

Award Number: W81XWH-05-2-0040

TITLE: Tools for Ultraspecific Probe/Primer Design

PRINCIPAL INVESTIGATOR: Dr. Yuriy Fofanov

CONTRACTING ORGANIZATION: University of Texas at Houston
Houston, Texas 77030

REPORT DATE: April 2006

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE (DD-MM-YYYY) 01-05-2006			2. REPORT TYPE Final		3. DATES COVERED (From - To) 15 Apr 2005 – 14 Apr 2006	
4. TITLE AND SUBTITLE Tools for Ultraspecific Probe/Primer Design					5a. CONTRACT NUMBER	
					5b. GRANT NUMBER W81XWH-05-2-0040	
					5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Dr. Yuriy Fofanov E-Mail:					5d. PROJECT NUMBER	
					5e. TASK NUMBER	
					5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Texas at Houston Houston, Texas 77030					8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012						
10. SPONSOR/MONITOR'S ACRONYM(S)					11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
13. SUPPLEMENTARY NOTES						
14. ABSTRACT We offer a novel methodology for rapidly identifying superior-performance DNA probes/primers for use in detecting emerging or engineered pathogens. Our approach will deliver DNA probes and PCR primers that have an unprecedentedly low probability of false positives or confusion by environmental background, and which resist evasion by threat agent engineering. Any detection method that utilizes DNA or RNA probes or primers will benefit greatly by using probes/primers designed with our methodologies. This technology is made possible by novel insights into statistical properties of useful probes, primer pairs, and targets. Such findings have become possible because of dramatic advances in the computational analysis of genomic sequence data. Using our novel approach, background sequences are rigorously (not heuristically, e.g., BLAST) discriminated against. Thus, probes and primers developed using these tools can be known to be at least three mismatches away from the nearest other sequence in an entire set of DNA sequences employed in the calculations. The Phase I studies will demonstrate the advantages of our design technology. In this phase we will (1) perform extensive analysis of several Category A and B pathogens and produce(deliver) the database of all human and/or "background" 1, 2, 3, and 4 mismatches blind 16-22-mers present in their genomes; (2) transform in house scientific software into a Windows-based application that allows users to perform similar calculations for any custom sequence for 16-19-mers with up to 3 mismatches blind; and (3) perform intensive experimental validation in order to verify candidate sequences and experimentally estimate false discovery rate.						
15. SUBJECT TERMS No subject terms provided						
16. SECURITY CLASSIFICATION OF:				17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	19b. TELEPHONE NUMBER (include area code)			

Table of Contents

Introduction.....4

Body.....4

Key Research Accomplishments.....8

Reportable Outcomes.....10

Conclusions.....12

Introduction

We offer a novel methodology for rapidly identifying superior-performance DNA probes/primers for use in detecting emerging or engineered pathogens. Our approach will deliver DNA probes and PCR primers that have an unprecedentedly low probability of false positives or confusion by environmental background, and which resist evasion by threat agent engineering. Any detection method that utilizes DNA or RNA probes or primers will benefit greatly by using probes/primers designed with our methodologies. This technology is made possible by novel insights into statistical properties of useful probes, primer pairs, and targets. Such findings have become possible because of dramatic advances in the computational analysis of genomic sequence data. Using our novel approach, background sequences are rigorously (not heuristically, e.g., BLAST) discriminated against. Thus, probes and primers developed using these tools can be known to be at least three mismatches away from the nearest other sequence in an entire set of DNA sequences employed in the calculations. The Phase I studies will demonstrate the advantages of our design technology. In this phase we will (1) perform extensive analysis of several Category A and B pathogens and produce(deliver) the database of all human and/or "background" 1, 2, 3, and 4 mismatches blind 16-22-mers present in their genomes; (2) transform in house scientific software into a Windows-based application that allows users to perform similar calculations for any custom sequence for 16-19-mers with up to 3 mismatches blind; and (3) perform intensive experimental validation in order to verify candidate sequences and experimentally estimate false discovery rate.

Body

During Phase I of the project, we assembled our entire team and began work on all four specific tasks. Specifically:

Task 1: Calculations of host blind 16-22-mers

1.1 Data collection and preparation: We acquired and updated all of the publicly available genome data for the calculations.

1.2 Existing software modification and testing: We completed the working version of the software which was used for all computations (subtask 1.3 and 1.5).

