

IN VIVO GENE EXPRESSION PROFILING OF THE PLAGUE BACILLUS, *YERSINIA*
PESTIS

APPROVED BY SUPERVISORY COMMITTEE

Stephen Albert Johnston, Ph.D. _____

Harold Ray (Skip) Garner, Ph.D. _____

Eric J. Hansen, Ph.D. _____

Robert Sims Munford, III, M.D. _____

DEDICATION

Science Influences

I would like to thank my mentor Dr. Stephen Albert Johnston, whose enthusiasm and creative thinking about science has been both an invaluable lesson and a fun experience. I am lucky to be trained by a scientist who appreciates the potential, the fluidity, and the responsibility of our chosen field. My committee members Dr. Gale, Dr. Garner, Dr. Hansen, and Dr. Munford for their guidance and support throughout my graduate career. I would also like to thank Dr. Michael Norgard for his willingness to allow me to take a graduate course in his department before I started graduate school, as well as his personal support as a Department Chair. Additionally, I must acknowledge Dr. Orhan K. Oz, for whom I worked with prior to graduate school. Dr. Oz's personal and professional support was instrumental in me pursuing science as a career. I thank him for never treating me like "just a technician", and challenging me to grow by treating me as a colleague. During my undergraduate studies at Baylor University I was well prepared for graduate school by many people, and I will mention Dr. Kearney and Dr. Baldrige for the tremendous amount of personal scientific

training they provided. Mr. Bittner was my High School Biology teacher for two years and the fundamental biological concepts he taught in such creative ways provides a lasting foundation for my science knowledge. Finally, I need to thank my parents for not only tolerating, but also fostering my childhood curiosity about nature. Understanding science would have been impossible without the experiences out in God's great earth that you gave me growing up. Thank you all.

My Love

I must begin and end with my wife Lisa, who agreed to marry me after I gave her a cheap gold ring on the shores of Whiterock Lake in Dallas in April of 2000. This was two weeks before I had any idea I would be accepted to graduate school. This is just one of many times you have demonstrated your love to me through unwavering confidence and faith in my potential. Not that my career has ever been the most important thing in our relationship, but I have needed your confidence when mine own periodically waned. I can say with the greatest of confidence that I would not have been able to reach this milestone had it not been for your constant love and support. I share all the positive things that may come from this experience, because you have earned them with me. I love you so much.

My Family

Again I begin with my wife Lisa. We became a family on our wedding day before our God, our family, and our friends. May we always be a three-stranded cord that cannot easily be broken. This last August 25th you provided me with the most amazing experience of my life

to date when Hannah was born. This new family adventure we have embarked on would be filled with hesitation and fear if it were not for the confidence and love I have in you and our marriage.

Mentioned earlier was my parents who have been the constant supporters, sometimes financial and always loving, of my aspirations. Thank you again. My siblings David, Lara and Rebecca, thanks for reminding me that although we are not able to pick our family, we can be so lucky as to be blessed with one filled with such wonderful friends. Furthermore, I would like to say that in this day and age it is rare to know that you have the love and support of your extended family and to be so frequently reminded with cards and calls.

To the Mitchell family, I would like to say thank you for helping to shape Lisa into the wonderful woman she is today. Before we were married I already felt a part of your family, and that love and support has only grown over time. Jim and Sheree, your proximity has been a blessing that which I cannot express. Graduate life with your sourdoughs and frequent phone calls is so much more tolerable. Jason and Gena, who are also engaged in writing their thesis, it is great to share and complain about common experiences. To the rest of the Mitchell clan I thank you for your continuing love and support, even if I am a Macintosh user.

I would like to thank my church family at Marsh Lane Baptist Church of Dallas for the support, love, laughs, and listening ears they provide. Lisa and I truly feel part of a family.

This would not be complete if I did not thank the newest member of my family, Hannah Louise. Although you do not know it, or realize that you are doing it, you have and continue to be a great source of love and inspiration in my life. Before you were born and after, you

continue to humble me in my pursuit of science by reminding me how little we know about biological life and more importantly why I should work hard to move science forward so that it may continue to provide help through medicine, education, and quality of life.

My Friends

I have been lucky to develop personal friendships throughout my life that have been untouched by distance or time apart. So many of you have been outlets for my frustrations and supporters of my successes. I always know that even if we have been separated by time and distance that when we get together we are right back where we left off. If I were to thank all of you personally I could easily tack on another 10 pages to this book none of you will read or should be made to for that matter, unless your having trouble sleeping. So here is a list in no particular order: Dr. Dan Hendrick, Dr. Brent (Berntly) Weiland, Gavin Jackson, Jerrett Morris, Warren Langley, William Spencer, Jeff Milush, Jill Mooney, Jody and Xandra Smith, Stephanie Little and the Dement family, Rick and Kay Kingland (thanks for Big Blue), all of Lisa's ASL friends.

I have also been lucky to develop many wonderful professional friendships. These have helped me both to develop professionally but also socially in this social profession of science. I would like to mention Dr. Adel Talaat, Dr. Mike McGuire, Dr. Ross Chambers, Dr. Xandra Smith, Dr. Kevin Luebke, Dr Kathlyn Brown, Dr. Reddy Moola, Dr. Dave Fancy, Eunice Webb, Qihua Sun and many others for contributing so much to my graduate training. I would like to thank all the past and current students of the Molecular

Microbiology Graduate program, as well as all the many administrative support staff that have helped me during my training.

My Faith

None of this would be possible without the knowledge of my salvation through Jesus Christ our Lord. In a profession where God is not a part of science because of a lack of empirical evidence I am comforted by the data I experience in my heart

IN VIVO GENE EXPRESSION PROFILING OF THE PLAGUE BACILLUS, *YERSINIA*
PESTIS

by

JONATHAN NEALE LAWSON

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

July 2006

Copyright

by

Jonathan Neale Lawson 2006

All Rights Reserved

IN VIVO GENE EXPRESSION PROFILING OF THE PLAGUE BACILLUS, *YERSINIA*
PESTIS

Publication No. _____

Jonathan Neale Lawson, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2006

Supervising Professor: Stephen Albert Johnston, Ph.D., Ph.D..

Yersinia pestis, the causative agent of plague, can be transmitted by infected-flea bite or inhaled aerosol. Both routes of infection have a high mortality rate, and pneumonic infections of *Yersinia pestis* represent a significant fear as a tool of bioterrorism.

Understanding the transcriptional program of this pathogen during pulmonary infection is valuable to better understand plague pathogenesis. Improved understanding of the mechanisms employed by *Y. pestis* to cause illness may lead to new vaccines or therapies.

Using a long-oligonucleotide microarray to the plague bacillus I evaluated the expression profiles of *in vitro* and *in vivo* models of *Y. pestis*. The change in temperature from ambient (e.g. aerosol or flea midgut) to the mammalian host has been used as a model to understand the effects of temperature as a signal for virulence gene expression. I have profiled transcription during an aerosol delivered mouse infection. By amplifying the *Y.*

pestis RNA from individual mouse lungs, I was able to map the transcriptional profile of plague at post-infection days 1 to 3. My data suggest a dramatically altered transcriptional profile relative to the *in vitro* model, suggesting *Y. pestis* is responding to a variety of host signals during infection. Of note was the number of genes found in genomic regions with altered %GC content that are up-regulated within the mouse lung environment. These data suggest these regions may provide promising targets for future therapy design and vaccine target discovery.

