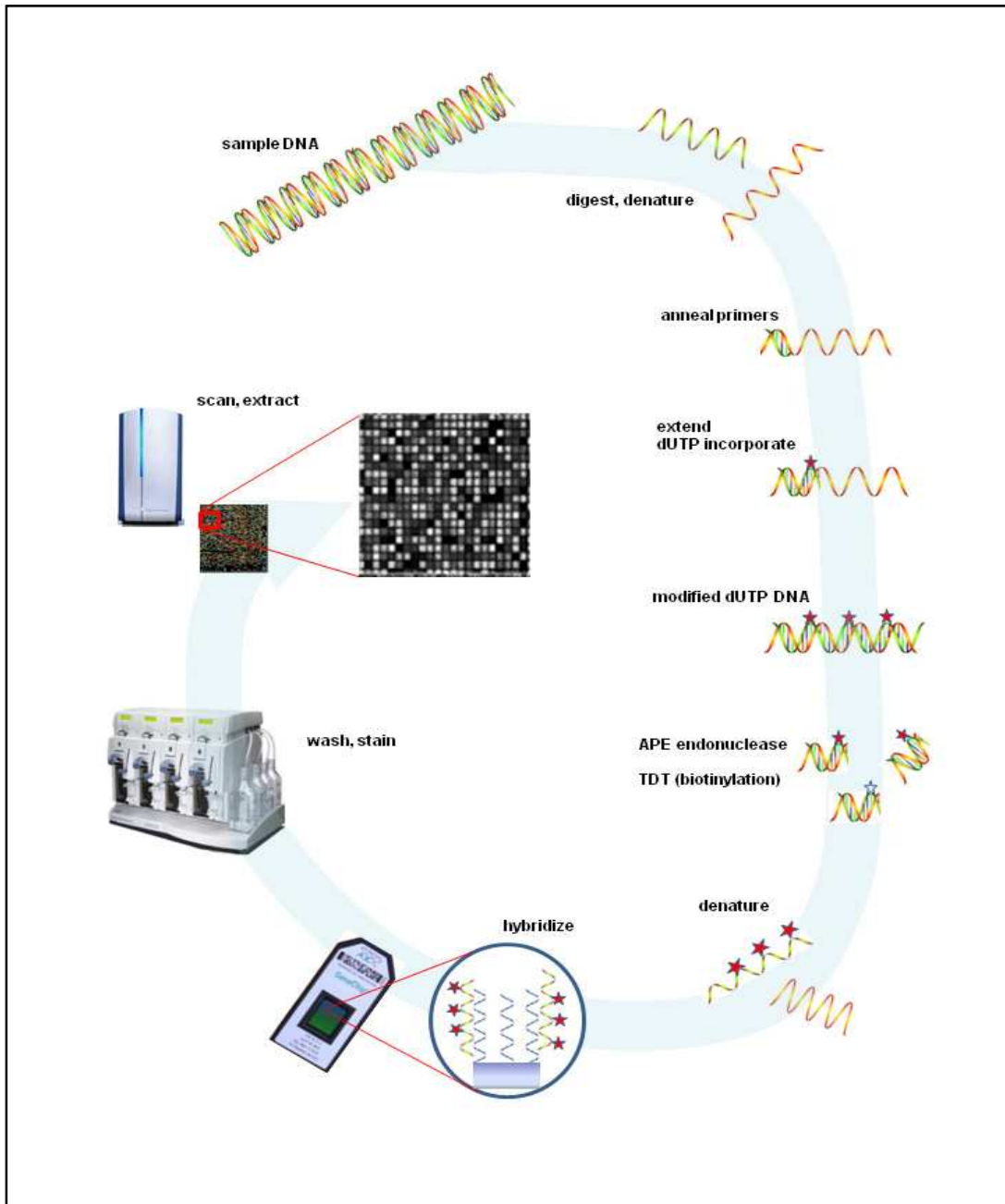
Each microarray can carry from a few hundred to a few hundred thousand individual target DNA sequences. Choice of specific array platform is driven by a combination of cost, density, and usability. For maximal utility, the ideal microarray should have as many features as possible, each feature representing one unique DNA sequence fragment. In the current work, the Affymetrix platform was exploited towards the development of a broad spectrum multi-species, multi-strain microarray, on a single microarray chip containing over 200,000 individual features. The Affymetrix system is closed source, meaning that the applied technologies for array fabrication, labeling, and data extraction are integrated into a pre-packaged system purchased from Affymetrix. Basic methods are established by the vendor, such that standardization of techniques takes much less time than with open source platforms. There are additional hardware costs for choosing Affymetrix relative to open source platforms, which are largely balanced by the reduced hands-on time for the pre-hybridization and post-hybridization manipulations, as well as for the extraction of data from the scanned microarray image. Figure 1 is a summary of the Affymetrix microarray processing system used at DRDC Suffield in support of this work. The protocol is itemized in detail in the Annexes.

In order to assess the utility of DNA microarrays for identification of bacterial pathogens to the species and strain level, a multipathogen chip was designed for the Affymetrix platform. Organisms included on the chip were derived from the National Institute of Allergy and Infectious Diseases Category A and B list of priority pathogens [4]. Also selected were *Haemophilus influenzae*, *Acinetobacter baumannii*, *Chaetomium* species, *Rickettsia* species, plasmids pBC16 and pLS1. Sequences representing bacterial toxins and antimicrobial resistance (*e.g.* antibiotic markers) were also sampled. Targets for viral pathogens were not included in this chip. The sequences thus chosen constituted approximately 16,000 individual sequence targets, which, allowing for sequence variants and internal controls, included over 81,000 unique probes. The remaining capacity of the chip surface was used to deploy some 140,000 probes from the Affymetrix "antigenomic library" to serve as non-targeted probes, essentially a random target library.

The number of microbial genomic targets thus did not equal the number of individual probes on the array. This is due to the redundancy built into the Affymetrix microarray technology, wherein variants of specific probe sequence differing by one or a few bases from the specific probe, are used to assess non-specific or variant binding to probe sites. In general, each specific target is represented by 3–20 individual probe sequences, varying by length, sequence, or single base pair

differences. In typical applications, only one signal is reported from a probe set, the remaining features serving as quality assurance and quality control indicators. For genomic fingerprinting, the variants related to the primary probe may also contain useful signals, and are also reported.

Figure 1: Overview of the Affymetrix microarray system.



APE : apurinic endonuclease; TDT : terminal deoxynucleotidyl transferase

## 2 Methods

---

### 2.1 Microarray platform selection

The two types of custom Affymetrix DNA microarray formats which could be used for genotyping are the Tiling Arrays and Resequencing Arrays. Choice of format in general governs the default reagent and protocols which can be accessed directly from Affymetrix. Probes on tiling arrays are spaced at a specified regular distances across a given genomic sequence (*e.g.* one probe every 500 bp). For resequencing arrays, the sequence for a region of interest is provided and four 25 base probes are designed for every base pair for the entire length of the sequence. Each member of a four probe set differs at the central (13th) bp at which an A, C, G or T will be incorporated. Probes are designed for both strands so that each base pair within the sequence is interrogated eight times.

Since the purpose of the multipathogen chip is to use probes derived from multiple source organisms, neither of the two standard custom designs fit our direct requirements. Through discussions with Affymetrix, a modified Tiling Array format was selected such that up to five probes would be provided for each sequence submitted for evaluation. In addition to control probes, whatever space remained on the array design would be filled with nonspecific probes selected from the existing Affymetrix probe library. The design contract with Affymetrix (executed as a subcontract with Canada West Biosciences) allows DRDC to retain ownership of our own probe designs, while using part of the Affymetrix probe library under licence.

### 2.2 Selection of targets

In principal, given the 25 base pair size of the oligonucleotide probes on the Affymetrix microarray, an ideal array could sample any possible sequence (known or unknown) if all possible 25-base oligonucleotides were spotted on the array. This would be A,C,G or T at all possible positions, which is  $4^{25}$  oligonucleotide sequences, or  $\sim 1.126 \times 10^{16}$  individual sequences. Current maximal capacity of the Affymetrix array system is approximately 1 million probes per array. It would require  $1 \times 10^{10}$  microarrays to cover most of the possible sequences. Thus, designing all possible 25 base pair sequences was not a practical nor an affordable approach.

Targeted probes are categorized as SNP (single nucleotide polymorphism) or non-SNP. SNP probes are included to better differentiate between strains of the same species. The target sequence submitted is 49 bp or less and contains one or more SNPs. For a sequence with one SNP, a set of 5 probes covering different segments of the target sequence is created for each of A, C, G and T at the variant base, resulting in a total of 20 probes. Thus, an organism with 'A' at the target site would register high intensity signal for the 5 'A' probes and low intensity for the remaining 15 probes. For sequences containing more than one variant within a 49 base pair region, the number of probes increases accordingly. The mismatch SNP probe variants of a specific probe sequence (differing by one or a few bases from the specific probe), are used to assess non-specific or variant binding to probe sites and are useful for genomic fingerprinting.

Non-SNP probes have little sequence commonality and are used to differentiate at the species level. The target sequences are much longer than those used for SNP design and the probes, ranging from 1 to 15 unique sequences, can be spread over a large section of genomic DNA. Multiple target sequences (SNP and non-SNP) were submitted for each organism of interest to

ensure detection even if some probes failed to perform as expected. Ideally, all probes designed using a specific organism's DNA sequence should produce high intensity signal while the remaining probes (off-target) should have little to no signal.

In order to rationally develop a library of probes suitable for identifying the maximal number of agents, we focused on regional microvariation within sequenced genomes of interest. Sequences for probe design included those that encompassed regions that differed between strains of the same species, especially those from Category A bacteria. Also included were regions that were constant within a species but differed between species, virulence genes, and antibiotic resistance genes.

## 2.3 Target sequence extraction

The first step in identifying regions of interest was to review the existing literature on bacterial microarray genotyping and strain differentiation. This provided a partial list of genes to include in our search. Next, various online databases were investigated for genes of interest. Initially, the NCBI Protein Clusters database [5] was used. Protein Clusters provides curated and non-curated clusters of related protein reference sequences. The database was searched by species and the protein clusters of that species were targeted by the level of conservation. For example, strain variants in *Bacillus anthracis* were identified by selecting clusters conserved to the *Bacillus cereus* or *Bacillus anthracis* group level. Selecting a cluster of interest revealed the list of all strains included in the cluster. Variants within the sequence were then identified by viewing the detailed alignment. When variants were found, the DNA sequence was retrieved by clicking on the locus tag, then on the sequence viewer.

Antibiotic resistance genes were obtained from the Antibiotic Resistance Genes Database [6]. The majority of the sequences used for probe selection were obtained from VFDB, the Virulence Factors of Pathogenic Bacteria database [7]. This database provided FASTA formatted (plaintext for database searches) sequences of virulence genes and sequences that can be used for comparative genomics. Additional strains of interest for inclusion in the microarray design were provided by Dr Kingsley Amoako (Canadian Food Inspection Agency, Lethbridge, Alberta). In addition, the coding sequences for hypoxanthine guanine phosphoribosyltransferase (HPT) and adenine phosphoribosyltransferase (APRT) from multiple species of origin were included for future applications. Species and strains represented on the microarray are itemized in Annex A.

A master Excel file was created in order to manage the selected sequence segments. This file contained a number of data points: probe name, organism used to obtain the sequence, gene ID/Accession Number/Locus ID used to locate the gene, start and end base positions of the sequence used, length of the sequence segment used, first and last 8 bp of the selected sub-sequence, gene name and description, the strain the sequence was derived from, and the complete sequence segment. To determine which strains matched which sequence (beyond and including the sequence source strain), sequences were initially tested using nucleotide BLAST tool [8] against nucleotide reference sequences, then whole-genome shotgun sequences. For strains that differed by single base pair variations, single-nucleotide polymorphism (SNP) target sequences were prepared. These probe sequences were 49 bases in length with the variant, designated by an "n", in the 25th (central) position.

From this master file, two files were prepared for Affymetrix. The first was the instruction file listing the probe name, start and end positions of the probe sequence, first and last 8 bp of the

probe sequence and a description of the probe. The second file included all the probe sequences in FASTA format, each identified by the probe name provided in the instruction file. Once Affymetrix received the instruction and FASTA sequence files, five 25-mer probes were designed for each probe sequence submitted, using Affymetrix proprietary software. Degenerate or redundant probes were removed and a list of the proposed microarray design was returned for evaluation. The final microarray design was assembled using 81,678 probes from 11,516 unique microbial sequences, 24,660 probes from 264 SNP sequences, and approximately 140,000 non-specific probes along with controls to fill in the 220,678-probe chip. Annex A contains the listing of species and strain-specific represented, and how many probes that were used in the final design.

## **2.4 Microarray in silico verification**

The straight text listing of oligonucleotide sequences printed onto the array was analyzed by testing all the sequences in the design against the entire NCBI non-redundant nucleotide database using iterative BLAST searches. PERL scripts were developed to run serial segments of the dataset (being too large to submit as a single set), and to log the returned sequence data, predicted species (and strain), and general annotations including accessions, of all hits against the submitted oligonucleotide sequences.