1.3 Calculation of 16-19-mers: All human as well as three synthetic backgrounds (genomes of microorganisms which have a high probability of being present in the air, drinking water, and human associated) 1-3 mismatches blind signatures of size 16 through 19 (16-19-mers) were calculated for the following organisms:

Category A Microbial and Viral Genomes:

- ***Bacillus anthracis*:** 3 complete genomes, *B.s anthracis* str. Ames (NC_003997), *B. anthracis* str. Sterne (NC_005945), and *B. anthracis* str. 'Ames Ancestor' (NC_007530), and 2 plasmid sequences.
- ***Yersinia pestis*:** 3 genomes, *Y. pestis* CO92 (NC_003143), *Y. pestis* KIM (NC_004088), and *Y. pestis* biovar *Mediaevalis* str. 91001 (NC_005810), and 8 plasmid sequences.
- **Dengue Virus:** 111 available complete genomes.

Category B Microbial and Viral Genomes:

- ***Shigella flexneri*:** 2 complete genomes, *S. flexneri* 2a str. 2457T (NC_004741) and *S. flexneri* 2a str. 301 (NC_004337), and 1 plasmid sequence.
- ***Escherichia coli*:** 3 complete genomes, *E. coli* CFT073 (NC_004431), *E. coli* O157:H7 (NC_002695) and *E. coli* O157:H7 EDL933 (NC_002655)
- **West Nile Virus:** 38 available complete genomes.
- **Japanese Encephalitis Virus:** 39 available complete genomes.

Category C Microbial and Viral Genomes:

- **Tickborne Encephalitis Virus:** 9 available complete genomes.
- **Yellow Fever Virus:** 18 available complete genomes.

Near-neighbor sequences:

- ***Yersinia pseudotuberculosis*:** 1 complete genome, *Y. pseudotuberculosis* IP 32953 (NC_006155), and 2 plasmid sequences.
- ***Shigella sonnei*:** 1 complete genome, *S. sonnei* Ss046 (NC_007384), and 1 plasmid sequence.

- **Other Flaviviruses:** 44 complete genome strains.

For each of the genomic sequences above, the genomic location and melting temperature (T_m) of all occurrences of all human blind n -mers were calculated in order to make them available through the database (Task 3).

1.4 Hardware upgrade and installation: 1 16Gb RAM workstation (HP) was purchased. All necessary software was installed. The 16Gb RAM machine was used primarily for calculations performed in subtasks 1.3 and 1.5.

1.5 Calculation of 20-22-mers: These calculations were performed in parallel with subtask 1.3. Due to the increased length of the signatures being designed in this subtask, the majority of the computations were required to be run on the 16Gb RAM workstation and run longer than the computations of subtask 1.3.

Task 2: Experimental validation

Biosafety: We established and certified a BSL-2 facility for handling the agents of interest in this work. Laboratory workers have recently undergone (refresher) training in BSL-2 operations.

Target nucleic acids: We obtained full-length yeast-*E. coli* shuttle vector cDNA clones of DEN1 (Western Pacific), DEN2 (NGC) and DEN4 from Drs. Barry Falgout and Robin Lewis at the FDA. We also received a generous supply of sterile genomic DNA of Sterne strain of *B. anthracis* from the laboratory of Prof. William Widger, University of Houston. We have extensively characterized this material for sterility, identity, and plasmid status; this testing has recently been completed.

Array hybridization testing of computationally-derived human-blind probes for *B. anthracis*: In order to validate the quality of human blind microarray probes we designed several Combimatrix CustomArray™ 12K microarrays. These assays were tested using human and *B. anthracis* (Sterne strain) DNA (DNA was fragmented using a 4-cutter restriction enzyme and then Cy3 or Cy5 labeled using the non-enzymatic ULS technology (Kreatech Biotechnology)). As one can see in the example of Fig. 1, the number of probes hybridized using of *B. anthracis* target DNA is significantly greater than in human DNA samples which do not contain anthrax. Considering that the human genome is 1000+ times longer, the fact that relatively few probes hybridize to human confirms the high quality of the probes produced by our computational approach.