TABLE OF CONTENTS

BACKGROUND AND LITERATURE REVIEW	25
Introduction.....	25
<i>Yersinia pestis</i> background and Historical perspective	26
General Biology	29
Plague is a zoonotic disease.....	29
Clinical Presentations of Plague in Humans	32
Clinical Aspects of <i>Yersinia pestis</i>.....	34
Diagnosis of <i>Yersinia pestis</i> infections	34
Therapy against the plague	36
Immunity to plague.....	36
Prophylaxis against the plague.....	38
Plague pathogenesis	41
Animal Models of Plague Infection.....	41
Pathogenesis in the Flea.....	41
Pathogenesis in mammalian hosts.....	42
Pulmonary plague pathogenesis.....	43
Known Virulence factors	44
Plasmid pCD1	44
Plasmid pMT1.....	45
Plasmid pPCP.....	45
Chromosomally encoded virulence factors.....	46
Targeting of immune cells during infection.....	46
Plague as a biothreat agent	47
<i>In vivo</i> expression technologies	49
Background and Introduction.....	49
Unique features of bacterial microarray studies	52
<i>In vivo</i> microarray studies.....	52
MATERIALS AND METHODS	54
Introduction.....	54
Microarray Resource and technology development	55
General Design considerations.....	55

Construction of <i>Y. pestis</i> Microarray	57
Summary	57
Probe design.....	59
Probe synthesis.....	65
Purification of Oligos.....	66
Concentration Determination of Oligos.....	67
Mechanical spotting	68
Microarray validation.....	70
General quality.....	70
Specificity of <i>Y. pestis</i> array	71
Reproducibility.....	72
RNA isolation Protocols.....	72
Quality evaluation of RNA and cDNA.....	72
RNAwiz or Trizol methods for RNA extraction.....	73
<i>Y. pestis</i> RNA from plate cultures.....	74
Protocol for CFU count and Enrichment of <i>Yersinia pestis</i> from mouse lungs.....	74
Analyte Labeling Protocols	75
Generation of Genomic standards for Microarrays.....	75
Generation of labeled cDNA.....	78
Direct labeling of RNA.....	79
Indirect labeling of RNA.....	81
Generation of GFP RNA template.	83
LAPT protocol for RNA amplification	83
Summary	83
Detailed LAPT Protocol.....	84
<i>In vitro</i> Transcription.	86
Amplification of <i>in vivo</i> samples	87
Microarray Hybridization and Analysis.....	88
Data analysis.....	88
Quantitative Real-Time PCR (QRT-PCR).....	89
Models of <i>Yersinia pestis</i> infection	90
Bacterial growth and RNA preparation.	90
The <i>Yersinia pestis</i> clinical isolate used in these studies.....	91
<i>In vitro</i> Growth of <i>Y. pestis</i> and RNA isolation.....	91
Pulmonary infection model of <i>Y. pestis</i>	92
CFU determination of infected mouse tissues.	92

Sample collection and RNA isolation from mouse pulmonary infection model. .	92
AMPLIFICATION OF SENSE-STRANDED PROKARYOTIC RNA.....	94
Summary.....	94
Introduction.....	95
Results	98
Outline of LAPT process.	98
Addition of a functional T7-promoter to the 5' end of GFP RNA.	99
Microarray construction and amplification of complex RNA samples.	101
Maintenance of biological information of a complex RNA sample.	103
Selective amplification from a mixed sample.....	106
EXPRESSION PROFILING OF <i>YERSINIA PESTIS</i> DURING MOUSE PULMONARY INFECTION.....	108
Summary.....	108
Introduction.....	109
Results	112
Characterization of the arrays using temperature-regulated gene expression <i>in vitro</i> . .	112
Characterization of pulmonary infection model	115
Amplification enables <i>in vivo</i> expression profiling.....	118
Global comparison of <i>in vitro</i> and <i>in vivo</i> expression profiles	122
Regulation of <i>Y. pestis</i> gene expression during pulmonary infection.....	124
Preferential <i>in vivo</i> expression of chromosomal pathogenicity island (PI) and PI-like regions.....	125
DISCUSSION	130
LAPT Discussion	130
Future directions	133
<i>In vivo</i> Expression Analysis Discussion.....	136
Future Directions.....	139

PRIOR PUBLICATIONS

Ng, Kar-wai, **Jonathan Lawson**, Harold Garner. 2004. "PathoGene: A Pathogen Coding Sequence Discovery and Analysis Resource". Biotechniques, vol. 37 no. 2 (2004)

Yuri Y. Belosludtsev, Dawn Bowerman, Ryan Weil, Nishanth Marthandan, Robert Balog, Kevin Luebke, **Jonathan Lawson**, Stephen A. Johnston, C. Rick Lyons, Kevin O'Brien, Harold Garner, and Thomas F. Powdrill. 2004. "Organism Identification Using a Genome Sequence-Independent Universal Microarray Probe Set". Biotechniques, 37(4):654-8.

Orhan K. Oz, Gen Hirasawa, **Jonathan Lawson**, Lydia Nanu, Anca Constantinescu, Peter P. Antich, Ralph P. Mason, Edward Tsyganov, Robert W. Parkey, Joseph E. Zerwekh, Evan R. Simpson. "Bone Phenotype of the Aromatase Deficient Mouse". Journal of Steroid Biochemistry and Molecular Biology, 79 (2001), pg.49-59.

Manuscripts Submitted:

M. Muralidhar Reddy*, **Jonathan Lawson***, Stephen Albert Johnston and Thomas Kodadek. "Toward A Minimal Small Molecule Microarray For Protein Fingerprinting: The Utility of Promiscuous Protein-Binding Agents." Molecular Biosystems (Submitted).

Lawson, Jonathan, Stephen Albert Johnston. 2005. "Amplification of Sense-Stranded Prokaryotic RNA". DNA and Cell Biology (In Review).

Lawson, Jonathan, C. Rick Lyons, Stephen Albert Johnston. 2005. "Temporal Expression of *Yersinia pestis* in Mice". Infection and Immunity (In Review)

LIST OF FIGURES

Figure 1. Timeline of Classical Plague Pandemics.....	27
Figure 2 Zoonotic life cycle of <i>Yersinia pestis</i>	30
Figure 3 Overall process of microarray construction.....	59
Figure 4 Graphical representation of Genomic data manipulation prior to probe design.	64
Figure 5 A typical quality control gel of 70mer oligos.....	66
Figure 6 General Quality of mechanically spotted microarrays.	70
Figure 7 Specificity of <i>Y. pestis</i> microarrays	71
Figure 8 Microarray reproducibility.	72
Figure 9 gDNA labeling using amino-allyl modified nucleotide.	76
Figure 10 Microarray sample handling for RNA processing through data analysis.....	78
Figure 11 Models of <i>Y. pestis</i> infection used in these studies.....	91
Figure 12 Depiction of the steps in the LAPT process.	99
Figure 13 Verification of addition of a functional T7-promoter to the 5' end of GFP transcripts.	100
Figure 14 Characteristic bioanalyzer pattern of amplified RNA.	102
Figure 15 Analysis of microarray signals from un-amplified and LAPT amplified RNA.	104
Figure 16 LAPT amplification maintains expression information on microarrays.	106
Figure 17 Scatter plots of Early (0.5 hr) and Late (2 hr) gene expression differences...114	
Figure 18 Correlation of RT-PCR and Microarray Data.	115
Figure 19 Survival curve and dissemination of YPMN1 in mice.....	117
Figure 20 Validation of <i>in vivo</i> array expression data with QRT-PCR	121
Figure 21 Cluster analysis of <i>in vivo</i> and <i>in vitro</i> expression patterns of <i>Y. pestis</i>	122
Figure 22 Temporal <i>Y. pestis</i> gene expression changes from mouse lungs.....	125
Figure 23 In vivo expression of a gene cluster within a predicted pathogenicity island. 126	
Figure 24 Differential expression of gene clusters in vivo.	127

Figure 25 Pathogenicity island with predicted outer-membrane proteins expressed <i>in vivo</i>	128
Figure 26 <i>In vivo</i> expressed genomic island of <i>Y. pestis</i>	129
Figure 27 Summary of <i>B. anthracis</i> microarray features.	145
Figure 28 Specificity of <i>B. anthracis</i> array.....	146
Figure 29 Reproducibility of <i>B. anthracis</i> arrays using a closely related <i>Bacillus</i> species.	146
Figure 30 Diagram of the hybridization based annotation method.....	152
Figure 31 Example of the <i>Mtb</i> 25kb array hybridized with labeled cDNA from H37Rv grown to log phase.	153
Figure 32 Graph of positive strand/negative strand probe intensities.....	154
Figure 33 General schema for the functional annotation of microbial genomes using D.O.C. high-density arrays.	155
Figure 34 Signal patterns of an annotation array hybridized.	156
Figure 35 Genomic region of <i>Y. pestis</i> used for validation of Functional annotation. ..	156
Figure 36 Resolving power of the arrays was tested using the ribosomal operon.....	158
Figure 37 Single base-pair resolution of the ribosomal operon is enhanced with mis- matched probe normalization.....	158
Figure 38 Functional evaluation of probes.	159
Figure 39 Bioinformatic strategy for identifying identical peptide sequences between multiple genomes.	163
Figure 40 Clustal W alignments of gene pairs with common peptides.	164
Figure 41 Proposed experimental approach to identify cross-protective epitopes.	167
Figure 42 Summary of PEPSEL process.	172
Figure 43 Comparison of tripeptide frequency in mouse compares to the gram-negative pathogen <i>Y. pestis</i>	174
Figure 44 Organism input portion of PathoGene user interface	180
Figure 45 Flowchart describing PathoGene processing.....	181
Figure 46 Experimental approach to presymptomatic serum proteomic changes following influenza A infection.....	190