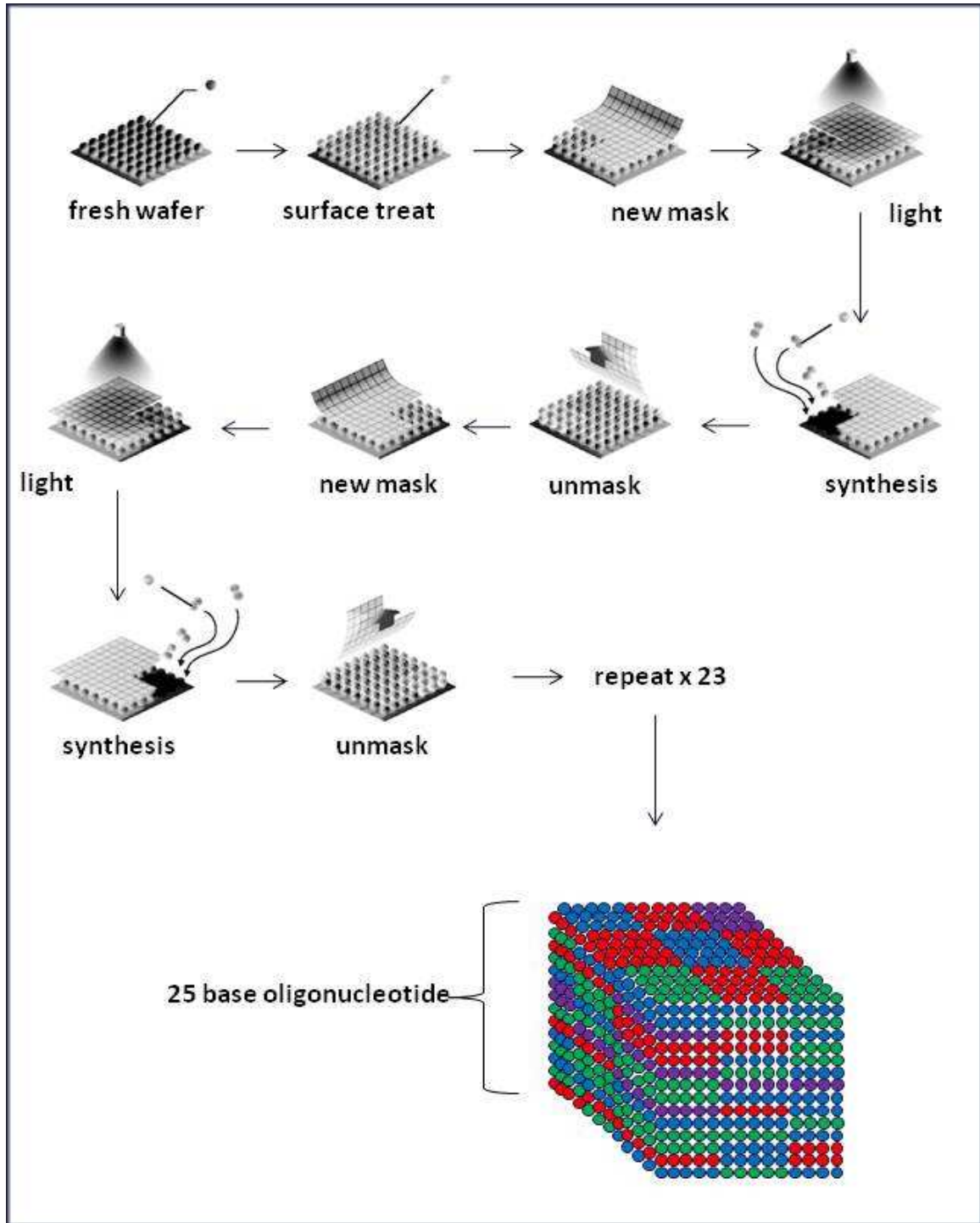
Since the entire feature sequence set was designed using publicly available databases, of which NCBI comprises a large, if not exhaustive aggregation, it was anticipated that *in silico* testing would recapitulate the designed species and strain identifications. It was also expected that due to sequence data errors and accession-specific variations within the database, that some sequences designed as unique probes (single species, single target) would actually align to accessions other than the record of origin.

## **2.5 Microarray fabrication**

### **Affymetrix photolithography**

The fabrication of the microarray per se is one key to the Affymetrix product package. Driven by the availability of photomicro lithography in microfabrication of microcircuitry, Affymetrix developed the method to template microarray chips using multiple lithographic overlays combined with photo-activatable oligonucleotide synthesis chemistry (Figure 2, adapted from Affymetrix Inc.). The technique allows for submicron precision in placement of oligonucleotide synthesis reactions on the surface of the silicon microarray wafer. Using multiple overlays, each site can be photoactivated differentially, and the different oligonucleotides synthesized stepwise. The more features on the array, the more overlays are required. Although photomicro lithography has been reported to produce some truncated probe sequences within each feature, the chip design includes truncation variants which can be used to verify signals from the features, or the absence of signal from the truncated probes as a set. This is generally done within the signal extraction software.

Figure 2: Photomicrolithography of microarrays. (adapted from Affymetrix Inc., Santa Clara, CA)



## 2.6 Microarray testing

### Microbial DNA samples

Table 1 lists the microbial DNA samples that were used for preliminary testing of the custom microarray design. DNA samples from *E. coli* were prepared by the contractor, while DNA from level 2 and level 3 microbes were prepared by DRDC Suffield staff in the DRDC Suffield BSL2 or BSL3 labs respectively. BSL2 and BSL3 DNA extracts were tested for product sterility using standard procedures in the containment facility prior to release for microarray testing.

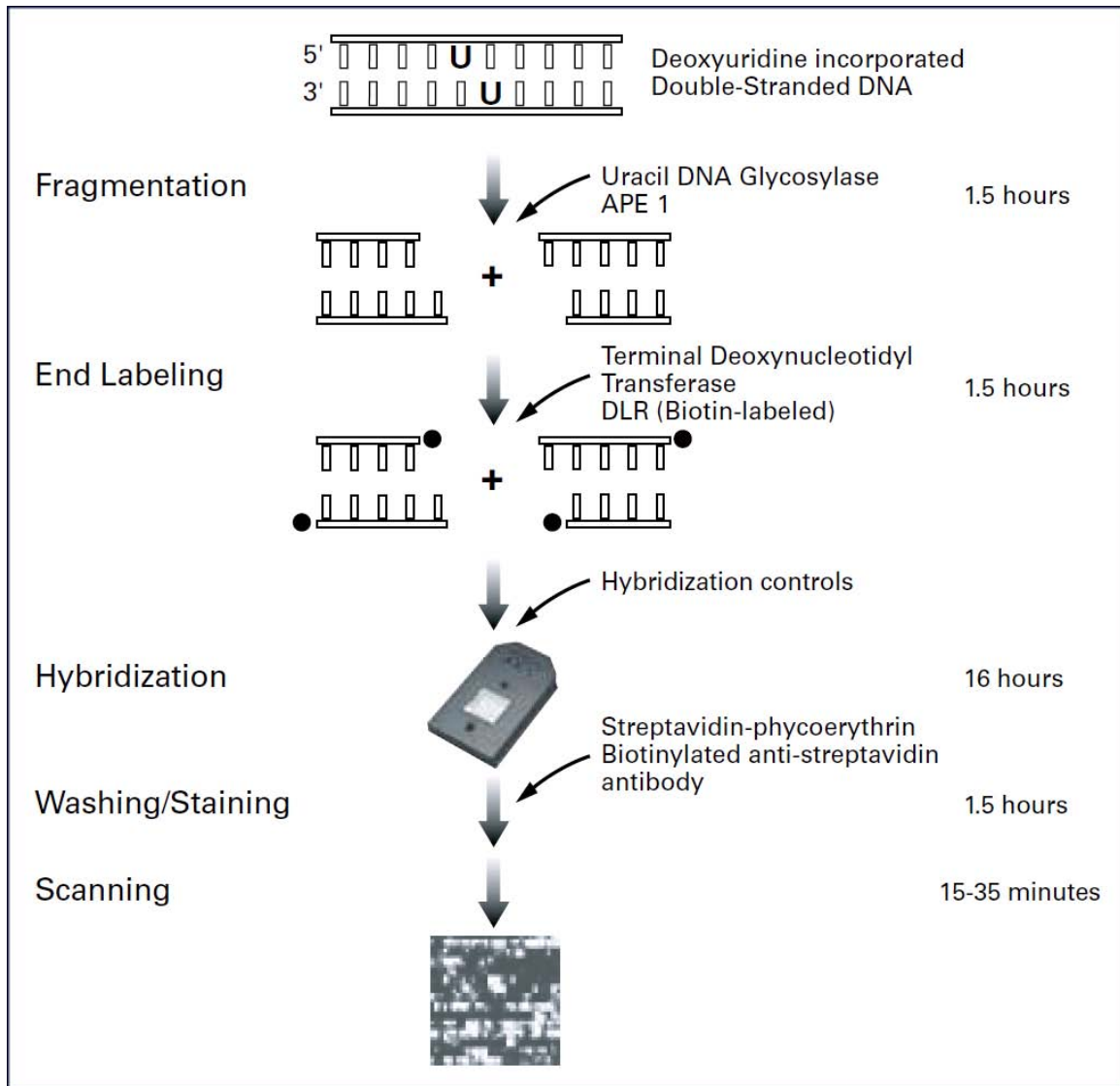
*Table 1: DNA extracts used in initial testing of the microarray*

| <b>Genus</b>       | <b>Species</b>        | <b>Strain / Isolate</b> |
|--------------------|-----------------------|-------------------------|
| <i>Escherichia</i> | <i>coli</i>           | JM109                   |
| <i>Bacillus</i>    | <i>anthracis</i>      | RP42                    |
| <i>Bacillus</i>    | <i>cereus</i>         | ATCC 11778              |
| <i>Yersinia</i>    | <i>pestis</i>         | 19428                   |
| <i>Yersinia</i>    | <i>enterocolitica</i> |                         |

### Affymetrix DNA labeling

The sample labeling method used in this study involved preparing a random-primed synthesis reaction incorporating uracil instead of thymine into the newly synthesized DNA using genomic DNA as the template, followed by direct end-labeling of the product DNA with biotinylated nucleotide, as shown in Figure 3. The detailed protocol used to label target DNA (per sample) is reproduced in Annex B:

Figure 3: End labelling sample DNA with terminal deoxynucleotide transferase (TdT).



## 2.7 Data reduction and analysis

Much of the data reduction work involved with microarrays involves feature extraction from the image to a spreadsheet, aligning the signals to the annotated target list, and data curation. These elements are contained within and managed by the Command Console software. Subsequent analysis involves normalization of data sets within the experimental series (between arrays), followed by comparison of test data to control data. In the case of genomic identification or fingerprinting, as on our array design, comparison to control arrays is not required for initial

assignment of genera and species. For the purposes of preliminary array design testing, no exhaustive comparison was undertaken.

Using tools developed for the open source microarray system (Chromablast) [9], data were reviewed without requiring prior normalization of signals. We compared data sets from *E. coli*, *B. anthracis*, *Y. pestis* and *Y. enterocolitica*, to determine qualitatively whether the target-specific array elements could discriminate between the samples. Data from the non-specific probe sets were not considered in this initial assessment. Student's t tests were performed pairwise on data sets to determine whether this analysis was informative relative to the heat map display.

### 3 Results and Discussion

---

Data were obtained on the Affymetrix custom-designed microarray for DNA from *Escherichia coli*, *Yersinia pestis*, *Yersinia enterocolitica* and *Bacillus anthracis*. Although a number of analytical tools are available for comparing and estimating distance in genomic fingerprint data [10,11], for this verification of concept and function, qualitative comparison was sufficient.

Using Chromablast [9], a heat map representing relative signal values was developed for a series of technical replicates of *E. coli* used for the microarray testing. An excerpt showing a region of the resultant heat map is shown in Figure 4. Uniform heat map colour across the replicates would indicate perfect concordance between replicates. The excerpted region shows some examples of this, as well as some replicates with varying colour, indicating some variation across the replicates. In Figure 4, green represents low value intensities (*i.e.* background to about 12% of maximum intensity, 0 to 6 in log base 2), and bright red indicates maximal intensity, as indicated in the scale below the heat map. The absolute scale of variation between non-normalized array data sets is thus seen to be about  $\pm 30\%$  within individual probe sets. This is verified by numerical analysis of the raw intensity data. Most of this variation is concentrated within the lower intensity values, where the standard deviation as a fraction of the mean is maximal. Above the mean signal intensity ( $\sim 7.0$  in log 2), the maximum signal variation per probe set is about  $\pm 5\%$  (Figure 5).