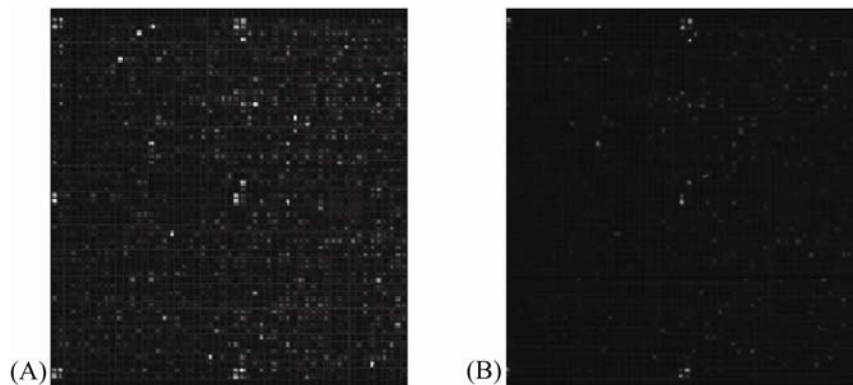


Fig. 1. Identical custom Combimatrix 12K arrays. Chip hybridized to: (A) *B. anthracis*-Cy3/control oligo-Cy5, (B) human-Cy3/control oligo-Cy5.

Experimental Validation of Computationally-derived Human Blind PCR Primers:

- **Cross-reactivity. Dengue genomes:** The first human-blind dengue primers designed for DEN1, DEN2, and DEN4 were tested for cross reactivity. DEN1 primers were tested with DEN2 and DEN4 genomes, DEN2 primers were tested with DEN1 and DEN4 genomes, and DEN4 primers were tested with DEN1 and DEN2 genomes. It was found that one set of DEN1 primers amplified a section of the DEN2 genome. Also various DEN4 primers amplified sections of DEN1 and DEN2 genomes. Therefore, in designing primers one must consider uniqueness between dengue types as well as from the human genome.

Human genome: Experiments were designed to test all dengue primers with human genomic DNA. Initially, no amplification was seen with all primers. However, the positive control also did not amplify. Several Mg²⁺ concentrations were tested along with four different polymerases. The human genomic target was successfully amplified with 5mM Mg²⁺ and Pfu DNA polymerase. Next, the dengue primers will be tested using these conditions.

The effect of the presence of the human genome in a dengue PCR reaction was studied. Four PCR reactions using equal amounts of DEN1 and a set of

human-blind primers were assembled each with increasing human DNA concentrations. It was found that the PCR reaction was inhibited only when a great excess - 200 ng - of human DNA was present.

- **Primer Length.** The effect of increasing primer length was investigated. A set of DEN1 primers was chosen and extended to 25, 30, 35, 40 and 45 base pairs (keeping the amplicon the same size). The PCR reaction was successful using all lengths of primers. We took delivery of the chip laser scanner and real-time PCR machine which will supplement our existing equipment as the heavy experimental phase of this project begins. We also developed RT-PCR protocols useful with RNA viruses such as Dengue.



Fig. 2. RT-PCR results of human total RNA. Lane 1: gDNA PCR using intra Alu Yd6 primers. Amplicon is 200 bp. Lane 2: cDNA PCR using Yd6 primers. Amplicon is 200 bp. Lane 3: No RT control. Lane 4 and 5: Non-template controls. Lane 6: gDNA PCR using beta-actin primers. Amplicon is slightly larger than the expected 208 bp. Lane 7: cDNA PCR using beta-actin primers. Amplicon is 208 bp. Lane 8: No RT control. Lane 9 and 10: Non-template controls.

- **Real-time PCR.** Eight sets of computationally derived primers were tested for DEN2 (Fig. 3) and DEN4 (Fig. 4) strains using real-time PCR. As it is shown, each primer set gave efficient amplification of the target DNA. Some cross amplification was obtained between DEN2 and DEN4 strains due to their similarity under the PCR conditions tried. New PCR conditions are being designed to eliminate nonspecific amplification. In addition, the primers were tested using genomic human DNA. Some nonspecific amplification was seen at later cycles. Design of new PCR conditions should correct this.

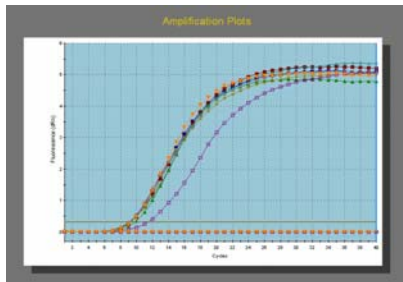


Fig. 3. PCR amplification of DEN2 cDNA. with human-blind PCR primer sets

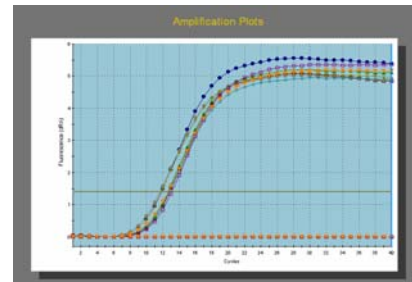


Fig. 4. PCR amplification of DEN4 cDNA with human-blind PCR primer sets.