Figure 47 Scatter plot of serum protein changes 48hrs post-infection with influenza A.	191
Figure 48 Generation of TCR diversity.	196
Figure 49 Outline of experimental approach to immunosignatures.....	198
Figure 50 Proposed system for validation of TCR-based immunosignatures	199
Figure 51 Peptoid library construction.....	206
Figure 52 Raw images of peptoid library microarrays after hybridization with three labeled proteins.	206
Figure 53 Venn diagram to identify promiscuous binding peptoids from three unrelated proteins.....	208
Figure 54 Hybridization of AAT to the peptoid array.	209
Figure 55 Inverse correlation of variance with peptoid signal intensity.....	210
Figure 56 Scatter plots of peptoid array data for the protein AAT.....	211
Figure 57 Graphical representation of peptoid array data.....	212
Figure 58 Blinded samples are identified based on hybridization patterns to the peptoid array.	214

LIST OF TABLES

Table 1 Sample table of genomic information for <i>Y. pestis</i>	61
Table 2 Summary of 70mer probe designed to <i>Y. pestis</i> genome.....	65
Table 3 Typical yields from gDNA labeling reaction.....	77
Table 4 Typical RNA yields from LAPT.....	103
Table 5 Spearman correlations of amplified <i>Yersinia</i> samples relative to an unamplified control.	107
Table 6 LD ₅₀ of YPNM1 in four strains of mice.....	116
Table 7 Amplification allows detection of more of the transcriptome.	119
Table 8 Differences between BioMarkers and BioSignatures.	188
Table 9 Proteins found to have the greatest change in the serum following influenza infection	192

APPENDICES

TABLE OF CONTENTS

BACILLUS ANTHRACIS MICROARRAYS	141
Objective summary	141
Introduction.....	141
Construction of a Microarray for the Anthrax bacillus, <i>Bacillus anthracis</i>	143
Genome annotation	143
Probe Design	144
Probe Synthesis	145
Microarray validation.....	145
Summary.....	147
FUNCTIONAL ANNOTATION THROUGH HYBRIDIZATION OF TRANSCRIBED PRODUCTS.....	149
Introduction.....	149
Methods and results.....	151
Basic concept	151
Initial Experiments in Functional Annotation.....	152
D.O.C. platform for functional annotation.....	154
Focused evaluation of a functionally annotated genomic region in <i>Y. pestis</i>	156
Resolving power of functional annotation	157
Functional evaluation of probe hybridization efficiency	159
Discussion.....	160
BIOINFORMATIC STUDIES	161
Identification of potential cross-protective antigens.....	161
Summary	161
Methods and Results	162
Proposal for the identification of a single-antigen and cross-protective vaccine	165
Statement	165
Bioinformatics.....	167

Vaccine construction.....	168
The Challenge – Protection Experiment	169
PEPSEL - Rapid screening of sequence space for the identification of vaccine targets	171
Introduction.....	171
Intellectual property disclosure.....	171
Results and Application	174
PathoGene: A Pathogen Coding Sequence Discovery and Analysis Resource.....	176
Summary	176
Introduction.....	176
Materials and Methods.....	178
Computational Resources	178
Validation Resources	178
Sequence Analysis and Algorithm.....	179
Output Format.....	182
Validation Method	183
Results and Discussion.....	184
BIOSIGNATURES OF INFECTION	187
Serum Protein Signature of Pre-Symptomatic Influenza Infection	187
Methods.....	189
Results.....	191
Summary	193
Immunosignatures - High-throughput Patterning of T-cell Repertoires	195
Statement.....	195
Research Impact.....	196
Research Plan.....	197
TOWARD A MINIMAL SMALL MOLECULE MICROARRAY FOR PROTEIN FINGERPRINTING: THE UTILITY OF PROMISCUOUS PROTEIN-BINDING AGENTS.....	201
Abstract.....	201
Introduction.....	202
Results	205
Identification of Promiscuous Peptoids In A Large Library.....	205
The Majority of the Promiscuous Peptoids Bind The Previously Unstudied Protein AAT.	208

The Promiscuous Peptoids Support Protein Fingerprinting.....	211
Discussion.....	214
Materials and Methods.....	217
Direct Labeling of Proteins.	217
Peptoid array construction.	217
Microarray hybridization and data acquisition.	217
Image/Data analysis.	218
Cell binding experiments to peptoid microarrays	219

LIST OF ABBREVIATIONS

AA	AMINO-ALLYL
AAT	ASPARTATE AMINOTRANSFERASE
ANOVA	ANALYSIS OF VARIANCE
BCR	B-CELL RECEPTOR
BLAST	BASIC LOCAL ALIGNMENT SEARCH TOOL
BP	BASE PAIR
BSA	BOVINE SERUM ALBUMIN
Ca ²⁺	CALCIUM DIVALENT CATION
CDNA	COMPLEMENTRY DEOXYRIBONUCLEIC ACID
CFU	COLONY FORMING UNITS
CO ₂	CARBON DIOXIDE
CY3	CYANINE 3
CY5	CYANINE 5
D.O.C.	DIGITAL OPTICAL CHEMISTRY
DATP	2'-DEOXYADENOSINE 5'-TRIPHOSPHATE
DCTP	2'-DEOXYCYTIDINE 5'-TRIPHOSPHATE
DFI	DIFFERENTIAL FLUORESCENCE INDUCTION
DGTP	2'-DEOXYGUANOSINE 5'-TRIPHOSPHATE
DMSO	DIMETHYL SULFOXIDE
DNA	DEOXYRIBONUCLEIC ACID

DNTP	DEOXYRIBONUCLEOTIDE TRIPHOSPHATE
DTT	DITHIOL THERETOL
DTTP	2'-DEOXYTHYMIDINE 5'-TRIPHOSPHATE
DUTP	2'-DEOXYURIDINE 5'-TRIPHOSPHATE
ELI	EXPRESSION LIBRARY IMMUNIZATION
EST	EXPRESSED SEQUENCE TAG
ETOH	ETHANOL
FACS	FLUORESCENCE ASSISTED CELL SORTING
GDNA	GENOMIC DNA
GDPS	GENOME DIRECTED PRIMERS
GFP	GREEN FLUORESCENT PROTEIN
GST	GLUTATHIONE-S-TRANSFERASE
HCL	HYDROCHLORIC ACID
IVET	<i>IN VIVO</i> EXPRESSES TECHNOLOGY
KJ	KILOJOULE
LAPT	LINEAR AMPLIFICATION OF PROKARYOTIC TRANSCRIPTS
LB	LURIA BROTH
LD50	50% LETHAL DOSE
LEE	LINEAR EXPRESSION ELEMENT
MACS	MAGNETIC ASSISTED CELL SORTING
MBP	MALTOSE BINDING PROTEIN
MHC	MAJOR HISTOCOMPATIBILITY COMPLEX

ML	MILLI LITTER
MM	MILLI MOLAR
MMLV	MOLONEY MURINE LUEKEMIA VIRUS
MUM	MINIMAL UNIQUE MATCH
NAOH	SODIUM HYDROXIDE
NCBI	NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION
NT GDNA	NICK TRANSLATED GENOMIC DEOXYRIBONUCLEIC ACID
O.N.	OVER NIGHT
OD	OPTICAL DENSITY
OH ₂	WATER
ORF	OPEN READING FRAME
OVA	OVALBUMIN
PBS	PHOSPHATE BUFFERED SALINE
PCR	POLYMERASE CHAIN REACTION
PI	PATHOGENICITY ISLAND
POLY(A)	POLYADENYLATED
QRT-PCR	QUANTITATIVE POLYMERASE CHAIN REACTION
R.T.	ROOM TEMPERATURE
RACE	RAPID AMPLIFICATION OF CDNA ENDS
RNA	RIBONECLEIC ACID
RRNA	RIBOSOMAL RIBONUCLEIC ACID
RT	REVERSE TRANSCRIPTION