In practice, this suggests that a pruning of low-intensity signals may be useful to refine discrimination between samples versus knowns. Alternatively, a weighting factor could be applied to bias discriminatory decisions towards higher intensity signals. One method to compensate for signal variation between replicate arrays is to use the Student's t test to compare knowns to unknowns, or to detect outliers within the replication set for a known sample. In the case of the *E. coli* replicated data set, for the complete data set, including the lowest value probe intensities (15,533 probe sets), less than 2% of all signals in a pairwise comparison have a t-test value of less than 0.05. If only the signals greater than background are considered (9335 probe sets), the number of t-test values less than 0.05 falls to  $\sim 1\%$  (72 probe sets). "Significant" t-test results obtained for low-intensity signals (low confidence) are removed by this strategy. Small occurrences of outlier or systematically unreliable signal sets, as indicated by this analysis, are unlikely to interfere with discrimination between different genera or species, but may complicate detailed discrimination between closely related strains.

An initial survey of *E. coli*, *B. anthracis*, *Y. pestis* and *Y. enterocolitica* on the microarray revealed that even at a high level view, the array could easily discriminate *E. coli* from *B. anthracis* (Figure 6-A,C). Note that these plots are of data not normalized post extraction, since the Affymetrix signal processing software applies in-process normalization using the internal controls. Comparison of the *Y. pestis* versus *E. coli* data suggested that the *Y. pestis* DNA sample contained DNA from *E. coli* or a related species.

Figure 4: Comparison of DNA from various *E. coli* replicates.  
See text for detailed description.

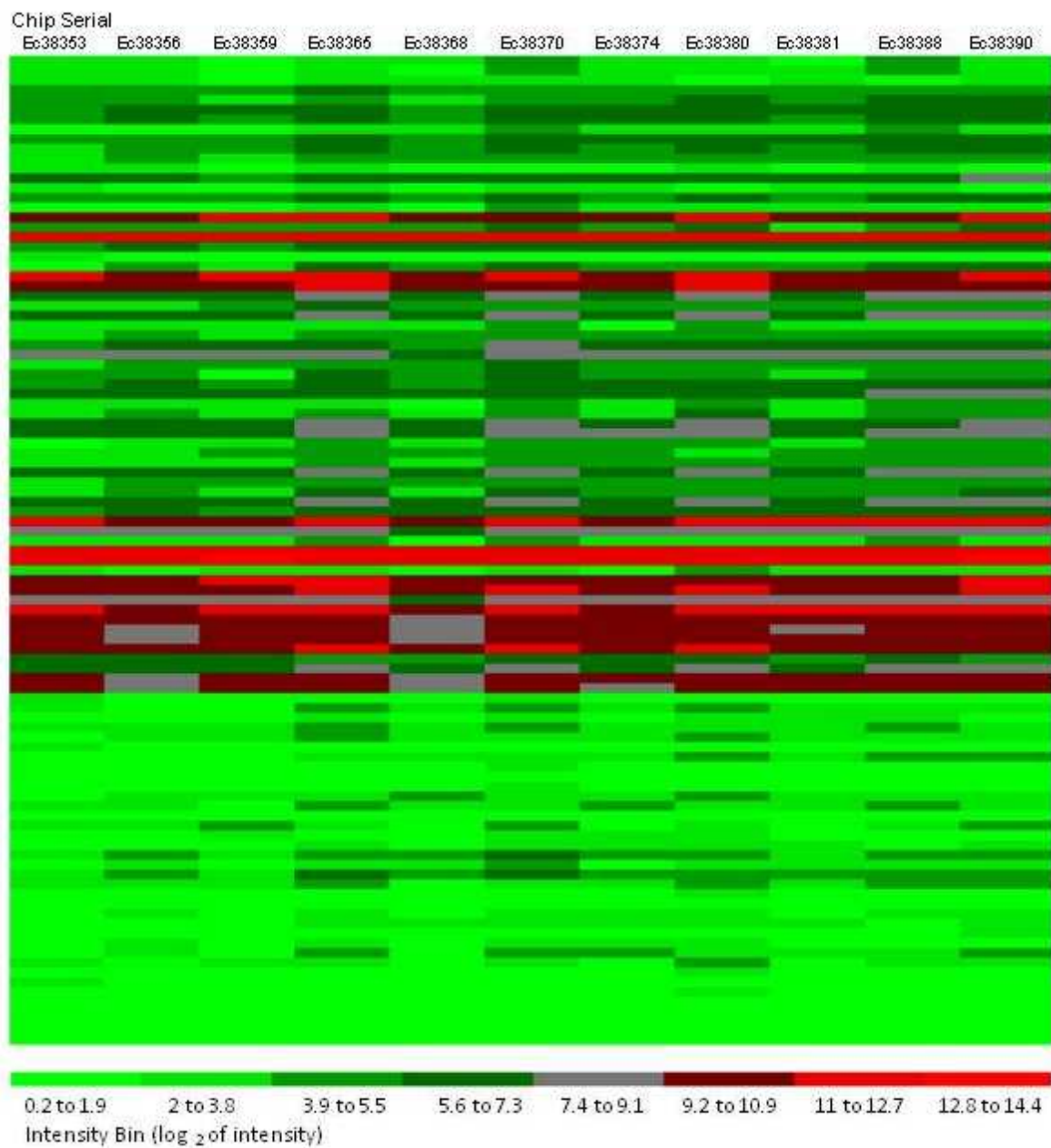
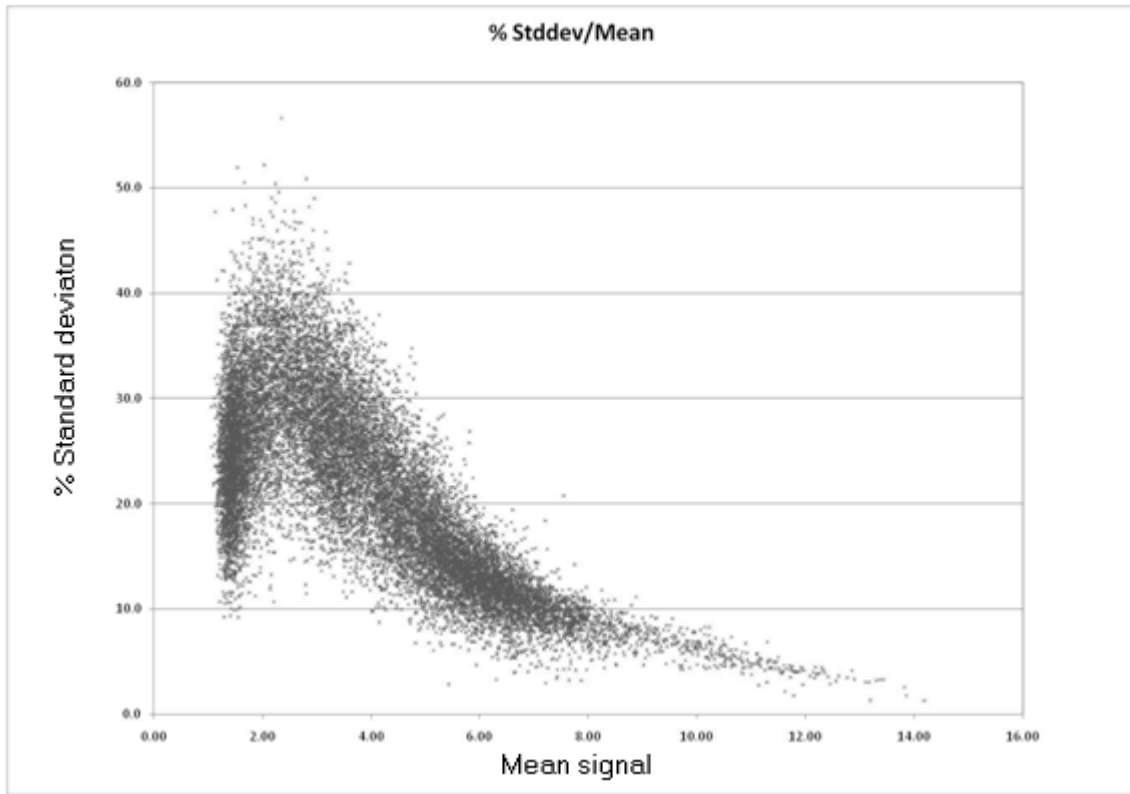
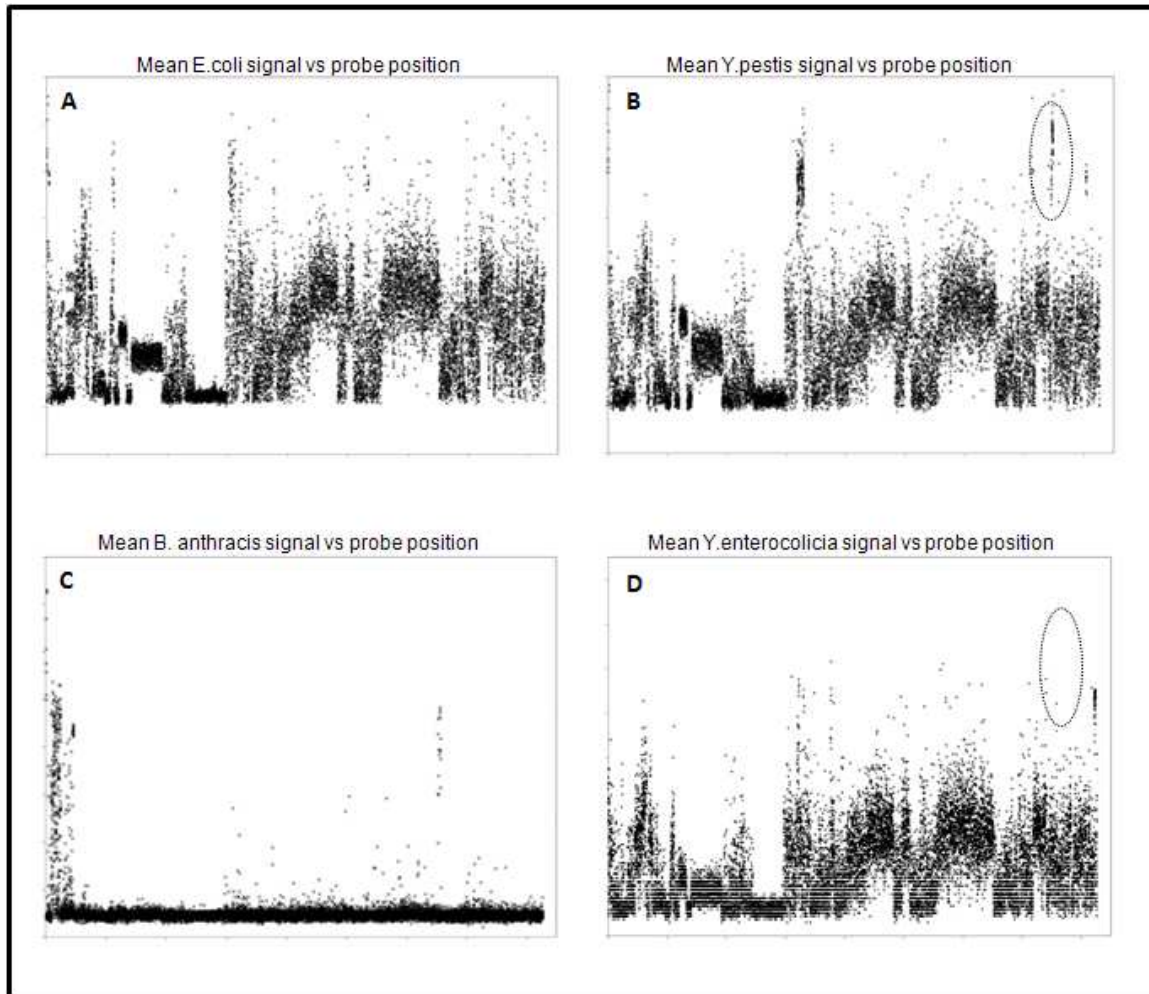


Figure 5: Standard deviation versus mean signal value.



Closer inspection of the plot in Figure 6-B showed the presence of signals not seen in the *E. coli* or the *Y. enterocolitica* samples (circled region in 6-B,D), indicating that this sample contained some unique DNA, and thus was either contaminated during preparation, or in the original culture stock. Upon detailed review of the signal data, it was apparent that even though the putative *Y. pestis* sample did indeed contain many signals similar to the *E. coli* pattern, a unique series of features designed to yield *Y. pestis*-specific signals, verified by in silico analysis, did indeed result in uniquely high signals for the *Y. pestis* data set, but not for any other data set. Our conclusion is that this sample did indeed contain *Y. pestis* DNA, but had suffered some contamination event either in original culture or in subsequent DNA preparation. Signals from both the intended DNA (*Y. pestis*) and from the contaminant were identified in this sample. In these plots, the *Y. enterocolitica* sample also appears to very like the *E. coli* sample, but differs noticeably from the *Y. pestis*.

Figure 6: Plots of signal intensity versus position.  
See text for discussion.