- **New PCR conditions.** In order to minimize the nonspecific amplification seen with DEN2 and DEN4 'human blind' primers, the primer annealing temperature was optimized. First, touch-down PCR was attempted. In touch-down PCR, the early PCR cycles are at a higher annealing temperature decreasing nonspecific amplification. However, this was not as effective as thought. Therefore, the annealing temperature was increased 5 degrees for all cycles. This was effective and resulted in the elimination of nonspecific products as can be seen in the comparison of Fig. 5 and Fig. 6.

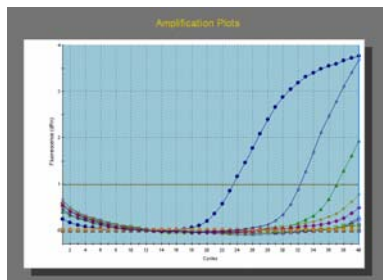


Fig. 5. DEN2 Primers w/Human DNA
T_{annealing} = 55° C

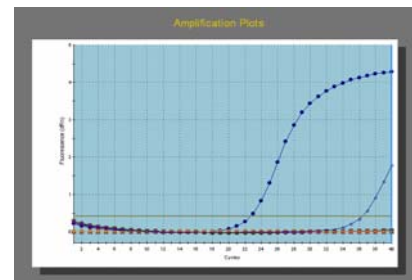


Fig. 6. DEN2 Primers w/Human DNA
T_{annealing} = 60° C

- **Efficiency.** The efficiency of each primer set for DEN1, DEN2, and DEN4 was calculated by creating standard curves (e.g. Fig. 7) of dilutions of dengue DNA. The experiments were performed in triplicate and the percent efficiency calculated.

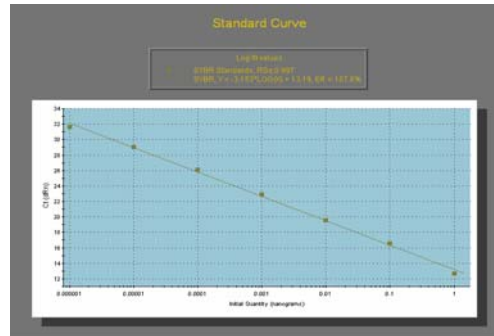


Fig. 7. DEN1 Standard Curve

- **Mixed Samples.** The human blind primers were tested with samples containing human DNA as well as dengue DNA. Gel electrophoresis was used to ensure the correct products were obtained.

Task 3: Database of host-blind sequences

3.1 Requirements Specifications and Design Document: Requirements necessary for the design of the host-blind database (HBDB) were identified and a design document was drafted and distributed to all members involved with Task 3 as well as DHS representatives present at our 1st Quarterly meeting in College Station, TX.

3.2 Implementation: The HBDB was developed using SQL Server 2005 with a web front end using .NET and PHP.

3.3 Testing: The implementation was extensively tested by a select group of team members and potential users within the UH community.

3.4 Hardware upgrade and installation: Two 2Gb RAM desktops (Dell) were purchased. All necessary software was installed. The two desktops were used for the database development and hosting.

3.5 Importing data to the database: As the data computed by the computing software developed in Task 1 became available it was imported into the database.

Background type	Number of signatures stored in the database		
	2MM	3MM	4MM+
Human genome	69,630,059	39,452,36	415,401
Air sample associated synthetic background (51 genomes)	100,389,144	12,747,032	1,092
Water sample associated synthetic background (23 genomes)	67,223,166	9,287,953	2,537
Human/indoor environmental sample associated synthetic background (63 genomes)	34,558,375	1,345,291	3

Table 1. Database statistics as of Feb. 15, 2006. The millions of sequences 1MM from their respective backgrounds are not included in the database. 20-, 21-, and 22-mers blind to the three environmental samples are also excluded from the database due to the fact that every n-mer is 4MM+ blind. This is a consequence of the relatively small size of the backgrounds with respect to the pathogen genome size, in particular the microbial pathogens.

3.6 Documentation (help, user manual): Help documentation including a quick reference guide and expanded users manual were both written and can be accessed through the web interface.

3.7 Database Delivery and Maintenance: This subtask is in progress. We are and will continue to make this database accessible to users approved by HSARPA.