RT-PCR	REVERSE TRANSCRIPTION POLYMERASE CHAIN REACTION
RT-TS	REVERSE TRANSCRIPTION TEMPLATE SWITCHING
SDS	SODIUM DODECYL SULFATE
SELDI	SURFACE ENHANCED LASER DESORPTION IONIZATION
SSC	SALINE SODIUM CITRATE
STM	SIGNATURE TAGGED MUTAGENESIS
TBST	TRIS-BUFFERED SALINE TWEEN-20
TCR	T-CELL RECEPTOR
TE	TRIS ETHYLENEDIAMINETETRAACETIC ACID
U	UNIT
UB	UBIQUITIN
μG	MICROGRAM
μL	MICROLITER
μM	MICROMOLAR

CHAPTER ONE

BACKGROUND AND LITERATURE REVIEW

Introduction

The plague bacillus is a gram-negative coccobacillus of the *Enterobacteriaceae* family (134). *Yersinia pestis* has played a role in shaping who we are today. This disease has impacted and molded our world both negatively through its human toll but also positively through its influence on science and medicine. As a naturally occurring infection, the plague demonstrates its great flexibility as an adaptive microorganism, however its use as a biothreat agent constitutes the majority of its current concern. Despite the availability of antimicrobial therapies, there is significant progress to be made in prophylaxis and new tools to fight this disease. Furthermore, the majority of previous studies have not studied plague pathogenesis following pulmonary infection. Efforts to better understand pathogenesis are often guided by transcriptional profiling, however, a common hindrance to global transcriptional profiling has been the inability to evaluate the pathogen in the natural environment of the host. There are many technical barriers to approaching these types of studies. My efforts have been toward the development and application of global expression tools to enable the study of *Yersinia pestis* during early periods following pulmonary infection.

***Yersinia pestis* background and Historical perspective**

As the molecular and epidemiological origins of the plague bacillus continue to be studied and debated today, it is important that the historical impact of this pathogen not be lost. The plague, or “black death”, has played a significant role in the shaping of countries, wars, and populations (134). Even as questions arise as to a definitive role of *Yersinia pestis* as the causative agent of the pandemics attributed to it, the absolute potential to create a significant public health problem today justifies continued scientific investigations (99, 165). The present fear of *Yersinia pestis* as a bioterrorism agent is validated by not only what is known currently about its pathogenesis, but also by the global recognition of this pathogen. The plague’s historical role makes the fear inflicted by its potential use greater (87). Therefore, as these current questions and debates continue I feel it is important to present not only what is perceived historically, but also more importantly to present what is attributed to the plague bacillus, *Yersinia pestis*.

Historically, plague had been credited with over 200 million deaths, which makes it the most lethal infectious disease man has known (40, 51). Only recently, with influenza and the HIV pandemic, has plague’s historical role as most lethal pathogen been challenged. Even as these diseases continue to kill tremendous numbers of victims, it does not yet communicate the visceral fear that plague continues to today. This was evidenced as recently as 1994 during an epidemic outbreak of the disease where fear caused a mass exodus of over half a million people from Surat, India (142). The history of the plague appears to never leave the public consciousness.

Classically, the history of the plague is divided into three pandemics (134)(Figure 1). Where there are references to plagues that may be caused by *Yersinia pestis* scattered in ancient texts, I will focus on the historical impact of the three classical pandemics. The first pandemic, known as the Justinian plague, occurred in the middle of the 6th century and consisted of 11 known epidemics. It is believed to have originated in Ethiopia and to have spread quickly through the Mediterranean basin and the Middle East. After the first epidemic spread, subsequent epidemics occurred in 8-12 year cycles and it is unknown what contributed to this periodicity. It is impossible to ascribe numbers to the deaths caused by the plague during this pandemic, however, it is estimated that during the 6th century the population declined 50-60% and plague is credited with a major role in this depopulation (69).

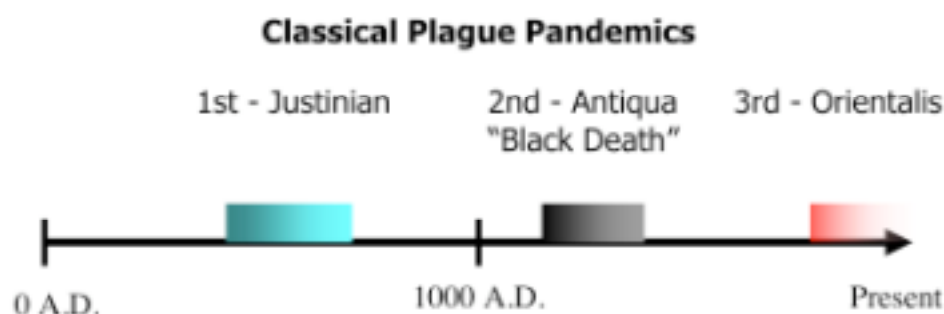


Figure 1. Timeline of Classical Plague Pandemics.

Pandemic outbreaks of the plague are divided into three time-periods, and within each multiple epidemics occurred. The third pandemic is considered ongoing although in decline.

After an approximately 600-year absence of epidemic plague, the second pandemic began in the mid 14th century arriving from Asia via trade routes. This

epidemic would spread through the known world and would be known as the infamous “black death”. The massive mortality incurred during this first epidemic would claim 30-40% of the European population, with subsequent epidemics claiming 5-15% of the population (69). Interestingly, the periodicity of the subsequent epidemics was reduced during this pandemic to 2-5 years. This periodicity lasted until the middle of the 15th century, where epidemics continued into the 17th century and were less frequent. The massive depopulation during this time, of which plague was a primary contributor, dramatically shaped all aspects of its history from political to scientific systems. Importantly, the scientific and medical advancements in this era spurred on by the plague helped to usher in modern medicine and scientific study.

The third pandemic started in the middle of the 19th century and continues today, although it is in decline (126). Whereas this pandemic has affected all populated continents, it had major impacts on Asia and India. India is estimated to have lost over 12 million people to the plague during a 20-year period starting in 1898. Importantly, it was during this pandemic, specifically the Hong Kong epidemic in 1894, that the plague bacillus was isolated (32). Subsequently, in 1897, the flea was described as an important vector for this disease (32). These and other discoveries quickly lead to a greater understanding of the epidemiology of plague and public health measures to combat the deadly epidemics. This has helped to reduce the geographical size and number of people affected by epidemics and more recently the discovery of antibiotics has further reduced their affect. Currently, plague has generally been isolated to small foci around the world and the threat of large natural epidemics has largely been

removed from developed countries. This has not necessarily alleviated the fear of the disease, mentioned earlier when as recently as 1994 fear of plague cause mass panic in India (142). What has been learned in the last century about the plague as a zoonotic disease propels us forward as we begin to approach plague as a biological weapon and not only a devastating natural disease.

General Biology

Plague is a zoonotic disease

Although *Yersinia pestis* may be now considered a greater threat as a biological weapon, its life as a zoonotic disease is important to study. Studying the zoonotic life of plague continues to lead us toward a better understanding of its pathogenesis and will be important to understanding environmental controls needed if this pathogen is ever used as a weapon.

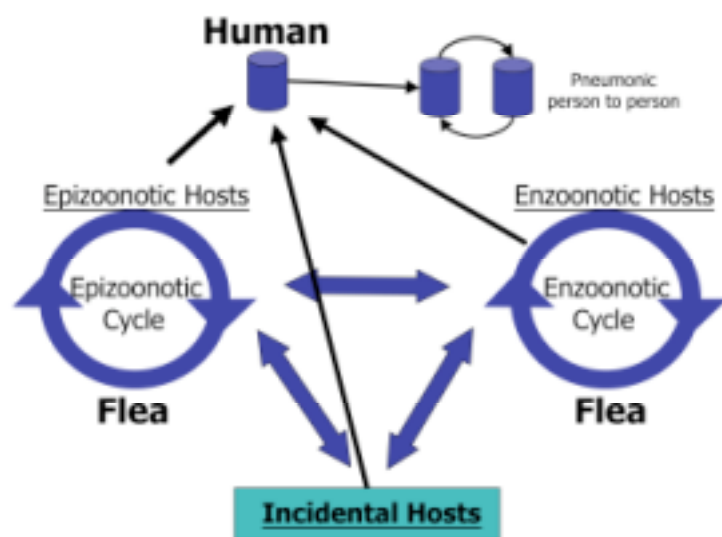


Figure 2 Zoonotic life cycle of *Yersinia pestis*.