## 4 Conclusions

---

The application of microarrays to microbial genotyping or fingerprinting is a technical compromise of time and difficulty versus data density. Single target or multiplex real-time PCR assays are faster and can be quantitative. Real-time PCR assays can in principle detect 2–4 targets per assay reaction, based on positive detection of specific sequences in known genomic targets. Routine PCR assays however, are not the best method of choice for detecting recombinants, variants, or the presence of non-target organisms. If an assay system could run hundreds of PCR reactions for each test sample, the analytical density of the microarray could be equaled. Typical microarray open source platforms can detect 20–50,000 targets per array, using a single labeling or amplification reaction. Open source microarray systems typically take 18–26 hours for a single execution, but each run encompasses the equivalent of 1–2 thousand multiplex PCR reactions.

Compared to PCR or gel electrophoresis assays, microarrays appear to be very expensive [13–15]. Microarray platforms are, for now at least, clear winners when the multiplex capabilities of the array systems are compared to comparable efforts on other platforms. Operating costs between open source microarray platforms and the Affymetrix system are similar, despite the higher hardware and consumables costs for the Affymetrix system. For example, the Affymetrix software package Command Console contains an integrated suite of tools for feature extraction, system quality assurance, and data curation. Required hands-on time for the feature extraction (from image to signal data on a spreadsheet) is measured in minutes, compared to the open source microarray system, where each array requires 1-3 hours of manual image data extraction. With the minimal hands-on time required for the post-hybridization, the automation features of the Affymetrix system represent its greatest operational advantage over the open source microarray platforms. Although initial costs are greater with the Affymetrix system, it seems likely that the cost differential will be very small once the accumulated savings in time and labour are considered.

The number of assays executed per microarray has the drawback that for some material sources, DNA from multiple species is likely to be present and may contribute to the measured signals [15]. If the microarray contains sufficient numbers of features and has a high degree of automation, endemic species are always going to give a signal, thus the mere presence of a signal of such a species in a given environmental or clinical context is not in itself meaningful [14-16]. Assays must be combined with other indices of suspicion (clinical signs, known exposure, suspect samples) in order to determine whether a given positive represents a real diagnosis or threat [1, 2, 17]. This is also true for most other molecular or microbiological assays currently in use. Simple detection of agent is insufficient to establish a diagnosis in a clinical setting.

In addition, as the sensitivity of assay systems improves (due to non-specific genomic amplification for example), out-of-context true positive signals (not within the normal range of endemics) may be detected [3,15,17]. Such signals may be due to sample contamination by workers, gratuitous sampling of infrequent (but locally intense) organism populations, or previously undetected genetic similarity between lab strains and endemic strains. Use of confirmatory assays of high-specificity (e.g. real-time PCR) will complement such data.

Given the requirement for technical expertise in operating a microarray system, and given the sensitivity to multiple targets in some samples, microarray systems will continue to require laboratory support. Microarray systems are in use in clinical centers, but point-of-care microarray systems are not imminent. On the other hand, time-to-result times are comparable to or better than conventional microbiology. Detailed testing of the current microarray and comparison to other microarray systems is underway. Additional testing with an expanded library of DNA samples and a wider sampling of species is required to fully assess the value of the microarray as a tool for diagnosis, detection, and identification of microbial samples.

## 5 References

---

- [1] Avashia, S., Riggins, W., Lindley, C., Hoffmaster, A., Drumgoole, R., Nekomoto, T., Jackson, P., Hill, K., Williams, K., Lehman, L., Libal, M., Wilkins, P., Alexander, J., Tvaryanas, A., Betz, T. 2006. Fatal pneumonia among metalworkers due to inhalation exposure to *Bacillus cereus* containing *Bacillus anthracis* toxin genes. *Clin Infect. Dis.* 44: 414-416.
- [2] Hoffmaster, A., Ravel, J., Rasko, D., Chapman, G., Chute, M., Marston, C., Barun, B., Sacchi, C., Fitzgerald, C., Mayer, L., Maiden, M., Priest, F., Barker, M., Jinag, L., Cer, R., Rilstone, J., Peterson, S., Weyant, R., Galloway, D., Read, T., Popovic, T., Fraser, C. 2004. Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. *Proc. Nat. Acad. Sci.* 101: 8449-8454.
- [3] Tucker, J.B. ed. 1997. The utility of sampling and analysis for compliance monitoring of the Biological Weapons Convention. Workshop Proceedings, The Center for Global Security Research, Livermore, CA.
- [4] Biodefense Category A,B,C Pathogens (online), National Institute of Allergies and Infectious Diseases, <http://www.niaid.nih.gov> (Access date: 2 Feb 2008).
- [5] Protein Clusters Home (online), National Center for Biotechnology Information, U.S. National Library of Medicine, <http://www.ncbi.nlm.nih.gov/sites/entrez?db=proteinclusters> (Access date: 25 Jan. 2008).
- [6] ARDB - Antibiotic Resistance Genes Database (online), Center for Bioinformatics and Computational Biology, University of Maryland, <http://arbd.cbcb.umd.edu/> (Access date: 24 Jan. 2008)
- [7] VFDB: Virulence Factors of Bacterial Pathogens (online), Institute of Pathogen Biology, State Key Laboratory for Molecular Virology and Genetic Engineering, Beijing, <http://www.mgc.ac.cn/VFs/main.htm> (Access date: 2 Feb. 2008).
- [8] BLAST: Basic Local Alignment Search Tool (online), National Center for Biotechnology Information, U.S. National Library of Medicine, (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>, (Access date: 10 Jan. 2008).)
- [9] Ford, B.N., Shei, Y., Bjarnason, S., Richardson, C. 2006. ChromaBlast—A Data Visualization Tool. DRDC Suffield TM 2006-049. Defence R&D Canada - Suffield.
- [10] Willse, A., Chandler, D., White, A., Protic, M., Daly, D., Wunschel, S. 2005. Comparing bacterial DNA microarray fingerprints. *Sat. Appl. Genet. Mol. Biol.* 4: 1-36.
- [11] Doran, M., Raicu, D., Furst, J., Settimi, R., Schipma, M., Chandler, D. 2007. Oligonucleotide microarray identification of *Bacillus anthracis* strains using support vector machines. *Bioinformatics* 23: 487-492.

- [12] Borucki, M., Kim, S., Call, D., Smole, S., Pagotto, F. 2004. Selective discrimination of *Listeria monocytogenes* epidemic strains by a mixed-genome DNA microarray compared to discrimination by pulsed-field gel electrophoresis, ribotyping, and multilocus sequence typing. *J. Clin. Microbiol.* 42: 5270-5276.
- [13] Malanoski, A., Lin, B., Wang, Z., Schnurr, J., Stenger, D. 2006. Automated identification of multiple micro-organisms from resequencing DNA microarrays. *Nuc. Acids. Res.* 34: 5300-5311.
- [14] Tembe, W., Zavaljevski, N., Bode, E., Chase, C., Geyer, J., Wasieloski, L., Benson, G., Reifman, J. 2007. Oligonucleotide fingerprint identification for microarray-based pathogen diagnostic assays. *Bioinformatics* 23: 5-13.
- [15] Kim, B., Park, J., Gu, M. 2005. Multiple and simultaneous detection of specific bacteria in enriched bacterial communities using a DNA microarray chip with randomly generated genomic DNA probes. *Anal. Chem.* 77: 2311-2317.
- [16] Rehrauer, H., Schönmann, S., Eberl, L., Schlapbach, R. 2008. PhyloDetect: a likelihood-based strategy for detecting micro-organisms with diagnostic microarrays. *Bioinformatics* 24: i83-i89.
- [17] Chandler, D., Alferov, O., Chernov, B., Daly, D., Golova, J., Perov, A., Protic, M., Robison, R., Schipma, M., White, A., Willse, A. 2006. Diagnostic oligonucleotide microarray fingerprinting of *Bacillus* isolates. *J. Clin. Microbiol.* 44: 244-250.

This page intentionally left blank.

## Annex A Species and strain-specific probes in the final array design.

---

| Organism                           | Strain / details           | Probes on Array |
|------------------------------------|----------------------------|-----------------|
| <i>Acinetobacter baumannii</i>     | ACICU                      | 38              |
| <i>Acinetobacter baumannii</i>     | ATCC 17978                 | 54              |
| <i>Acinetobacter baumannii</i>     | AYE                        | 143             |
| <i>Acinetobacter baumannii</i> SNP | baumannii                  | 20              |
| <i>Acinetobacter baumannii</i>     | baumannii                  | 15              |
| <i>Acinetobacter baumannii</i>     | plasmid pSUN-5             | 5               |
| <i>Acinetobacter baumannii</i>     | SDF                        | 44              |
| <i>Acinetobacter baumannii</i> HPT | ATCC 17978                 | 5               |
| <i>Acinetobacter baumannii</i> HPT | AYE                        | 5               |
| <i>Acinetobacter baumannii</i> HPT | baumannii                  | 5               |
| <i>Bacillus anthracis</i>          | Ames ancestor              | 140             |
| <i>Bacillus anthracis</i>          | Ames ancestor plasmid pX01 | 5               |
| <i>Bacillus anthracis</i>          | Ames ancestor plasmid pX02 | 25              |
| <i>Bacillus anthracis</i>          | anthracis                  | 45              |
| <i>Bacillus anthracis</i>          | Australia 94               | 6               |
| <i>Bacillus anthracis</i>          | Kruger                     | 5               |
| <i>Bacillus anthracis</i>          | Sterne                     | 55              |
| <i>Bacillus anthracis</i> APRT     | A2012 plasmid pX01         | 5               |
| <i>Bacillus anthracis</i> APRT     | Ames                       | 5               |
| <i>Bacillus anthracis</i> HPT      | A0442                      | 5               |
| <i>Bacillus anthracis</i> HPT      | anthracis                  | 5               |
| <i>Bacillus anthracis</i> HPT      | anthracis                  | 10              |
| <i>Bacillus anthracis</i> plasmid  | Sterne plasmid pX01+pX02-  | 10              |
| <i>Bacillus anthracis</i> SNP      | A2012                      | 20              |
| <i>Bacillus anthracis</i> SNP      | anthracis                  | 780             |
| <i>Bacillus anthracis</i> SNP      | other anthracis            | 200             |
| <i>Bacillus anthracis</i> SNP      | W. North America           | 20              |
| <i>Bacillus cereus</i>             | ATCC 10987                 | 179             |
| <i>Bacillus cereus</i>             | ATCC 14579                 | 200             |
| <i>Bacillus cereus</i>             | B. cereus plasmid pBCX01   | 5               |
| <i>Bacillus cereus</i>             | E33L                       | 45              |
| <i>Bacillus cereus</i>             | G9241                      | 5               |
| <i>Bacillus cereus</i> group SNP   | Bacillus                   | 1800            |
| <i>Bacillus cereus</i> HPT         | E33L                       | 5               |
| <i>Bacillus cereus</i> SNP         | ATCC 10987                 | 80              |
| <i>Bacillus cereus</i> SNP         | ATCC 14579                 | 320             |

|  |                          |     |
|--|--------------------------|-----|
| <i>Bacillus cereus</i> SNP             | B. cereus plasmid pBCXO1 | 320 |
| <i>Bacillus cereus</i> /anthracis SNP  | B. cereus plasmid pBCXO1 | 20  |
| <i>Bacillus amyloliquefaciens</i> APRT | FZB42                    | 5   |
| <i>Bacillus clausii</i>                | KSM-K16                  | 5   |
| <i>Bacillus clausii</i> APRT           | KSM-K16                  | 5   |
| <i>Bacillus halodurans</i>             | C-125                    | 5   |
| <i>Bacillus halodurans</i> APRT        | C-125                    | 5   |
| <i>Bacillus licheniformis</i>          | ATCC 14580               | 35  |
| <i>Bacillus licheniformis</i> APRT     | ATCC 14580               | 5   |
| <i>Bacillus pumilus</i> APRT           | SAFR-032                 | 5   |
| <i>Bacillus subtilis</i>               | 168                      | 25  |
| <i>Bacillus subtilis</i> APRT          | 168                      | 5   |
| <i>Bacillus thuringiensis</i>          | 97-27                    | 115 |
| <i>Bacillus thuringiensis</i>          | AI Hakam                 | 130 |
| <i>Bartonella bacilliformis</i>        | ATCC 35685               | 175 |
| <i>Bartonella henselae</i>             | Houston-1                | 270 |
| <i>Bartonella quintana</i>             | Toulouse                 | 245 |
| <i>Bartonella tribocorum</i>           | CIP 105476               | 330 |
| <i>Bordetella</i> SNP                  | Bordetella               | 20  |
| <i>Bordetella avium</i>                | 197N                     | 440 |
| <i>Bordetella avium</i> APRT           | 197N                     | 5   |
| <i>Bordetella bronchiseptica</i> APRT  | RB50                     | 5   |
| <i>Bordetella bronchiseptica</i>       | RB50                     | 75  |
| <i>Bordetella parapertussis</i>        | 12822                    | 268 |
| <i>Bordetella pertussis</i> APRT       | Tohama I                 | 5   |
| <i>Bordetella pertussis</i>            | Bordetella               | 5   |
| <i>Bordetella pertussis</i>            | Tohama I                 | 615 |
| <i>Bordetella petrii</i> APRT          | DSM 12804                | 5   |
| <i>Borrelia afzelii</i> APRT           | PKo                      | 5   |
| <i>Brucella</i> SNP                    | 9-941                    | 20  |
| <i>Brucella</i>                        | all brucella             | 250 |
| <i>Brucella</i> HPT                    | all brucella             | 5   |
| <i>Brucella abortus</i>                | 9-941                    | 30  |
| <i>Brucella abortus</i>                | S19                      | 45  |
| <i>Brucella abortus</i> SNP            | melitensis/abortus       | 40  |
| <i>Brucella abortus</i> APRT           | 9-941                    | 5   |
| <i>Brucella abortus/melitensis</i> SNP | abortus/melitensis       | 20  |
| <i>Brucella abortus/suis</i> SNP       | abortus/suis             | 20  |
| <i>Brucella canis</i>                  | ATCC 23365               | 15  |
| <i>Brucella canis</i>                  | S19                      | 5   |
| <i>Brucella canis</i> HPT              | ATCC 23365               | 10  |
| <i>Brucella melitensis</i>             | 16M                      | 427 |
| <i>Brucella melitensis</i>             | 2308 bv Abortus          | 210 |
| <i>Brucella melitensis</i>             | bv Melitensis            | 10  |
| <i>Brucella melitensis</i>             | bv Suis 686              | 5   |
| <i>Brucella ovis</i>                   | ATCC 25840               | 82  |