Task 4: Windows-based application to compute host blind sequences

4.1 Requirements Specifications and Design Document: Requirements necessary for the design of the Windows-based application, called Host-blind Sequence Sleuth, were identified and a design document was drafted and distributed to all members involved with Task 4 as well as DHS representatives present at our 1st Quarterly meeting in College Station, TX.

4.2 Implementation: The application was implemented using C#.NET.

4.3 Testing: The implementation was extensively tested by a select group of team members and potential users within the UH community.

4.4 Documentation: Help documentation including a quick reference guide and expanded users manual were both written and can be accessed through the application's Help menu.

Task 4 Extension: Create a UNIX-based application allowing users to compute genomic location and melting temperature (T_m) of subsequences present in custom genomic sequence(s).

4.1 Requirements Specifications and Design Document: The design documentation created for the Windows-based application (Phase 1, Task 4.1) was used to develop the design documentation for the UNIX-based application. The requirements necessary to implement the increased functionality of the UNIX-based application in comparison with that of the Windows-based application were identified.

4.2 Implementation: The migration of the Windows-based computations that have already been developed served as an alpha version. Based upon the specifications outlined in the design document, the application was written and compiled.

4.3 Testing: The implementation was extensively tested by a select group of team members and potential users within the UH community.

4.4 Documentation: A user's manual as well as technical documentation was written for the application.

4.4 Application Delivery and Maintenance: The application is now freely available to all HSARPA bidders approved by HSARPA. Application delivery includes the application and all documentation. Source code is freely available for all users upon request and is documented thoroughly such that users can further develop or customize the application in the future.

KEY RESEARCH ACCOMPLISHMENTS

In addition to the results of the specific tasks listed above, several additional key research accomplishments emanated from this research, specifically:

- **Development of a new generation of long (22 - 60 nucleotides) ultra-specific pathogen signatures.** One of the most intriguing scientific results of Phase I of the project was the development of a **new generation of long ultra-specific signatures**, allowing us to design even more robust and specific (blind to the background) pathogen signatures of **virtually any length**. The basic idea of this new approach is to design genomic signatures to be not only blind (several mismatches away) from the host/background, but also make sure that **every subsequence of particular length present in the signature is also significantly blind to the background** (Figure 8). For instance, instead of designing a signature of 25 nucleotides which is 4 mismatches away from the nearest background sequence, a much more specific signature of same length can be designed by insuring that each of the (in this particular example six) 20-mers present in the signature is at least 3 mismatches away from the background. This new design technique can significantly improve the specificity and sensitivity of the signature because: (1) it **increases the quality of probes and primers** by excluding unreliable cases where all mismatches are located on the head or tail of the probe (and primer); (2) it is **computationally much less expensive**; and (3) it supports the **design of probes and primers of virtually any length**. To design these longer ultra-specific signatures, we have developed new algorithms and data structures for the computation and storage of the distance (number of mismatches) from the background for each short (16-20 nucleotides long) n -mer at each position in the target pathogen genome.

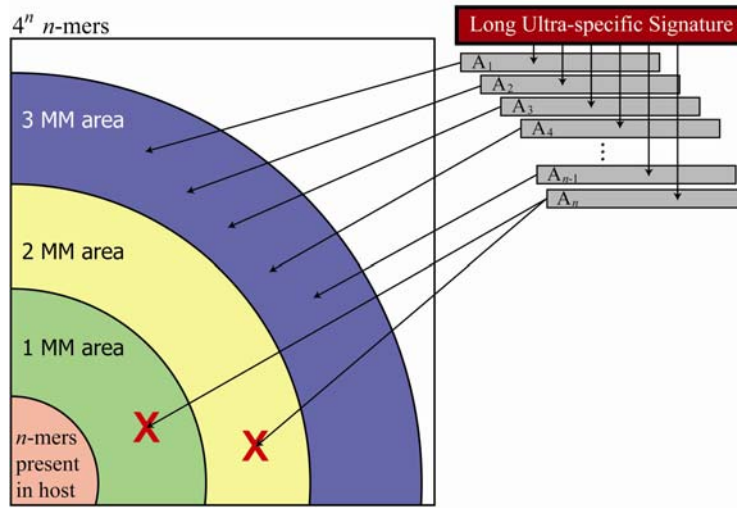


Fig. 8. Long ultra-specific signatures: A_1 - A_n represent all subsequences of a particular size which are present within the longer signature.