The diversity of potential hosts is demonstrated by the potential cycling of *Y. pestis* through multiple host pools, primarily rodent species but can include larger mammals such as domestic cats. All hosts have the potential to transmit the bacteria to humans.

Plague, as a zoonotic disease, cycles through fleas and rodent hosts (60)(Figure 2). It is generally accepted that this basic enzootic cycle is what maintains the pathogen in the environment. Within this cycle, there is little observed host mortality. Outside of this cycle is believed to be another less frequent cycle between more susceptible hosts termed an epizootic cycle. Within this cycle, the rodent hosts are generally more susceptible to *Yersinia pestis* and serve as a means to amplify the spread of disease. From both of these zoonotic cycles, incidental hosts may be exposed to *Yersinia pestis*. Humans can enter this cycle through exposure to enzootic, incidental or epizootic host contact, an example of each being the flea itself, and incidentally infected cat, or a prairie dog species respectively (108). Interestingly, in the United States, the rate of plague infection is greater in women and children and may be related

to the fact that they more frequently contact household cats that have become incidental hosts for the plague (59). Although this concept of multiple enzoonotic and epizoonotic cycles appears to explain the maintenance and initiation of plague epidemics there is evidence that has lead to alternative and supporting hypotheses (60).

The primary contributing factor to maintenance and spread of zoonotic plague may be other environmental factors that affect not only the zoonotic hosts of plague but also the plague bacillus itself. Climatic changes including precipitation can obviously affect the environment for the propagation and spread of hosts susceptible to the plague (33, 131). This has lead to a hypothesis about a threshold level of hosts required for transmission to incidental hosts. Other studies have proposed temperature as an important factor in plague transmission as lower temperatures lead to increased flea survival and possible transmission (54). Temperature can also influence the physiology of the plague bacteria themselves, which may have influences on their survival within zoonotic cycles and within incidental hosts (134). It has even been proposed that plague may survive in the soil either within flea feces, dead animals, soil protozoan species, or even in a latent state outside of a host (50). All of these states can be influence by climate and other environmental conditions.

In the genomic age one must also consider environmental factors that influence the genomic content of the plague bacillus. The sequence of the plague genome revealed a highly dynamic structure with large numbers of mobile genetic elements as well as genetic structures that contribute to rearrangement (130). These elements may play a vital role in adjusting the pathogenic “rheostat” over time that influence all

aspects of the zoonotic cycle. An interesting idea proposed for *Vibrio* species is that the burden of pathogenic genetic elements not required in zoonotic hosts may be “distributed” throughout a population of bacteria and only rarely be brought back together to subsequently infect alternate hosts (personal communication Bill Costerton).

The plague has been well established as a zoonotic disease and this knowledge has lead to measures to reduce transmission as well as stem epidemics in progress. The number of environmental factor that contribute to the dynamics of this cycle, or even the extent and diversity of the cycle itself, is not completely understood. Better understanding of plague’s zoonotic life will allow better models to predict future epidemics and ways to reduce their impacts. This will lead us to better understand the steps and measures to address non-natural releases of plague into the environment.

Clinical Presentations of Plague in Humans

Although clinical cases of the plague are rare in the United States, they do occur (<http://www.cnn.com/2006/HEALTH/04/19/bubonic.plague.ap/index.html> link describes a reported case in L.A in April of 2006). Plague can have one of three primary presentations in human cases, bubonic, septicemic, or pneumonic. None of these are mutually exclusive and if left untreated the disease will progress to include features of all of these forms. Fortunately in the U.S. plague is lethal only 10-20% of the time due to successful intervention with antibiotics (45); however, it is also rarely diagnosed early in non-endemic areas. This is due to symptoms commonly shared with

other bacterial infections and it infrequently makes it onto a differential diagnosis as a cause of patient symptoms in non-endemic regions. The mortality seen in the U.S. is frequently related to patients not seeking treatment early enough, making patient education not only diagnosis and treatment important.

Bubonic plague is considered the classical presentation of the disease and is associated with *Yersinia pestis* infecting the lymphatic system. This often leads to swelling of the lymph nodes resulting in the bubo from which the name is derived. Clinically, the patient presents with fever, headache, and frequently gastrointestinal symptoms 2-6 days after exposure (134). The distinguishing clinical feature is the sensitivity of the lymph nodes to palpitation and frequently the swelling of these organs. The bacteria often enter the vascular circulation and become septicemic and/or pneumonic (61). The bubonic form of the disease provides the greatest amount of time for successful antibiotic intervention.

Primary septicemic plague is defined as bacteremia in the absence of presenting lymph node sensitivity (86). Although less common, septicemic plague is often more lethal with mortality rates as high as 50%(46). The increased mortality can be attributed to both the time to intervention but more frequently to the selection of non-effective antibiotic for treatment prior to accurate diagnosis.

The most infrequent presentation of the plague is the pneumonic form and it is also the most lethal. In the United States primary pneumonic plague can be contracted from household animals such as cats, and although over 10% of plague patients develop secondary pneumonia the last case of human to human transmission in the U.S. was >80

years ago (120). Factors contributing to the primary pneumonic forms high lethality are the associated short incubation time (1-3 days) and the overwhelming pneumonia it causes. Whether a primary or secondary infection, the pneumonia cause by *Yersinia pestis* is associated with a poorer prognosis.

The three primary presentations of this disease are not exclusive as mentioned above. Along with clinical tests, a thorough patient history may be the best tool for getting patients the appropriate treatment in time. The observation of a high mortality rate associated with the pneumonic form of the disease, while clinically the most infrequent presentation, is also the goal of a malicious attack. This presents a tremendous problem for clinical personnel in the event of such an attack.

Clinical Aspects of *Yersinia pestis*

Diagnosis of *Yersinia pestis* infections

Diagnosis of *Yersinia pestis* begins with the patient's history of potential exposure to the pathogen. Presently there are few regions in the United States, such as the four corners area where *Yersinia pestis* remains endemic, that maintain a heightened sense of awareness of the disease. In the absence of the characteristic buboes or lymph node sensitivity, characteristic of *Yersinia pestis* infection, it is difficult to distinguish plague infection from other gram-negative pathogens, and is generally low on the list of

agents responsible for the patient's symptoms. This fact, delaying appropriate therapy, contributes greatly to the increased mortality of septicemic and pneumonic plague.

Presumptive diagnosis is normally achieved in clinical setting through bacteriological tests, either through blood cultures and/or direct immuno-detection of the plague F1 capsular antigen. The problems with both of these approaches are numerous and can delay therapy and isolation through false negative results. Biochemical tests require cultures of the bacteria, which in the case of *Yersinia pestis* can take 2 days due to its slow growth rate (165). Furthermore the automated clinical culture identification equipment commonly used today to run these biochemical tests might not be set to identify *Yersinia pestis*. The direct immuno-fluorescence test using anti-F1 capsular antigen antibodies is used in many local and regional health departments but can provide false negative results due to low F1 expression at temperatures below 37°C, and can be lost in clinical settings because of sample storage at lower temperatures (46). Furthermore the F1 capsular antigen is not required for pathogenesis in many models of the disease and may not be present (187). In the U.S. the CDC continues to use phage lysis to definitively identify *Yersinia pestis*, but this resource is not practical in standard clinical microbiology laboratories (13).

Although not currently used in the clinic, many other technologies have been used to identify the presence of *Yersinia pestis*. These include ELISA, PCR-based, and hybridization technologies (11, 68, 177). Additionally, variable-number of tandem repeats (VNTR) has been used as an epidemiological tool to identify the source of *Yersinia pestis* strains (108). It is highly likely that many of these technologies have

been employed for environmental surveillance purposes but none have been approved for clinical use.

Therapy against the plague

Yersinia pestis remains sensitive to many antibiotics and this constitutes the primary means of treatment for all forms of plague infection outside of supportive therapy (27). Of the effective antibiotics streptomycin is the most commonly used for initial therapy, although its bactericidal activity requires careful attention against endotoxic shock. Other antibiotics such as tetracyclins and chloramphenicols have been used effectively along with newer generation antibiotics. The continued sensitivity of *Yersinia pestis* to antibiotics makes the plague a treatable and containable infection.