|  |                     |      |
|--|---------------------|------|
| <i>Brucella ovis</i>                   | bv Abortus 2308     | 35   |
| <i>Brucella suis</i>                   | 1330                | 25   |
| <i>Brucella suis</i>                   | ATCC 23445          | 5    |
| <i>Brucella suis</i>                   | ATCC 23447          | 5    |
| <i>Brucella suis</i>                   | ATCC 25840          | 10   |
| <i>Brucella suis</i>                   | bv. 4 str. 40       | 15   |
| <i>Brucella suis/abortus</i> SNP       | suis/abortus        | 80   |
| <i>Burkholderia</i> SNP                | Burkholderia        | 1160 |
| <i>Burkholderia</i> HPT                | all burkholderia    | 5    |
| <i>Burkholderia mallei</i> APRT        | ATCC 23344          | 5    |
| <i>Burkholderia mallei</i>             | ATCC 23344          | 40   |
| <i>Burkholderia mallei</i>             | PRL-20              | 5    |
| <i>Burkholderia multivorans</i> APRT   | ATCC 17616          | 5    |
| <i>Burkholderia pseudo/mallei</i> SNP  | pseudomallei/mallei | 20   |
| <i>Burkholderia pseudo/mallei</i>      | Burkholderia        | 15   |
| <i>Burkholderia pseudomallei</i>       | 668                 | 10   |
| <i>Burkholderia pseudomallei</i>       | 1710b               | 5    |
| <i>Burkholderia pseudomallei</i>       | 392f                | 5    |
| <i>Burkholderia pseudomallei</i> SNP   | B7210               | 40   |
| <i>Burkholderia pseudomallei</i>       | K96243              | 75   |
| <i>Burkholderia pseudomallei</i> SNP   | pseudomallei        | 340  |
| <i>Burkholderia pseudomallei</i>       | pseudomallei        | 5    |
| <i>Burkholderia pseudomallei</i>       | T18-1984            | 5    |
| <i>Burkholderia pseudomallei</i> HPT   | 91                  | 5    |
| <i>Burkholderia pseudomallei</i> HPT   | 668                 | 5    |
| <i>Burkholderia pseudomallei</i> HPT   | NCTC 13177          | 5    |
| <i>Burkholderia pseudomallei</i> APRT  | 668                 | 5    |
| <i>Burkholderia thailandensis</i> APRT | E264                | 5    |
| <i>Campylobacter concisus</i> APRT     | 13826               | 5    |
| <i>Campylobacter fetus</i>             | 82-40               | 440  |
| <i>Campylobacter hominis</i> APRT      | ATCC BAA-381        | 5    |
| <i>Campylobacter jejuni</i> APRT       | doylei 269.97       | 5    |
| <i>Campylobacter jejuni</i>            | 269.97 ss doylei    | 476  |
| <i>Campylobacter jejuni</i>            | 81116 (NCTC 11828)  | 351  |
| <i>Campylobacter jejuni</i>            | 81-176              | 349  |
| <i>Campylobacter jejuni</i>            | jejuni              | 60   |
| <i>Campylobacter jejuni</i>            | NCTC 11168          | 560  |
| <i>Campylobacter jejuni</i>            | plasmid pCjA13 t    | 5    |
| <i>Campylobacter jejuni</i>            | RM 1221             | 304  |
| <i>Campylobacter jejuni</i> APRT       | 81-176              | 5    |
| <i>Campylobacter jejuni</i> plasmid    | 81-176 plasmid pVir | 5    |
| <i>Chaetomium atrobrunneum</i>         | atrobrunneum        | 5    |
| <i>Chaetomium funicola</i>             | funicola            | 29   |
| <i>Chaetomium funicola</i>             | OC13                | 5    |
| <i>Chaetomium funicola</i>             | olrim130            | 5    |
| <i>Chaetomium thermophilum</i>         | CT2                 | 20   |

|                                     |                              |      |
|-------------------------------------|------------------------------|------|
| <i>Chaetomium thermophilum</i>      | MTCC 6350                    | 5    |
| <i>Chaetomium thermophilum</i>      | thermophilum                 | 85   |
| <i>Chlamydia abortus</i>            | S26/3                        | 115  |
| <i>Chlamydia caviae</i>             | GPIC                         | 115  |
| <i>Chlamydia felis</i>              | Fe/C-56                      | 120  |
| <i>Chlamydia muridarum</i>          | Nigg (MoPn)                  | 115  |
| <i>Chlamydia pneumoniae</i>         | AR39                         | 115  |
| <i>Chlamydia pneumoniae</i>         | CWL 029                      | 5    |
| <i>Chlamydia trachomatis</i>        | D/UW-3/CX                    | 175  |
| <i>Chlamydia trachomatis</i>        | HAR-13                       | 15   |
| <i>Chlamydia trachomatis</i>        | trachomatis                  | 5    |
| <i>Clostridium botulinum</i> APRT   | Alaska E43                   | 10   |
| <i>Clostridium botulinum</i> APRT   | ATCC 3502                    | 5    |
| <i>Clostridium botulinum</i> APRT   | Eklund 17B                   | 5    |
| <i>Clostridium botulinum</i> APRT   | Okra                         | 5    |
| <i>Clostridium botulinum</i>        | A str. ATCC 19397            | 5    |
| <i>Clostridium botulinum</i>        | ATCC 3502                    | 40   |
| <i>Clostridium botulinum</i>        | B str. Eklund 17B            | 5    |
| <i>Clostridium botulinum</i> SNP    | B1 str. Okra plasmid pCLD    | 20   |
| <i>Clostridium botulinum</i>        | B1 str. Okra plasmid pCLD    | 5    |
| <i>Clostridium botulinum</i>        | Bf                           | 5    |
| <i>Clostridium botulinum</i> SNP    | botulinum                    | 1860 |
| <i>Clostridium botulinum</i>        | C str. Eklund                | 5    |
| <i>Clostridium botulinum</i> SNP    | C. botulinum A strains       | 100  |
| <i>Clostridium botulinum</i>        | C. botulinum A strains       | 5    |
| <i>Clostridium botulinum</i>        | <i>Clostridium botulinum</i> | 15   |
| <i>Clostridium botulinum</i>        | Hall 183                     | 5    |
| <i>Clostridium botulinum</i> HPT    | Alaska E43                   | 15   |
| <i>Clostridium botulinum</i> HPT    | Eklund 17B                   | 10   |
| <i>Clostridium botulinum</i> HPT    | Loch Maree                   | 20   |
| <i>Clostridium botulinum</i> HPT    | Okra                         | 5    |
| <i>Clostridium botulinum</i>        | A3 str. Loch Maree           | 5    |
| <i>Clostridium acetobutylicum</i>   | ATCC 824                     | 25   |
| <i>Clostridium beijerinckii</i>     | NCIMB 8052                   | 20   |
| <i>Clostridium difficile</i>        | 630                          | 45   |
| <i>Clostridium difficile</i>        |                              | 15   |
| <i>Clostridium difficile</i> HPT    | difficile                    | 5    |
| <i>Clostridium kluyveri</i> APRT    | DSM 555                      | 5    |
| <i>Clostridium novyi</i>            | ATCC19402                    | 45   |
| <i>Clostridium novyi</i>            | NT                           | 40   |
| <i>Clostridium perfringens</i> APRT | SM101                        | 5    |
| <i>Clostridium perfringens</i>      | 13                           | 111  |
| <i>Clostridium perfringens</i>      | ATCC 13124                   | 66   |
| <i>Clostridium perfringens</i> S    |                              | 20   |
| <i>Clostridium perfringens</i>      | SM101                        | 65   |
| <i>Clostridium perfringens</i> HPT  | 13                           | 5    |

|  |                             |      |
|--|-----------------------------|------|
| <i>Clostridium perfringens</i> HPT     | ATCC 13124                  | 10   |
| <i>Clostridium perfringens</i> HPT     | SM101                       | 10   |
| <i>Clostridium perfringens</i> plasmid | plasmid pCP13               | 5    |
| <i>Clostridium tetani</i>              | E88                         | 55   |
| <i>Clostridium tetani</i> HPT          | tetani                      | 5    |
| <i>Clostridium thermocellum</i>        | ATCC 27405                  | 15   |
| <i>Corynebacterium diphtheriae</i>     | diphtheriae                 | 5    |
| <i>Corynebacterium diphtheriae</i>     | NCTC 13129                  | 165  |
| <i>Corynebacterium efficiens</i>       | YS-314                      | 65   |
| <i>Corynebacterium glutamicum</i>      | ATCC 13032                  | 20   |
| <i>Corynebacterium glutamicum</i>      | R                           | 69   |
| <i>Corynebacterium glutamicum</i> APRT | ATCC 13032                  | 5    |
| <i>Corynebacterium jeikeium</i>        | K411                        | 110  |
| <i>Coxiella burnetii</i>               | CbuG Q212                   | 15   |
| <i>Coxiella burnetii</i>               | Dugway 5J108-111            | 25   |
| <i>Coxiella burnetii</i>               | MSU Goat Q117               | 29   |
| <i>Coxiella burnetii</i>               | RSA 331                     | 15   |
| <i>Coxiella burnetii</i>               | RSA 334                     | 5    |
| <i>Coxiella burnetii</i>               | RSA 493                     | 178  |
| <i>Coxiella burnetii</i> HPT           | Dugway                      | 5    |
| <i>Coxiella burnetii</i> HPT           | burnetti                    | 10   |
| <i>Enterococcus faecalis</i>           | faecalis                    | 5    |
| <i>Enterococcus faecalis</i>           | MMH594                      | 5    |
| <i>Enterococcus faecalis</i>           | V583                        | 145  |
| <i>Enterococcus faecalis</i> APRT      | V583                        | 5    |
| <i>Enterococcus faecalis</i> HPT       | faecalis                    | 5    |
| <i>Escherichia coli</i>                | 536                         | 1035 |
| <i>Escherichia coli</i>                | 1226                        | 5    |
| <i>Escherichia coli</i>                | 1334                        | 5    |
| <i>Escherichia coli</i>                | 55989                       | 20   |
| <i>Escherichia coli</i>                | 042                         | 70   |
| <i>Escherichia coli</i>                | 17-2                        | 25   |
| <i>Escherichia coli</i>                | 536 (UPEC)                  | 30   |
| <i>Escherichia coli</i>                | B171                        | 85   |
| <i>Escherichia coli</i>                | C1845                       | 5    |
| <i>Escherichia coli</i>                | CFT 073 (UPEC)              | 516  |
| <i>Escherichia coli</i>                | coli                        | 182  |
| <i>Escherichia coli</i>                | coli/shigella               | 5    |
| <i>Escherichia coli</i>                | E. coli plasmid pC15-1a_016 | 5    |
| <i>Escherichia coli</i>                | E/99 3-2 SHV                | 10   |
| <i>Escherichia coli</i>                | E2348/69                    | 285  |
| <i>Escherichia coli</i>                | E45035                      | 5    |
| <i>Escherichia coli</i>                | EC7372                      | 5    |
| <i>Escherichia coli</i>                | EU2657                      | 5    |
| <i>Escherichia coli</i>                | EU4855 plasmid              | 5    |
| <i>Escherichia coli</i>                | H11128                      | 25   |