- New approach to designing sets of signatures with the increased ability to detect false positives.** The significance of using a set of frequently-present orthogonal (not correlated by appearance between target genomes) probes/primers instead of a few unique common probes/primers for pathogen identifications is that the unique presence/absence pattern provides the additional ability for pathogen classification. For example, the parsimony tree on Figure 9 was created using the presence/absence pattern of 4000 randomly selected probes in "in silico" hybridization experiments with 89 strains of the Dengue virus of serotypes 1-4. Based solely on their presence/absence pattern, the Dengue strains clearly cluster according to their serotype and are further grouped by their origins and the time at which the samples were taken. The presence/absence pattern of a new strain is expected to fall into its respective serotype cluster, and to be closer in proximity within the tree to strains of the same origin. In the event that the new pattern does not fit within the serotype cluster (e.g., the branch labeled with "?" in Figure 9), it is highly likely that the organism detected is in fact not Dengue. Thus the presence/absence pattern provides an additional tool for detecting false positives. Taking into consideration the combinatorial nature of the presence/absence patterns, such an approach is expected to increase our ability to recognize false positive calls well beyond what is currently possible.

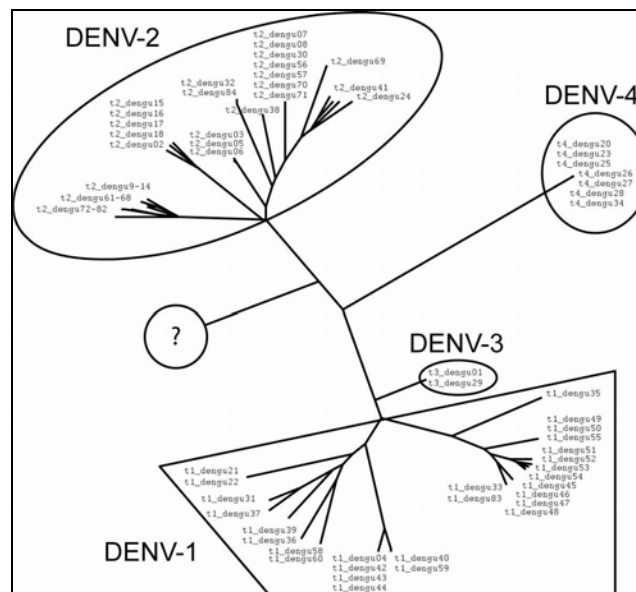


Fig. 9. Parsimony tree created using presence/absence pattern of 4000 probes in "in silico" hybridization experiments with 89 serotype 1-4 strains of Dengue virus. Branch marked by "?" signifies potential outliers (false positive call).

- **Sets of primers with the ability to detect emerging or specifically engineered pathogens.** This new approach is based on recently discovered statistical properties of microbial/viral genomes³⁶. The basic idea is to increase the size of the set of background blind signatures in order to make it robust to mutations in the target genomes (Figure 10). We believe that this new strategy will allow a shift in the design strategy of diagnostic tests from identification of sets of known (sequenced) genomes to detecting sets of pathogens which may evolve or be engineered with certain mutation rates from the set of "landmark genomes" used in the original design. We are planning to explore this possibility further during Phase II of the project.

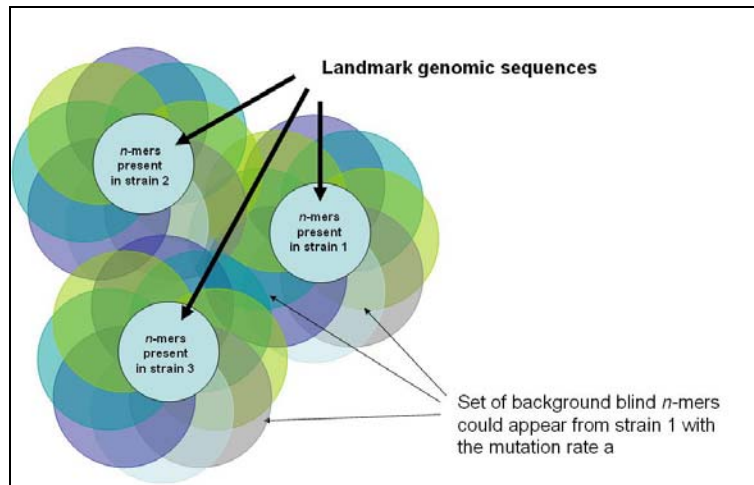


Fig. 7. Sets of primers to detection unsequenced naturally occurring relatives of the landmark strains or specifically-engineered pathogens.