Immunity to plague

The immune system is broadly divided into innate and adaptive systems with the adaptive system further classified into cellular and humoral responses. Although, when studied these systems are often compartmentalized, it is recognized that these systems frequently interact with each other. To date, both adaptive and innate immune responses have been shown to be involved in host defenses against the plague. Recently the roles of the adaptive responses to plague are beginning to be understood.

The innate systems focus on common pathogen motifs, such as LPS. *Yersinia pestis* has demonstrated that it can alter its LPS to subvert host innate defenses (97).

The modifications appear to be modulated in part due to temperature-regulated factors and result in varying susceptibility to innate defenses such as the cationic anti-microbial peptide, polymyxin B (10). The innate system also involves local and systemic signaling, via cytokines, to prime and alert additional innate and adaptive systems. The *Yersinia pestis* virulence factor LcrV appears to be largely responsible for the general attenuation of pro-inflammatory signaling through systemic Interleukin-10 up-regulation (30). The targeting of *Yersinia pestis* to the immune cells responsible for amplification of pro-inflammatory signals further attenuates the innate defense (116). The ability of Yop proteins H and J/P to ultimately down-regulate NF- κ B to suppress pro-inflammatory signals, is a localized effect relative to more systemic effects of LcrV (25, 26). YopM has been shown to interfere with the innate immune system through targeted depletion of NK cells (95). Whereas, the innate system is available to assist in the immune response to plague, *Yersinia pestis* has also developed multiple mechanisms to either subvert or attenuate its effectiveness.

The adaptive responses, broadly divided into humoral and cellular arms, have both been shown to play a role in protection against *Yersinia pestis* infection. By far the most extensively studied is the humoral response to plague antigens. Many reports have described a correlation with antibody titers and protection, as well passive protection to the plague (7, 67, 90, 184). The antibody-mediated protection appears to be effected through opsonization of the bacteria for subsequent phagocytosis. Recently, the role of CD4 T-cells has been explored in both the support of the humoral response and their direct secretion of type-1 cytokines (129). Another study demonstrated that protection

could be mediated through T-cells alone in the absence of B-cells (128). Interestingly this study demonstrated a role for both CD4 and CD8 T-cells in protection, but the details of the mechanisms involved remain to be determined.

The frontline immune response to plague is the innate system, of which *Yersinia pestis* has demonstrated efficient mechanisms to avoid. This initial avoidance is generally all the time the pathogen requires to overpower the host. This makes the priming of an adaptive response for prophylaxis essential, and has been demonstrated with degrees of success. A better understanding of all the immune mechanisms involved in protection against this pathogen will lead to development of more effective immune prophylaxis toward the plague.

Prophylaxis against the plague

Prophylaxis against the plague has been achieved through vaccination and antibiotic therapy (165). Where antibiotics, primarily tetracycline, are used in cases of possible exposure, vaccines are used for individuals with potential exposure. There have been two clinically used vaccines against the plague; a live attenuated vaccine and a formalin killed formulation (119). The later was developed and used in the United States primarily for military purposes and has not changed since 1967. The effectiveness of the U.S. plague vaccine is evaluated based on the generated immunological response and indirect evidence from persons given the vaccine and exposed to plague prone environments. Most importantly it is questionable whether the

vaccine is effective against pneumonic plague as there is a report of a vaccinated individual acquiring pneumonic plague (39). Furthermore, animal models have shown various vaccine formulations to have only partial protection following aerosol challenges (67, 77, 156, 183).

The formalin-killed whole cell vaccines potential ineffectiveness against pulmonary challenge is not its only problem. As with all gram-negative whole cell vaccines the side effects from the endotoxin contained in the vaccine can range from mild to severe. Furthermore the antibody titers (F1 capsular antigen) are generally low and wane quickly requiring booster vaccinations every one to two years.

Much effort has been spent developing the F1 capsular antigen as a subunit vaccine. This has been reported in numerous studies using rodent models (9, 72, 154, 175). The primary challenge to its development will be its potential effectiveness against pulmonary infection, in which it has shown mixed results. The other challenge to its future use is the evidence that F1 capsular antigen is not required for virulence in mice as well as evidence that *Yersinia pestis* can down-regulate its F1 expression (57). The development of this vaccine may provide a level of comfort for the population that uses it, however, the primary challenge with all subunit vaccines is that they only require the pathogen to avoid a single immunological hurdle. In the case of the F1 subunit vaccine, *Yersinia pestis* has already demonstrated that this hurdle may not be difficult to overcome.

Most current vaccine studies have taken this into account and use multiple antigens, namely the addition of the LcrV gene product (V antigen) (55, 66, 67, 147,

151, 156). The use of both F1 and V antigens in immunization generally produces an additive protective effect. In a recent study the V antigen alone showed greater long term survival compared to F1 alone or F1-V fusion vaccinated animals, further highlighting that the F1 capsular antigen may be a poor target for immunization (156). It was not demonstrated if this difference in protection was due to levels of immune responses to these antigens or adaptation of *Yersinia pestis* to the immunological pressure.

There have been numerous other effective vaccines reported for *Yersinia pestis* in the scientific literature. These span the breadth of current vaccine platforms from genetic immunization to replication deficient bacterial vectors (21, 31, 55, 102, 150, 185). Many other antigens have been tested and shown to have varying degrees of protection against plague infection (20, 56, 101, 175). Of the antigens used they are almost exclusively associated with virulence effectors such as the Yop proteins or the virulence factor LcrV.

The focus of future prophylaxis to the plague remains on the generation of an effective vaccine. Of interest was a recent report using aerosolized antibodies to provide a protective response to pneumonic plague suggesting that prophylaxis may also be obtained through direct administration of antibodies to the site of infection (80). With the combination of antibiotic therapy, direct immune therapy and the development of long-term immune responses the current status of prophylaxis against the plague is promising.

Plague pathogenesis

Animal Models of Plague Infection

Taking advantage of the observation that rodents play a vital role of plagues zoonotic lifecycle, rodent animals models are predominantly employed to study this disease. Mice, rats, and guinea pigs have been most used in scientific studies with mice being the most prominent model today. All three are susceptible to *Yersinia pestis* and result in fatal disease when infected, although rats show a varying range of susceptibility that may be a consequence of their more prominent role in natural infection cycles (160). Unlike rats and guinea pigs, mice do not demonstrate the characteristic swollen lymph nodes of bubonic plague although the lymph nodes are infected. Most prominently bubonic plague, or those infected subcutaneously, have been studied in these models and only recently have pulmonary models regained interest, as this is one route of malicious release.

Pathogenesis in the Flea

The pathogenesis of the plague in the flea is vital to the maintenance of its zoonotic lifecycle. The dynamic differences in environment from the flea to a mammalian host demonstrates the rapid adaptability of the plague bacillus and highlights the ability of *Yersinia pestis* to coordinate gene expression relative to its

environment (134, 135). When infecting a flea, presumably from a blood meal from an infected animal, the bacteria take resident in the flea gut. To survive in this environment the bacteria secrete phospholipase D to prevent lysis (83). Furthermore, *Yersinia pestis* must also acquire and regulate iron uptake in this environment which it accomplishes utilizing a chromosomally encoded hemin storage locus *hms* (89, 96, 133). It is after colonization in the gut that *Yersinia pestis* begins to form aggregates in the gut and midgut and eventually forming a biofilm structure that extends into the proventriculus (82, 88). It is in the proventriculus where both the physical blockage of nutrients to the rest of the digestive system as well as the prevention of efficient blood cell lysis essentially starves the flea. This results in a more aggressive feeding behavior, and ultimately concludes in the death of the flea. Prior to this, transmission of the pathogen to a new host often occurs through the regurgitation of the biofilm mass and pathogen during an attempted feeding.

Pathogenesis in mammalian hosts

The pathogenesis of bubonic plague in mammalian hosts has been studied in rodent animal models and inferred from human cases. Upon inoculation via the feeding flea the organism travels into the lymphatic system via immune cells namely dendritic cells, macrophages, and neutrophils (116). The organism remains localized to the lymphatic system where it travels to the draining lymph node. It is here that the bacteria multiply, often causing the primary swollen lymph node (bubo), and

subsequently disseminate throughout the lymphatic system. They eventually reach the blood stream through the thoracic duct and subsequent spreading of the organism to the spleen, liver, and secondary lymph nodes via the vascular system. This results in the rapid overwhelming of the hosts defenses and death rapidly ensues by endotoxic shock, septicemia, and loss of vascular integrity (112, 160). During the septicemic spread of the organism the lung can become infected resulting in secondary pneumonia.