|                                       |                        |      |
|---------------------------------------|------------------------|------|
| <i>Escherichia coli</i>               | H11129                 | 5    |
| <i>Escherichia coli</i>               | K12                    | 38   |
| <i>Escherichia coli</i>               | K12 substr. MG1655     | 25   |
| <i>Escherichia coli</i>               | K983802                | 5    |
| <i>Escherichia coli</i>               | KS52                   | 5    |
| <i>Escherichia coli</i>               | O157:H7 EDL933         | 345  |
| <i>Escherichia coli</i>               | plasmid                | 15   |
| <i>Escherichia coli</i>               | plasmid p541           | 5    |
| <i>Escherichia coli</i>               | plasmid pEC365         | 5    |
| <i>Escherichia coli</i>               | plasmid pGR2439        | 5    |
| <i>Escherichia coli</i>               | plasmid pMEL2          | 3    |
| <i>Escherichia coli</i>               | plasmid RZA92          | 5    |
| <i>Escherichia coli</i>               | Sakai(EHEC O157:H7)    | 11   |
| <i>Escherichia coli</i>               | SMS-3-5                | 20   |
| <i>Escherichia coli</i>               | Str. 01 (APEC)         | 50   |
| <i>Escherichia coli</i>               | Toho-1                 | 5    |
| <i>Escherichia coli</i>               | UTI89 (UPEC)           | 65   |
| <i>Escherichia coli</i>               | YMC02/08/U310          | 5    |
| <i>Escherichia coli</i>               | SMS-3-5                | 5    |
| <i>Escherichia coli</i> APRT          | O157:H7 EDL933         | 5    |
| <i>Escherichia coli</i> HPT           | ATCC 8739              | 5    |
| <i>Escherichia coli</i> HPT           | E24377A                | 5    |
| <i>Escherichia coli</i> HPT           | F11                    | 4    |
| <i>Escherichia coli</i> HPT           | HS                     | 5    |
| <i>Escherichia coli</i> plasmid       | plasmid pAPEC-O1-ColBM | 40   |
| <i>Escherichia coli</i> strain EO 516 | EO 516                 | 5    |
| <i>Francisella holarctica</i> APRT    | OSU18                  | 5    |
| <i>Francisella holarctica</i>         | FTNF002-00             | 15   |
| <i>Francisella holarctica</i>         | holarctica             | 31   |
| <i>Francisella holarctica</i>         | LVS                    | 35   |
| <i>Francisella holarctica</i>         | OSU18                  | 25   |
| <i>Francisella holarctica</i> HPT     | holarctica             | 10   |
| <i>Francisella holarctica</i> SNP     | FSC022                 | 40   |
| <i>Francisella holarctica</i> SNP     | FTNF002-00             | 80   |
| <i>Francisella holarctica</i> SNP     | HOL 257                | 20   |
| <i>Francisella holarctica</i> SNP     | holarctica             | 240  |
| <i>Francisella holarctica</i> SNP     | LVS                    | 20   |
| <i>Francisella holarctica</i> SNP     | OSU18                  | 100  |
| <i>Francisella novicida</i>           | U112                   | 105  |
| <i>Francisella novicida</i> HPT       | U112                   | 5    |
| <i>Francisella novicida</i> SNP       | GA99-3548              | 700  |
| <i>Francisella novicida</i> SNP       | novicida               | 7480 |
| <i>Francisella novicida</i> SNP       | U112                   | 620  |
| <i>Francisella tularensis</i>         | ATCC 6223              | 46   |
| <i>Francisella tularensis</i>         | francisella            | 5    |
| <i>Francisella tularensis</i>         | fsc033                 | 5    |

|  |                              |     |
|--|------------------------------|-----|
| <i>Francisella tularensis</i>                | FSC198                       | 15  |
| <i>Francisella tularensis</i>                | plasmid pOM1                 | 5   |
| <i>Francisella tularensis</i>                | SCHU S4                      | 411 |
| <i>Francisella tularensis</i>                | tularensis                   | 52  |
| <i>Francisella tularensis</i>                | WY96-3418                    | 55  |
| <i>Francisella tularensis</i> SNP            | SCHU                         | 180 |
| <i>Francisella tularensis</i> SNP            | tularensis                   | 580 |
| <i>Francisella tularensis</i> SNP            | WY96                         | 100 |
| <i>Francisella tularensis</i> SNP            | WY96-3418                    | 20  |
| <i>Francisella</i>                           | Francisella                  | 15  |
| <i>Francisella holarctica/novicida</i>       | holarctica/novicida          | 5   |
| <i>Francisella holarctica/tularensis</i>     | holarctica/tularensis        | 25  |
| <i>Francisella novicida/tularensis</i>       | novicida/tularensis          | 30  |
| <i>Francisella tularensis/holarctica</i> SNP | tularensis/holarctica        | 20  |
| <i>Haemophilus ducreyi</i>                   | 35000 HP                     | 405 |
| <i>Haemophilus influenzae</i> APRT           | 86-028NP                     | 5   |
| <i>Haemophilus influenzae</i> APRT           | Rd KW20                      | 5   |
| <i>Haemophilus influenzae</i>                | 12                           | 30  |
| <i>Haemophilus influenzae</i>                | 1007                         | 89  |
| <i>Haemophilus influenzae</i>                | 3179B                        | 5   |
| <i>Haemophilus influenzae</i>                | 86-028NP                     | 336 |
| <i>Haemophilus influenzae</i>                | AM30                         | 25  |
| <i>Haemophilus influenzae</i>                | C54                          | 5   |
| <i>Haemophilus influenzae</i>                | influenzae                   | 5   |
| <i>Haemophilus influenzae</i>                | N187                         | 5   |
| <i>Haemophilus influenzae</i>                | Pitt EE                      | 275 |
| <i>Haemophilus influenzae</i>                | Pitt GG                      | 299 |
| <i>Haemophilus influenzae</i>                | Rd                           | 95  |
| <i>Haemophilus influenzae</i>                | Rd KW20                      | 375 |
| <i>Haemophilus somnus</i>                    | 2336                         | 205 |
| <i>Haemophilus somnus</i>                    | 129 PT                       | 380 |
| <i>Helicobacter acinonychis</i>              | Sheeba                       | 279 |
| <i>Helicobacter hepaticus</i>                | ATCC 51449                   | 250 |
| <i>Helicobacter pylori</i> APRT              | J99                          | 5   |
| <i>Helicobacter pylori</i>                   | 26695                        | 438 |
| <i>Helicobacter pylori</i>                   | HPAG1                        | 377 |
| <i>Helicobacter pylori</i>                   | J99                          | 484 |
| Human  | Human                        | 100 |
| <i>Klebsiella pneumonia</i> APRT             | MGH 78578                    | 5   |
| <i>Lactobacillus delbrueckii</i> APRT        | subsp. bulgaricus ATCC 11842 | 5   |
| <i>Legionella pneumophila</i>                | Philadelphia 1               | 793 |
| <i>Legionella pneumophila</i> HPT            | Corby                        | 3   |
| <i>Legionella pneumophila</i> HPT            | Lens                         | 5   |
| <i>Legionella pneumophila</i> HPT            | Paris                        | 10  |
| <i>Legionella pneumophila</i> HPT            | Philadelphia 1               | 5   |
| <i>Legionella pneumophila</i>                | Corby                        | 296 |

|  |                          |      |
|--|--------------------------|------|
| <i>Legionella pneumophila</i>          | Lens                     | 378  |
| <i>Legionella pneumophila</i>          | Paris                    | 399  |
| <i>Legionella pneumophila</i>          | pneumophila              | 5    |
| <i>Listeria innocua</i>                | Clip 11262               | 105  |
| <i>Listeria ivanovi</i>                | ATCC 19119               | 5    |
| <i>Listeria monocytogenes</i>          | monocytogenes            | 10   |
| <i>Listeria monocytogenes</i> APRT     | EGD-e                    | 5    |
| <i>Listeria monocytogenes</i> HPT      | 4b 2365                  | 10   |
| <i>Listeria monocytogenes</i> HPT      | EGD-e                    | 5    |
| <i>Listeria monocytogenes</i>          | 4b 2365                  | 260  |
| <i>Listeria monocytogenes</i>          | EGD-e sv 1/2A            | 453  |
| <i>Listeria monocytogenes</i>          | F2365                    | 95   |
| <i>Listeria monocytogenes</i> APRT     | F2365                    | 5    |
| <i>Listeria monocytogenes</i> SNP      | J1-194                   | 1280 |
| <i>Listeria monocytogenes</i> SNP      | J2-064                   | 80   |
| <i>Listeria monocytogenes</i>          | J2-064                   | 5    |
| <i>Listeria monocytogenes</i> SNP      | monocytogenes            | 5180 |
| <i>Listeria welshimeri</i> APRT        | SLCC 5334                | 5    |
| <i>Listeria welshimeri</i>             | SLCC 5334                | 100  |
| <i>Mycobacterium avium</i> APRT        | K-10 ss paratuberculosis | 5    |
| <i>Mycobacterium avium</i>             | 104                      | 263  |
| <i>Mycobacterium avium</i>             | K-10 ss paratuberculosis | 743  |
| <i>Mycobacterium bovis</i> APRT        | BCG str. Pasteur 1173P2  | 5    |
| <i>Mycobacterium bovis</i>             | AF2122/97                | 15   |
| <i>Mycobacterium bovis</i>             | BCG Pasteur 1173P2       | 15   |
| <i>Mycobacterium gilvums</i>           | PYR-GCK                  | 619  |
| <i>Mycobacterium leprae</i> APRT       | TN                       | 5    |
| <i>Mycobacterium leprae</i>            | TN                       | 379  |
| <i>Mycobacterium marinum</i> APRT      | M                        | 5    |
| <i>Mycobacterium smegmatis</i>         | MC2155                   | 543  |
| <i>Mycobacterium tuberculosis</i> APRT | CDC 1551                 | 5    |
| <i>Mycobacterium tuberculosis</i>      | CDC 1551                 | 120  |
| <i>Mycobacterium tuberculosis</i>      | F11                      | 15   |
| <i>Mycobacterium tuberculosis</i>      | H37 Ra                   | 5    |
| <i>Mycobacterium tuberculosis</i>      | H37 Rv                   | 682  |
| <i>Mycobacterium tuberculosis</i>      | tuberculosis/bovis       | 5    |
| <i>Mycobacterium ulcerans</i>          | Agy 99                   | 504  |
| <i>Mycobacterium ulcerans</i> APRT     | Agy99                    | 5    |
| <i>Mycobacterium ulcerans</i> plasmid  | Agy99 plasmid pMUM001    | 20   |
| <i>Mycobacterium van baalenii</i>      | PYR-1                    | 702  |
| <i>Mycobacterium</i> sp.               | JLS                      | 650  |
| <i>Mycobacterium</i> sp.               | KMS                      | 120  |
| <i>Mycobacterium</i> sp.               | MCS                      | 45   |
| <i>Mycoplasma agalactiae</i>           | PG2                      | 45   |
| <i>Mycoplasma capricolum</i>           | ATCC 27343               | 10   |
| <i>Mycoplasma gallisepticum</i>        | R                        | 230  |

|                                      |                  |      |
|--------------------------------------|------------------|------|
| <i>Mycoplasma genitalium</i>         | G37              | 50   |
| <i>Mycoplasma hyopneumoniae</i> APRT | 7448             | 4    |
| <i>Mycoplasma hyopneumoniae</i> APRT | J                | 7    |
| <i>Mycoplasma hyopneumoniae</i>      | 232              | 70   |
| <i>Mycoplasma hyopneumoniae</i>      | 7448             | 35   |
| <i>Mycoplasma hyopneumoniae</i>      | J                | 30   |
| <i>Mycoplasma mobile</i>             | 163K             | 105  |
| <i>Mycoplasma mycoides</i> APRT      | PG1              | 5    |
| <i>Mycoplasma mycoides</i>           | PG1              | 90   |
| <i>Mycoplasma penetrans</i>          | HF-2             | 250  |
| <i>Mycoplasma pneumoniae</i> APRT    | M129             | 5    |
| <i>Mycoplasma pneumoniae</i>         | M129             | 50   |
| <i>Mycoplasma pneumoniae</i>         | pneumoniae       | 5    |
| <i>Mycoplasma pulmonis</i> APRT      | UAB CTIP         | 5    |
| <i>Mycoplasma pulmonis</i>           | UABCTIP          | 74   |
| <i>Mycoplasma synoviae</i>           | 53               | 10   |
| <i>Neisseria gonorrhoeae</i>         | FA 1090          | 205  |
| <i>Neisseria meningitidis</i>        | FAM18            | 188  |
| <i>Neisseria meningitidis</i>        | MC58             | 274  |
| <i>Neisseria meningitidis</i>        | neisseria        | 5    |
| <i>Neisseria meningitidis</i>        | str. 053442      | 164  |
| <i>Neisseria meningitidis</i>        | Z2491            | 281  |
| Plasmid pBC16                        | Plasmid pBC16    | 5    |
| Plasmid pLS1                         | Plasmid pLS1     | 5    |
| <i>Pseudomonas aeruginosa</i> HPT    | 2192 Paer2_01_70 | 5    |
| <i>Pseudomonas aeruginosa</i> HPT    | PA01             | 10   |
| <i>Pseudomonas aeruginosa</i> HPT    | PA7              | 5    |
| <i>Pseudomonas aeruginosa</i>        | aeruginosa       | 5    |
| <i>Pseudomonas aeruginosa</i>        | PA01             | 1274 |
| <i>Pseudomonas aeruginosa</i>        | PA7              | 1015 |
| <i>Pseudomonas aeruginosa</i>        | UCBPP-PA14       | 317  |
| <i>Pseudomonas entomophila</i> HPT   | L48              | 5    |
| <i>Pseudomonas entomophila</i>       | L48              | 558  |
| <i>Pseudomonas fluorescens</i> HPT   | Pf-5             | 5    |
| <i>Pseudomonas fluorescens</i> HPT   | PfO-1            | 5    |
| <i>Pseudomonas fluorescens</i>       | Pf-5             | 710  |
| <i>Pseudomonas fluorescens</i>       | PfO-1            | 590  |
| <i>Pseudomonas mendocina</i> HPT     | ymp              | 5    |
| <i>Pseudomonas mendocina</i>         | ymp              | 645  |
| <i>Pseudomonas putida</i> APRT       | KT 2440          | 5    |
| <i>Pseudomonas putida</i> HPT        | GB-1             | 5    |
| <i>Pseudomonas putida</i> HPT        | KT 2440          | 5    |
| <i>Pseudomonas putida</i>            | F1               | 430  |
| <i>Pseudomonas putida</i>            | GB-1             | 607  |
| <i>Pseudomonas putida</i>            | KT 2440          | 706  |
| <i>Pseudomonas putida</i>            | W619             | 560  |