REPORTABLE OUTCOMES

Manuscripts, abstracts, presentations:

PEER-REVIEWED ARTICLES

1. Putonti, C., Fofanov, V., Fofanov, Y. Computational Methods for Estimating the Size and Identifying the Members of Microbial Communities. In: Du, D.-Z. (ed.) *Proceedings of the IASTED International Conference on Computational and Systems Biology* (Dallas, Texas), 2006: 540-033.
2. Zhang, M., Putonti, C., Fofanov, Y. A New Set of Algorithms for Designing Ultra-specific Probes for High-throughput Allele Detection. In: Du, D.-Z. (ed.) *Proceedings of the IASTED International Conference on Computational and Systems Biology* (Dallas, Texas), 2006: 540-038.
3. Putonti, C., Chumakov, S., Mitra, R., Fox, G.E., Willson, R.C., Fofanov, Y. Human-blind probes and primers for Dengue virus identification: Exhaustive analysis of subsequences present in the human and 83 Dengue genome sequences. *FEBS J.* 2006; 273(2): 398-408.
4. Chumakov, S., Belapurkar, C., Putonti, C., Li, T.-B., Pettitt, B.M., Fox, G.E., Willson, R.C., Fofanov, Y. Theoretical Basis for Universal Identification Systems for Bacteria and Viruses. *Journal of Biological Physics and Chemistry.* 2005; 5(4); 121-128.

CONFERENCE ABSTRACTS (TALKS & POSTERS)

5. Putonti, C., Añez, M., Kourentzi, K., Fox, G.E., Willson, R.C., Fofanov, Y. Ultraspecific assays for the detection of pathogens in human samples. ASM Biodefense and Emerging Diseases Research Meeting (Washington, D.C.), 2007: 50.
6. Putonti, C., Feng, C., Quance, M., Fofanova, T., Fofanov, Y. Computational challenges in designing assays for the detection of infectious agents. 24th Annual Houston Conference on Biomedical Engineering Research (Houston, TX), 2007: 256.
7. Putonti, C., Añez, M., Fox, G.E., Willson, R.C., Fofanov, Y. Ultraspecific PCR Assays for Dengue Detection. Texas Branch Meeting of the American Society for Microbiology Fall Meeting (Galveston, TX), 2006: 34.