Pulmonary plague pathogenesis

The progression of plague introduced to the lungs can vary based on multiple factors including the growth temperature of *Yersinia pestis*, and the genotype of the host. Where these factors generally affect the infectious dose and replication rates of *Yersinia pestis*, the overall result and progression is conserved. Generally within the first 24-36 hrs post-infection there is a distinct lack of inflammation in the lungs. With the architecture generally intact there are few infiltrating immune cells, normal low levels of pro-inflammatory cytokines, and a general absence of any inflammatory response. This suggests either an active suppression of this response by *Yersinia pestis*, or a lack of recognition by the host. From 24-48 hrs some *Yersinia pestis* escapes the lung environment and become systemic. Concomitantly with this event an inflammatory state begins to develop within the lungs as bacteria return to the organ via the circulation. During this period the bacteria can be identified in all compartments of the lung tissue. The ability of *Yersinia pestis* to mask itself or suppress the immune system rapidly falls

away as neutrophils following a rapid increase in pro-inflammatory cytokines in the lung release defenses that are insufficient and counterproductive at this stage of the infection. It is unclear if the ultimate failure of the lung to adequately exchange oxygen is the primary reason for death as multiple other systemic insults contribute to disease progression.

Known Virulence factors

Many components of the *Yersinia pestis* genome contribute to its ability to cause disease in the mammalian host. Many of these are located on one of the three plasmids maintained by *Yersinia pestis* (130).

Plasmid pCD1

The 70 kb plasmid, pCD1, encodes a type III secretion system, which serves to deliver effector proteins into host cells (43, 135). Further, this plasmid encodes a low Ca^{++} response stimulon (LCR) that is thermally regulated by a transcriptional activator LcrF (190). Up-regulation of the LCR occurs at growth at 37°C or in a low Ca^{2+} environment, and results in transcription of Yop genes and the V antigen (LcrV) (49). These components are absolutely essential for virulence in a mouse model of *Yersinia pestis* (134). The Yop region encodes several proteins that can enhance virulence. For example, YopH and YopE are important for anti-phagocytic activity (44). YopE also down-regulates Rho GTPases and leads to actin microfilament disruption (103). LcrV is thought to act in Yop targeting at the extra-cellular surface.

Plasmid pMT1

The 100kb plasmid pMT1 of *Yersinia pestis* encodes for the fraction 1(F1) capsule-like antigen and the murine toxin (105). The production of F1 capsular antigen results in a large gel-like capsule around the organism. In contrast to YopE and YopH, which disrupt phagocytosis activity, the F1 appears to result in decreased adhesion of the *Yersinia pestis* to the macrophages (44).

The murine toxin encodes a protein with phospholipase D activity (153). This activity does not appear to be required for mammalian pathogenesis (84), and has also shown varied toxicity among vertebrates. This is in contrast to its requirement in the flea vector (83).

Plasmid pPCP

The smallest of the three plasmids at ~9.5-kb encodes the *pla* gene responsible for production of an outer-membrane protein. This protein constitutes multiple functions including plasminogen activation, C3 cleavage, and collagen binding. It is thought that these functions increase virulence of *Yersinia pestis* in the mammalian host though the presence of the protein is not required for pathogenicity.

Chromosomally encoded virulence factors

The chromosome encodes additional known virulence factors though less studied than many of the plasmid encoded virulence factors. Included in these are pilin-like genes (also known as the pH 6 antigen), and genes that contribute to serum resistance (134). Though not fully defined as virulence factors, a putative additional type three secretion system with no known effectors is present on the chromosome, as well as multiple iron uptake systems (130). Furthermore, a signature-tagged mutagenesis screen of 300 mutants in *Yersinia pestis* within a mouse sub-cutaneous infection model identified 16 putative-virulence factors of which 14 were located on the chromosome (56). Interestingly all 14 of these genes were related to hypothetical genes or global physiology. This study highlights the limitations of STM as less than 10% of the genes were screened in this assay. However this study reaffirms the need to investigate chromosomal gene expression during infection as the majority of the putative virulence-factors identified were located on the chromosome and did not have a previously described role in virulence.

Targeting of immune cells during infection

An aspect of plague pathogenesis is its apparent ability to target immune cells during infection (95, 112, 116). The dual threat of virulence mechanisms targeted to disrupt cellular function combined with preferential infection of immune cells acts as a two pronged offense on the host immune system. The ability to avoid host immune

attacks combined with destruction of cells responsible for potential protective functions makes this targeting an efficient mechanism to increase the plague's pathogenicity.

Plague as a biothreat agent

The history and introduction to plague above provides an outline for the attractive nature of *Yersinia pestis* as a biothreat agent. This potential of plague as a weapon has not been unnoticed and has historically been employed as a weapon as early as the second pandemic. Until president Nixon signed the treaty banning biological weapons in 1972 the U.S. had *Yersinia pestis* in its biological weapons programs (38). There is also strong evidence that the former Soviet Union continued developing *Yersinia pestis* as a biological weapon long past the bioweapons treaty (2). The present and future threat of *Yersinia pestis* as a biological weapon can be summarized by its reputation, pathogenesis, and gaps in detection and therapy.

As highlighted as recently as 1994, the fear of plague as a natural disease has not waned with history. Furthermore, this reputation only enhances the fear and panic potentially incurred with its use as a biothreat agent. The anthrax attacks of 2001 demonstrated that even a relatively small-scale attack could generate pervasive paralytic fear in the United States. If the goal of a biothreat attack is not only to kill citizens/soldiers, but also to generate a disruptive environment, the anthrax attacks of 2001 highlighted that this is certainly possible. The social moniker of plague as an unknown, awful and deadly disease would only allow it to parallel anthrax in its

effectiveness. As is the case with *Bacillus anthracis*, natural infections do occur in the U.S., however, any number of cases or cases outside the normal geographical distribution of the disease could easily be perceived as an attack.

Plague is frequently lethal and would be presumed to have a high mortality rate until prophylactic therapy was initiated. Depending on the route of infection (most likely aerosol) and the size of the attack, fatalities would be expected. Although pneumonia (bacterial and viral) is responsible for thousands of deaths each year in this country it is generally restricted to the elderly or immunocompromized population, the unbiased manner of plagues virulence in the population would further the reach of its effectiveness.

Finally, as was the case with the anthrax attacks of 2001, the absence of biothreat agents from the consciousness of mainstream medicine afforded more time for the attack to occur. Identifying the attack event requires first identifying the pathogen. As mentioned the rarity and lack of rapid resources to quickly identify *Yersinia pestis* at the front lines of an attack, medical facilities, makes the time to isolation and prophylaxis longer.

The future concerns of plague as a biothreat agent continue with its reputation, lethality, and long diagnosis time, but are magnified going forward by the potential that biotechnology has afforded this pathogen. Natural resistance to antibiotics has been described in *Yersinia sp.*; furthermore, non-natural resistance has been achieved in laboratory strains of *Yersinia* species (62, 118, 186). The potential harm of an antibiotic resistant strain is obviously greater. Furthermore, when exposed to hospital

environments and communities where reservoirs of resistance genes are bountiful, the natural acquisition of resistance is not unthinkable (73). More concerning is the fact that *Yersinia sp.* is genetically tractable in the lab and resistance to frontline antibiotics can easily be engineered into the pathogen prior to an attack (62). Even the F1 capsular gene can be manipulated to be absent or decreased in expression to avoid a F1-based vaccine or potentially delay identification. Following this line of thought forward there are many other potential applications of genetic engineering that would be effective in either masking the identification of the pathogen or increasing the virulence of the plague. It is clear that the efforts and knowledge we have gained about combating the plague may be easily compromised by common lab technologies and a malicious mind.