|                                       |                             |      |
|---------------------------------------|-----------------------------|------|
| <i>Pseudomonas stutzeri</i>           | A1501                       | 480  |
| <i>Pseudomonas stutzeri</i> HPT       | A1501                       | 5    |
| <i>Pseudomonas syringae</i> APRT      | pv. phaseolicola 1448A      | 5    |
| <i>Pseudomonas syringae</i>           | 1448A                       | 1042 |
| <i>Pseudomonas syringae</i>           | B728a                       | 1021 |
| <i>Pseudomonas syringae</i>           | DC3000                      | 1214 |
| <i>Pseudomonas syringae</i> HPT       | pv. phaseolicola 1448A      | 7    |
| <i>Pseudomonas syringae</i> HPT       | pv. syringae B728a          | 5    |
| <i>Pseudomonas syringae</i> HPT       | pv. tomato str. DC3000      | 5    |
| <i>Pseudomonas syringae</i> plasmid   | 1448A large plasmid         | 50   |
| <i>Pseudomonas syringae</i> plasmid   | plasmid pDC3000A            | 20   |
| <i>Ricinus communis</i>               | communis                    | 20   |
| <i>Rickettsia prowazekii</i>          | Madrid E                    | 55   |
| <i>Rickettsia prowazekii</i>          | prowazekii                  | 5    |
| <i>Rickettsia rickettsii</i>          | Iowa                        | 70   |
| <i>Rickettsia rickettsii</i> SNP      | rickettsiae                 | 60   |
| <i>Rickettsia rickettsii</i>          | rickettsii/africae/sibirica | 5    |
| <i>Rickettsia typhi</i>               | Wilmington                  | 55   |
| <i>Salmonella enterica</i> APRT       | Typhi str. CT18             | 5    |
| <i>Salmonella enterica</i>            | ATCC 9150 sv paratyphi A    | 168  |
| <i>Salmonella enterica</i>            | CT18                        | 332  |
| <i>Salmonella enterica</i>            | enterica                    | 5    |
| <i>Salmonella enterica</i>            | LT2                         | 520  |
| <i>Salmonella enterica</i>            | RSK 2980 ss arizona sv 62   | 544  |
| <i>Salmonella enterica</i>            | SC-B67 sv Choleraesuis      | 201  |
| <i>Salmonella enterica</i>            | SPB7 sv Paratyphi B         | 207  |
| <i>Salmonella enterica</i>            | sv typhimurium              | 239  |
| <i>Salmonella enterica</i>            | Ty2                         | 10   |
| <i>Salmonella enterica</i> plasmid    | pSN254                      | 125  |
| <i>Salmonella enterica</i> plasmid    | SC-B67 plasmid pSCV50       | 10   |
| <i>Salmonella typhimurium</i>         | LT2                         | 253  |
| <i>Salmonella typhimurium</i> plasmid | LT2 plasmid pSLT            | 5    |
| <i>Shigella dysenteriae</i>           | plasmid pmK105              | 5    |
| <i>Shigella boydii</i>                | 227                         | 43   |
| <i>Shigella boydii</i>                | 0-1392                      | 20   |
| <i>Shigella boydii</i>                | CDC 3083-94                 | 93   |
| <i>Shigella boydii</i>                | Sb227                       | 300  |
| <i>Shigella boydii</i> HPT            | boydii                      | 5    |
| <i>Shigella boydii</i> plasmid        | plasmid pSB4_227            | 15   |
| <i>Shigella dysenteriae</i> APRT      | Sd197                       | 5    |
| <i>Shigella dysenteriae</i>           | 197                         | 107  |
| <i>Shigella dysenteriae</i>           | Sd197                       | 130  |
| <i>Shigella dysenteriae</i> plasmid   | plasmid pSD1_197            | 171  |
| <i>Shigella flexneri</i>              | 301                         | 866  |
| <i>Shigella flexneri</i>              | 8401                        | 60   |
| <i>Shigella flexneri</i>              | 2457T                       | 80   |

|   |                          |     |
|---|--------------------------|-----|
| <i>Shigella flexneri</i>                | flexneri                 | 45  |
| <i>Shigella flexneri</i>                | M90T                     | 248 |
| <i>Shigella flexneri</i>                | multiple species         | 5   |
| <i>shigella flexneri HPT</i>            | flexneri                 | 5   |
| <i>Shigella flexneri plasmid</i>        | M90T plasmid pWR501      | 15  |
| <i>Shigella flexneri plasmid</i>        | plasmid pPCP301          | 35  |
| <i>Shigella sonnei</i>                  | Ss046                    | 66  |
| <i>Shigella sonnei plasmid</i>          | str. 046 plasmid pSS_046 | 15  |
| <i>Staphylococcus aureus APRT</i>       | N315                     | 5   |
| <i>Staphylococcus aureus HPT</i>        | aureus                   | 5   |
| <i>Staphylococcus aureus</i>            | aureus                   | 45  |
| <i>Staphylococcus aureus</i>            | COL                      | 140 |
| <i>Staphylococcus aureus</i>            | JH1                      | 15  |
| <i>Staphylococcus aureus</i>            | JH9                      | 15  |
| <i>Staphylococcus aureus</i>            | MRSA 252                 | 255 |
| <i>Staphylococcus aureus</i>            | MSSA 476                 | 2   |
| <i>Staphylococcus aureus</i>            | Mu3                      | 10  |
| <i>Staphylococcus aureus</i>            | Mu50                     | 140 |
| <i>Staphylococcus aureus</i>            | MW2                      | 350 |
| <i>Staphylococcus aureus</i>            | N315                     | 20  |
| <i>Staphylococcus aureus</i>            | NCTC 8325                | 29  |
| <i>Staphylococcus aureus</i>            | Newman                   | 15  |
| <i>Staphylococcus aureus</i>            | RF122                    | 203 |
| <i>Staphylococcus aureus</i>            | USA 300_TCH 1516         | 10  |
| <i>Staphylococcus aureus</i>            | USA 3000                 | 27  |
| <i>Staphylococcus epidermidis APRT</i>  | RP62A                    | 5   |
| <i>Staphylococcus epidermidis</i>       | ATCC 12228               | 62  |
| <i>Staphylococcus epidermidis</i>       | RP62A                    | 60  |
| <i>Staphylococcus epidermidis HPT</i>   | epideridis               | 5   |
| <i>Staphylococcus haemolyticus</i>      | JCSC 1435                | 80  |
| <i>Staphylococcus haemolyticus HPT</i>  | haemolyticus             | 5   |
| <i>Staphylococcus saprophyticus HPT</i> | saprophyticus            | 5   |
| <i>Staphylococcus saprophyticus</i>     | ATCC 15305               | 95  |
| <i>Streptococcus agalactiae APRT</i>    | A909                     | 5   |
| <i>Streptococcus agalactiae</i>         | 2603 V/R                 | 145 |
| <i>Streptococcus agalactiae</i>         | A909                     | 200 |
| <i>Streptococcus agalactiae</i>         | agalactiae               | 5   |
| <i>Streptococcus agalactiae</i>         | FM027022                 | 5   |
| <i>Streptococcus agalactiae</i>         | NEM316                   | 75  |
| <i>Streptococcus agalactiae HPT</i>     | agalactiae               | 5   |
| <i>Streptococcus agalactiae HPT</i>     | CJB111                   | 10  |
| <i>Streptococcus gordonii</i>           | Challis                  | 150 |
| <i>Streptococcus mutans</i>             | UA 159                   | 135 |
| <i>Streptococcus pneumoniae APRT</i>    | Hungary 19A-6            | 5   |
| <i>Streptococcus pneumoniae APRT</i>    | R6                       | 5   |
| <i>Streptococcus pneumoniae HPT</i>     | Hungary 19A-6            | 6   |

|                                       |                                  |     |
|---------------------------------------|----------------------------------|-----|
| <i>Streptococcus pneumoniae</i> HPT   | pneumoniae                       | 5   |
| <i>Streptococcus pneumoniae</i> HPT   | TIGR4                            | 2   |
| <i>Streptococcus pneumoniae</i>       | CGSP14                           | 87  |
| <i>Streptococcus pneumoniae</i>       | D39                              | 154 |
| <i>Streptococcus pneumoniae</i>       | Hungary 19A-6                    | 130 |
| <i>Streptococcus pneumoniae</i>       | pneumoniae                       | 5   |
| <i>Streptococcus pneumoniae</i>       | R6                               | 5   |
| <i>Streptococcus pneumoniae</i>       | TIGR4                            | 185 |
| <i>Streptococcus pyogenes</i> APRT    | M1 GAS                           | 5   |
| <i>Streptococcus pyogenes</i>         | Manfredo st M5                   | 50  |
| <i>Streptococcus pyogenes</i>         | MGAS 10270 st M2                 | 95  |
| <i>Streptococcus pyogenes</i>         | MGAS 10394 st M6                 | 95  |
| <i>Streptococcus pyogenes</i>         | MGAS 10750 st M4                 | 83  |
| <i>Streptococcus pyogenes</i>         | MGAS 2096 st M12                 | 57  |
| <i>Streptococcus pyogenes</i>         | MGAS 315 st M3                   | 85  |
| <i>Streptococcus pyogenes</i>         | MGAS 5005 st M1                  | 35  |
| <i>Streptococcus pyogenes</i>         | MGAS 6180 st M28                 | 80  |
| <i>Streptococcus pyogenes</i>         | MGAS 8232 st M18                 | 65  |
| <i>Streptococcus pyogenes</i>         | MGAS 9429 st M12                 | 10  |
| <i>Streptococcus pyogenes</i>         | pyogenes                         | 5   |
| <i>Streptococcus pyogenes</i>         | SF370                            | 150 |
| <i>Streptococcus pyogenes</i>         | SSI-1 st M3                      | 36  |
| <i>Streptococcus pyogenes</i> HPT     | MGAS 10394                       | 5   |
| <i>Streptococcus pyogenes</i> HPT     | MGAS 10750                       | 5   |
| <i>Streptococcus pyogenes</i> HPT     | MGAS 8232                        | 5   |
| <i>Streptococcus pyogenes</i> HPT     | pyogenes                         | 5   |
| <i>Streptococcus sanguinis</i>        | SK36                             | 232 |
| <i>Streptococcus sanguinis</i> HPT    | sanguinis                        | 5   |
| <i>Streptococcus suis</i>             | 05ZYH33                          | 138 |
| <i>Streptococcus suis</i>             | 98 HAH33                         | 65  |
| <i>Streptococcus thermophilus</i> HPT | LMG 18311                        | 4   |
| <i>Streptococcus thermophilus</i>     | CNRZ1066                         | 117 |
| <i>Streptococcus thermophilus</i>     | LMD-9                            | 150 |
| <i>Streptococcus thermophilus</i>     | LMG 18311                        | 135 |
| <i>Streptococcus thermophilus</i> HPT | thermophilus                     | 5   |
| <i>Treponema pallidum</i>             | Nichols                          | 5   |
| <i>Treponema pallidum</i>             | pallidum                         | 5   |
| <i>Treponema pallidum</i>             | SS14                             | 30  |
| <i>Ureaplasma parvum</i> APRT         | ATCC 27815                       | 5   |
| <i>Vibrio cholerae</i> APRT           | O395                             | 5   |
| <i>Vibrio cholerae</i> HPT            | 623-39                           | 5   |
| <i>Vibrio cholerae</i> HPT            | RC385                            | 5   |
| <i>Vibrio cholerae</i>                | 1587                             | 5   |
| <i>Vibrio cholerae</i>                | 623-39                           | 10  |
| <i>Vibrio cholerae</i>                | all other <i>Vibrio cholerae</i> | 60  |
| <i>Vibrio cholerae</i>                | cholerae                         | 5   |