8. Fofanov, Y. A novel approach in the design of primer and probe sequences for the detection of pathogens in the presence of complex backgrounds. Texas Branch Meeting of the American Society for Microbiology Fall Meeting (Galveston, TX), 2006: Invited Speaker.
9. Putonti, C., Willson, R.C., Fox, G.E., Fofanov, Y. Ultraspecific Assay Design for the Detection of West Nile virus in Human Samples. Advances in Microarray Technology (Amsterdam, Netherlands), 2006. *Invited speaker.*
10. Añez, M., Putonti, C., Fox, G.E., Fofanov, Y., Willson, R.C. Ultraspecific PCR Assays for Dengue Virus Detection. 62nd Southwest Regional Meeting of the American Chemical Society (Houston, TX), 2006: 66.
11. Putonti, C., DeVille, T., Fox, G.E., Willson, R.C., Fofanov, Y. The Host-Blind Database: A Portal to Ultraspecific Signatures for Assay Design. Chemical and Biological Countermeasure Technologies for Homeland Security (Boulder, CO), 2006.
12. Putonti, C., Travisano, M., Willson, R.C., Fofanov, Y. Improving Nucleic Acid-Based Assay Design through the Estimation of the Total Genomic Diversity of Complex Backgrounds. Chemical and Biological Countermeasure Technologies for Homeland Security (Boulder, CO), 2006.
13. Putonti, C., Zhao, Q., Willson, R.C., Fox, G.E., Fofanov, Y. The Host-Blind Database: A portal to ultraspecific signatures for assay design. 11th Structural Biology Symposium (Galveston, TX), 2006: 127.
14. Putonti, C., Añez, M., Willson, R.C., Fox, G.E., Fofanov Y. Ultraspecific PCR assays for the detection of the dengue virus in human samples. Bio-Sensors, Assays, and Informatics for Homeland Security (Monterey, CA), 2006: 45.
15. Fofanov, Y., Willson, R.C., Fox, G.E., Putonti, C. HBDB: A portal to ultraspecific signatures for nucleic acid-based assay design. Bio-Sensors, Assays, and Informatics for Homeland Security (Monterey, CA), 2006: 26.
16. Putonti, C., Fofanov, V., Reed, C., Fofanov, Y. A novel approach to assay design for pathogen discovery. Bugs, Drugs, and Vaccines: Securing Our Future Symposium (Galveston, TX), 2006: P39.
17. Añez, M., Putonti, C., Fox, G.E., Fofanov, Y., Willson, R.C. Background-resistant PCR assays for dengue detection. Bugs, Drugs, and Vaccines: Securing Our Future Symposium (Galveston, TX), 2006: P22.
18. Añez, M., Putonti, C., Fox, G.E., Fofanov, Y., Willson, R.C. Efficient 'host-blind' PCR Primers for Dengue Virus Detection. Sigma Xi Scientific Research Society Poster Competition (Houston, Texas), 2006: 16.
19. Putonti, C., Añez, M., Fox, G.E., Fofanov, Y., Willson, R.C. Pathogen identification using ultraspecific sequences "blind" to host DNA. DNA Structure, Genomic Rearrangements, and Human Disease (Houston, Texas), 2006: 69.
20. Anez, M., Putonti, C., Fox, G.E., Fofanov, Y., Willson, RC. Background-Resistant PCR Assays for Dengue Detection. XX Annual OCHEGS Symposium. Organization of Chemical Engineering Students (Houston, Texas), 2005: 29.
21. Anez, M., Putonti, C., Fox, G.E., Fofanov, Y., Willson, RC. Background-Resistant PCR Assays for Dengue Detection. The 2005 Annual Research Conference, hosted by the Keck Center for Interdisciplinary Bioscience Training (Houston, Texas), 2005: 33.
22. Putonti, C., Mitra, R., Fox, G.E., Chumakov, S., Willson, R.C., Fofanov, Y. Dengue and Yellow Fever Identification Using Ultraspecific Probes "Blind" to Hosts and Aedes aegypti DNA. The XIII International Congress of Virology (San Francisco, CA), 2005: 88. *Symposia Speaker.*
23. Putonti, C., Fofanov, V., Mitra, R., Fox, G.E., Chumakov, S., Willson, R.C., Fofanov, Y. Flavivirus Detection Using Ultraspecific Human-blind Probes. European Conference on Mathematical and Theoretical Biology (Dresden, Germany), 2005: 460. *Contributing Speaker.*
24. Putonti, C., Mitra, R., Fox, G.E., Chumakov, S., Willson, R.C., Fofanov, Y. Pathogen Detection Using Ultraspecific Probes "Blind" to Vector DNA. Bioinformatics and Disease SIG (Detroit, MI), 2005: 13.
25. Añez, M., Putonti, C., Fox, G.E., Fofanov, Y., Willson, R.C. Efficient 'host-blind' PCR Primers for Dengue Virus Detection. ACS National Meeting (San Francisco, California), 2006.

Degrees obtained that are supported by this award:

Catherine Putonti, Ph.D. Computer Science
 Qin Zhao, Ph.D. Biology and Biochemistry
 Qin Zhao, M.S. Computer Science
 Aaron Skewes, M.S. Electrical Engineering

Infomatics:

Host-blind Database (HBDB)

Personnel receiving pay from the research effort:

Dr. Yuriy Fofanov (PI),
Mr. Mark Fortner (Programmer),
Dr. George Fox (Co-PI),
Mr. Stephen Huff (Ph.D. student)
Dr. Katerina Kourentzi (Technician)
Mr. Roy Luo (Ph.D. student)
Dr. Catherine Putonti (Key Personnel)
Mr. Aaron Skewes (Ph.D. student)
Dr. Petri Urvil (Technician)
Mr. Max Willson (Undergraduate student)
Dr. Richard Willson (Co-PI)
Dr. Qin Zhao (Ph.D. student)
Mr. Dianhui Zhu (Ph.D. student)

CONCLUSIONS

The key computational challenge of Phase I was to transform in-house scientific software to several fully-developed end user applications. The key experimental challenge was to establish an infrastructure (including establishing BSL-2 facilities, installation of appropriate equipment and training of the personnel) and begin testing/validating the first set of probes/primers designed using our advanced computational methods. The Phase I studies demonstrated the advantages of the proposed design technology through extensive analysis of several Category A and B pathogens, the delivery of a database of all human and/or "background"-blind (2, 3, and 4 mismatches away) 16-22-mers present in the pathogens' genomes, and experimental validation with the goals of verifying candidate sequences and obtaining an experimental estimation of the false positive rate. The overall goal of Phase II is to advance this project from the "proof of concept" and "demonstration of advantages" of Phase I to a useful resource available to the scientific/business community focused on protecting the United States from the possible release of biological threats.