|  |                                |      |
|--|--------------------------------|------|
| <i>Vibrio cholerae</i>                 | MAK 757                        | 5    |
| <i>Vibrio cholerae</i>                 | MZO-2                          | 5    |
| <i>Vibrio cholerae</i>                 | MZO-3                          | 5    |
| <i>Vibrio cholerae</i>                 | N16961                         | 1144 |
| <i>Vibrio cholerae</i>                 | NCTC 8457                      | 5    |
| <i>vibrio cholerae</i>                 | O395                           | 145  |
| <i>Vibrio cholerae</i>                 | plasmid pTLC -1                | 5    |
| <i>Vibrio cholerae</i>                 | plasmid pTLC -2                | 5    |
| <i>Vibrio cholerae</i>                 | plasmid pTLC -3                | 5    |
| <i>Vibrio cholerae</i>                 | plasmid pTLC -4                | 5    |
| <i>Vibrio cholerae</i>                 | plasmid pTLC -5                | 5    |
| <i>Vibrio cholerae</i>                 | RC385                          | 5    |
| <i>Vibrio cholerae</i>                 | V51                            | 10   |
| <i>Vibrio cholerae</i> HPT             | cholerae                       | 5    |
| <i>Vibrio cholerae</i> HPT             | V51                            | 5    |
| <i>Vibrio fischeri</i>                 | ES114                          | 554  |
| <i>Vibrio parahaemolyticus</i>         | AQ3810                         | 5    |
| <i>Vibrio parahaemolyticus</i>         | parahaemolyticus               | 5    |
| <i>Vibrio parahaemolyticus</i>         | RIMD 2210633                   | 830  |
| <i>Vibrio vulnificus</i>               | CMCP6                          | 764  |
| <i>Vibrio vulnificus</i>               | Vibrio vulnificus              | 5    |
| <i>Vibrio vulnificus</i>               | YJ016                          | 443  |
| <i>Xanthomonas axonopodis</i> APRT     | pv. citri str. 306             | 5    |
| <i>Yersinia enterocolitica</i>         | 8081                           | 560  |
| <i>Yersinia enterocolitica</i>         | 84-50                          | 5    |
| <i>Yersinia enterocolitica</i>         | A127                           | 177  |
| <i>Yersinia enterocolitica</i>         | W1024                          | 10   |
| <i>Yersinia enterocolitica</i> APRT    | 8081                           | 5    |
| <i>Yersinia enterocolitica</i> HPT     | 8081                           | 10   |
| <i>Yersinia enterocolitica</i> plasmid | 8081 plasmid pYVe8081          | 94   |
| <i>Yersinia pestis</i>                 | 91001 bv Microtus              | 20   |
| <i>Yersinia pestis</i>                 | Angola                         | 38   |
| <i>Yersinia pestis</i>                 | Antiqua                        | 50   |
| <i>Yersinia pestis</i>                 | bv Microtus str. 91001         | 15   |
| <i>Yersinia pestis</i>                 | CO92                           | 614  |
| <i>Yersinia pestis</i>                 | KIM                            | 65   |
| <i>Yersinia pestis</i>                 | Nepal 516                      | 20   |
| <i>Yersinia pestis</i>                 | Pestoides F                    | 15   |
| <i>Yersinia pestis</i>                 | Y. pestis                      | 5    |
| <i>Yersinia pestis</i> APRT            | Angola                         | 5    |
| <i>Yersinia pestis</i> APRT            | CO92                           | 5    |
| <i>Yersinia pestis</i> APRT            | KIM                            | 5    |
| <i>Yersinia pestis</i> HPT             | CO92                           | 10   |
| <i>Yersinia pestis</i> plasmid         | pIP1202                        | 90   |
| <i>Yersinia pestis</i> plasmid         | 91001 bv Microtus plasmid pCD1 | 10   |
| <i>Yersinia pestis</i> plasmid         | Angola plasmid pCD             | 5    |

|   |                           |     |
|---|---------------------------|-----|
| <i>Yersinia pestis</i> plasmid                | Pestoides F plasmid pCD   | 13  |
| <i>Yersinia pseudotuberculosis</i>            | IP 31758                  | 115 |
| <i>Yersinia pseudotuberculosis</i>            | IP 32953                  | 68  |
| <i>Yersinia pseudotuberculosis</i>            | pseudotuberculosis        | 5   |
| <i>Yersinia pseudotuberculosis</i>            | YP111                     | 56  |
| <i>Yersinia pseudotuberculosis</i> HPT        | PB1/+                     | 10  |
| <i>Yersinia pseudotuberculosis</i> plasmid    | IP32953 plasmid YV        | 12  |
| <i>Yersinia pseudotuberculosis</i> plasmid    | plasmid pYps IP31758.1    | 195 |
| <i>Yersinia pseudotuberculosis</i> plasmid    | plasmid pYps IP31758.2    | 45  |
| <i>Yersinia pestis/pseudotuberculosis</i>     | pestis/pseudotuberculosis | 10  |
| <i>Yersinia pestis/pseudotuberculosis</i> SNP | IP 31758                  | 20  |
| <i>Yersinia pestis/pseudotuberculosis</i> SNP | pestis/pseudotuberculosis | 520 |
|   |                           |     |
|   |                           |     |

## **Annex B Detailed Protocols for microarray labeling**

---

### **dUTP Incorporation - Using ROCHE Random Priming kit (no amplification)**

1. Add 5 µg sample DNA (16 µl volume) to PCR tube. Incubate at 95 °C for 10 min.
2. During the above incubation, prepare the following (Incorporation mix):

0.8 µl dH<sub>2</sub>O  
0.8 µl of 1 mM dUTP  
1.6 µl of 0.5 mM dTTP  
2.0 µl of 0.5 mM dATP  
2.0 µl of 0.5 mM dCTP  
2.0 µl of 0.5 mM dGTP

3. Vortex the mixture briefly.
4. When the incubation in step 1 is almost finished, add the following to the incorporation mix:
  - 2.0 µl of the hexamer primer reaction mix
  - 1.0 µl of Klenow enzyme (keep on ice until use)
5. Vortex the mixture briefly and keep on ice until use.
6. Add 12.2 µl (total volume of the incorporation mixture) to the cooled sample DNA from step 1. Vortex briefly.
7. Incubate the sample at 37 °C for 2 hrs (program AFFY1).
8. After 2 hrs, take sample from incubator and add sufficient RNase-free H<sub>2</sub>O to make a final volume of 28.2 µl (if required).
9. Incubate sample tube at 95 °C for 10 min

### **Fragmentation - Using AFFY GeneChip WT Terminal Labelling Kit**

10. When step 9 is almost finished, prepare the following (Fragmentation mix):

10.0 µl RNase-free H<sub>2</sub>O  
4.8 µl of Fragmentation buffer  
1.0 µl of 10 U/µl UDG enzyme  
1.0 µl of 1000 U/µl APE1 enzyme  
Total volume 16.8 µl

11. Vortex briefly and keep on ice until needed.
12. Add 16.8 µl of Fragmentation mix to the cooled sample DNA. Total volume should be 45 µl.
13. Incubate mixture at 37 °C for 1 hour (program AFFY2)

### **Labeling - Using the AFFY GeneChip WT Terminal Labelling Kit**

14. When step 13 is almost finished, prepare the following (Labeling Mix):

10.0  $\mu$ l 5x TdT buffer  
2.0  $\mu$ l TdT  
1.0  $\mu$ l DNA labelling reagent  
Total volume 15.0  $\mu$ l

15. Vortex briefly and keep on ice until needed.

16. Add the total volume of labelling mix to the cooled sample mixture from step 13.

17. Incubate at 37 °C for 1 hour (program AFFY3)

### **Microarray hybridization, post-treatment, scanning, feature extraction**

After preparing the biotinylated sample probe, a hybridization mixture is prepared as follows:

60.0  $\mu$ l sample mixture  
11.0  $\mu$ l warmed (65 °C) 20x Eukaryote Hybridization Control  
3.7  $\mu$ l B2 Oligo control  
15.4  $\mu$ l DMSO  
110.0  $\mu$ l 2X Hybridization Buffer Mix  
20.0  $\mu$ l dH<sub>2</sub>O

The entire labeled sample reaction is denatured at 99 °C, then cooled to 45 °C for 5 minutes. 200  $\mu$ l is applied to the microarray, then the array is incubated in a hybridization oven in a rotating holder for 16–18 hours at 45 °C. After hybridization, the array is transferred to the fluidics station which performs a post-hybridization wash followed by an automated labeling with streptavidin-phycoerythrin, a fluorescent chromophore complex that binds to the biotin in the sample probe.

Post-hybridization arrays were scanned in the 3000-7G Affymetrix array imager. This unit scans the barcode of the array, then applies the appropriate colour and resolution settings automatically, scans the array, and downloads the recorded data to the workstation for analysis. Feature extraction takes place automatically following the array scan download.

## List of symbols/abbreviations/acronyms/initialisms

---

|       |   |
|-------|---|
| APE   | Apurinic endonuclease; cleaves DNA adjacent to apurinic sites   |
| APRT  | adenine phosphoribosyltransferase   |
| ATCC  | American Type Culture Collection; an organization supplying standard microbial strains and samples                                |
| BLAST | Basic Local Alignment Search Tool   |
| bp    | base pair   |
| DNA   | deoxyribonucleic acid   |
| DRDC  | Defence Research & Development Canada   |
| HPT   | hypoxanthine guanine phosphoribosyltransferase  |
| mM    | millimolar  |
| NCBI  | National Center for Biotechnology Information (also referred to as Genbank)   |
| PERL  | a high level programming language for scanning text files, extracting data, and generating reports from the data                  |
| R&D   | Research & Development  |
| SNP   | single nucleotide polymorphism; a sequence variant at one base position which may be different between populations or individuals |
| TDT   | terminal deoxynucleotidyl transferase   |

This page intentionally left blank.

### DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

|  |  |  |   |
|--|--|--|---|
| 1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)<br><br>Defence R&D Canada – Suffield<br>P.O. Box 4000, Station Main<br>Medicine Hat, Alberta T1A 8K6 |  | 2. SECURITY CLASSIFICATION<br>(Overall security classification of the document including special warning terms if applicable.)<br><br>UNCLASSIFIED |   |
| 3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)<br><br>An Affymetrix Microarray Design for Microbial Genotyping  |  |  |   |
| 4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)<br><br>Ford, B.N.; Bader, D.; Shei, Y.; Ruttan, C.; Mah, D.  |  |  |   |
| 5. DATE OF PUBLICATION<br>(Month and year of publication of document.)<br><br>October 2009   |  | 6a. NO. OF PAGES<br>(Total containing information, including Annexes, Appendices, etc.)<br><br>48  | 6b. NO. OF REFS<br>(Total cited in document.)<br><br>17 |
| 7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)<br><br>Technical Memorandum                            |  |  |   |
| 8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)<br><br>Defence R&D Canada – Suffield<br>P.O. Box 4000, Station Main<br>Medicine Hat, Alberta T1A 8K6   |  |  |   |
| 9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)  |  | 9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)   |   |
| 10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)<br><br>DRDC Suffield TM 2009-183  |  | 10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)                     |   |
| 11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)<br><br>Unlimited  |  |  |   |
| 12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.)<br><br>Unlimited            |  |  |   |

13. **ABSTRACT** (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

There is an ongoing requirement for development of high-density or multiplex assays for detection or identification of microbes. There is also a need to develop assays or toolsets that can detect or identify microbial threats without prior knowledge of the target microbe(s) in a given sample. Indeed, some samples that contain no culturable material (e.g. viable but non-culturable cells) will nonetheless contain detectable DNA fragments which might be of value with respect to forensics or attribution of source. For many pathogenic microbes, various specific tests already exist, but there are few general methods wherein a single adaptable tool can be applied to multiple species or to previously uncharacterized organisms. The high-density DNA microarray has the potential to address many of these requirements and thus complements existing identification tools. The microarray platform has for example, the ability to detect microbial DNA that is not a perfect match to known genomic DNA sequences, thus making it possible to detect microbial variants that might otherwise be missed. In this report, the design and preliminary testing of a high-density DNA microarray for the purpose of microbial identification and detection is described.

14. **KEYWORDS, DESCRIPTORS or IDENTIFIERS** (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

microbial genotyping; genomic fingerprinting; detection; identification; microarray



## **Defence R&D Canada**

Canada's Leader in Defence  
and National Security  
Science and Technology

## **R & D pour la défense Canada**

Chef de file au Canada en matière  
de science et de technologie pour  
la défense et la sécurité nationale



[www.drdc-rddc.gc.ca](http://www.drdc-rddc.gc.ca)