***In vivo* expression technologies**

Background and Introduction

There is little argument that microbiology has provided critical insights and technologies that now support a foundation for much of modern science. In the ontogeny of this process, often prokaryotes have been taken out of their natural environmental context and evolved into systems amenable to scientific study. For the study of pathogenic organisms this is problematic as pathogenic or virulence factors are “lost” in the lab. Whereas, *in vitro* systems designed to mimic *in vivo* environmental conditions have led to tremendous understanding of pathogenic mechanisms, there is

still a desire to study pathogens in the context of their natural environment. *In vivo* technologies have been developed to accomplish these studies. Methods such as signature-tagged mutagenesis (STM), *in vivo* expression technology (IVET), or differential fluorescence induction (DFI) all provide insights into gene expression *in vivo*, but are limited by the experimental biases introduced (37). This in turn limits the breadth of questions that can be addressed by these methods. Global gene expression analysis by microarrays presents the best opportunity to analyze *in vivo* gene activity without introducing non-natural selection bias (94, 109).

Gene expression arrays allow the survey of a pathogen's complete transcriptome in a single experiment. Microarrays have rapidly become a standard method in science. Unfortunately, the power of this method has had limited application to *in vivo* expression analysis of pathogens, in part due to the amount of RNA required for microarray analysis and secondly to the presence of host RNA in sample preparations (115). Despite these limitations *in vivo* microarray studies of pathogens have been done in a limited number of systems. These studies, however, relied on non-natural *in vivo* growth chambers, extensive bacterial enrichment procedures or large numbers of animals to obtain enough RNA for microarray analysis. Improving and demonstrating new technologies that would make *in vivo* gene expression studies of pathogens more universal would be advantageous (109). Application of data obtained from *in vivo* expression profiling can be used to address numerous fundamental questions regarding prokaryotic biology and pathogenesis.

While individual gene expression from an *in vivo* sample can be done using QRT-PCR, it is limiting in the scope of questions that can be addressed (1). Other indirect methods developed to study gene expression *in vivo* on a larger scale including STM, IVET, or DFI, however, have some inherent weaknesses. STM is limited to the study of genes required *in vivo* but that are not also required for *in vitro* growth. In addition to difficulties with calibrating infectious dose and pool complexity, gene products trans-complemented by other STM clones may be missed. IVET and DFI both require a threshold level of promoter activity to identify *in vivo* responsive promoter elements. All three of these examples require *ex vivo* genetic manipulation of the pathogen, and results can be strongly biased depending on the temporal expression of genes *in vivo*. The direct monitoring of bacterial expression through microarrays would bypass many of these limitations.

The development of microarrays enabled global gene expression studies. Since their development the technology continues to be refined. There are two general platforms used for expression studies; short-oligonucleotide arrays synthesized on the glass substrate itself (affymetrix format), or probes that are deposited on the substrate after their synthesis (64). The later provides greater probe flexibility due to the off substrate synthesis and often utilizes PCR products, or long oligonucleotides. Both array platforms are well established and have been used to study bacterial expression (109).

Unique features of bacterial microarray studies

There are unique challenges of bacterial expression microarrays relative to eukaryotic studies. While both systems require quality RNA that is prone to degradation from ubiquitous RNAases, bacterial arrays can further be limited by the abundance of RNA, and its diversity (109). While the decreased diversity of the RNA found in bacteria relative to eukaryotic organisms may seem an advantage as less material is needed to obtain adequate signals, the stochastic nature of bacterial gene expression can increase the error in these measurements (53). It is important to view expression studies in the context of a population average rather than individual cells. Eukaryotic expression studies have been able to overcome this limitation through amplification techniques and laser-capture microscopy, but these have not been applied to microbial studies to date (19, 123, 180)

***In vivo* microarray studies**

Challenges to applying bacterial microarrays studies to *in vivo* derived samples are numerous (115). They include the amount of bacterial RNA that can be obtained from *in vivo* samples and the presence of host RNA in the preparations. Prior to Linear Amplification of Prokaryotic Transcripts (see chapter three) there has been limited efforts to address the abundance of bacterial RNA and these generally included the pooling of multiple samples or non-linear amplification techniques (3, 148, 171). The

presence of host RNA from *in vivo* samples has been address in multiple ways, from implanting non-natural growth chambers to exclude host cells, to extensive enrichment procedures, or intentionally biased priming of bacterial RNA (148, 170, 171). These approaches are not exclusive and are often used concurrently to perform *in vivo* expression studies. Even with these tools many studies remain elusive, primarily due to the amount of RNA required for a single microarray experiment.

CHAPTER TWO

MATERIALS AND METHODS

Introduction

The Center for Biomedical Inventions was set up to identify technological barriers in science and strive to develop creative technological solutions to allow science to further expand into these areas. These efforts were pursuant to the scientific challenge and included hardware, molecular techniques, software developments, and high-throughput assays. I enjoyed the intellectual challenges of participating in many of these projects in the center. The primary technological development that I contributed to the center was the development of the Linear Amplification of Prokaryotic Transcripts (LAPT), addressing a fundamental deficiency in the area of prokaryotic expression technologies. This will be discussed in detail in chapter 3. Before this project was undertaken a significant number of resources were created to enable this and subsequent studies. The primary resource utilized in the subsequent studies was a long-oligonucleotide (70mer) microarray specific for the plague bacillus, *Yersinia pestis*. The process leading to the development of this resource will be discussed in detail. Other efforts resulting in the development of resources, technologies, and computational tools will be discussed in detail within the appendices.

Microarray Resource and technology development

As mention previously the goal of my thesis work revolved around global expression studies in *Yersinia pestis*. As I initiated this work microarray technology was only five years old and had yet to become the “routine assay” that it considered today. More significantly, application of this technology had not reached the increasing number of microbial pathogens whose sequences were rapidly becoming available. Therefore, as I initiated this resource development process my goal was to make a process applicable to other pathogens. This process was later applied to another pathogen, *Bacillus anthracis*. However, I will focus specifically on the process as it applied to the development of the *Yersinia pestis* microarray.

General Design considerations

As described in chapter one there are two general platforms for expression microarrays, although new platforms have recently been developed, the bulk of the technologies can be characterized as mechanically spotted or on substrate synthesized (85). This presented a major decision point and required a thorough evaluation. One consideration was cost, as at the time the only available means to create an Affymetrix style microarray was to go through the company itself. This has since changed and will be mentioned in the functional annotation appendix (64), but at the time there was a significant investment required to have a short oligonucleotide array constructed. Also,

the stepwise yield of the synthesis process and the inability to purify the oligonucleotides on the substrate restricted the length of the oligonucleotides to under 28bp. This fact in itself was not bad, as there is an inverse correlation in sensitivity and specificity with the probe length. Meaning the shorter the oligonucleotide the more specific the probe/target interaction but the sensitivity is less than that of a longer probe. This is attributed to the melting temperature difference and ability to use temperature to balance these parameters. Where longer oligonucleotides lose some specificity based on their higher melting temperature, that makes them more prone to non-complementary binding, the higher melting temperature also allows for more sensitive detection of target sequences due to larger Gibbs free energy of binding. The higher T_m also allows the reaction to run at higher temperatures to compensate for the specificity to some degree. As my goal was to develop a microarray that I could use to perform expression studies from *in vivo* samples, I weighed the sensitivity greater than the specificity, as I did not know the amount of material I could expect from *in vivo* samples and further I would be able to evaluate the specificity empirically and manipulate the hybridization conditions to optimize this parameter.

With the decision to use long-oligonucleotides the question was then how long? Two factors contributed to this decision, first being a report evaluating the specificity and sensitivity of a variety of lengths of probes in microarray experiments and our ability to synthesize long oligonucleotides. Reports empirically demonstrated the inverse correlation between sensitivity and specificity in relation to probe length, however they employed static hybridization conditions in their study (91). They

concluded that probes of 50bp provided the best balance of specificity and sensitivity for their purposes. In house studies using the MerMade oligonucleotide synthesizer (143) demonstrated our ability to make long oligonucleotides up to 120bp in length (CBI unpublished). As specificity can be influence by hybridizations conditions I opted to synthesis 70mer probes in favor of their greater sensitivity.

Construction of *Y. pestis* Microarray

Summary

70bp probes were designed for each ORF of CO92 genome using the program ProbeSelect (104). To further validate the specificity of all designed probes, I ran each probe through BLAST using a mouse EST database to evaluate potential cross-hybridization to mouse sequences (182). Only 64 probes returned an E-score less than 0.01. The 70mer probes were either synthesized at the Center for Biomedical Inventions using a MerMade oligonucleotide synthesizer (143) (Bioautomation Plano, TX), or purchased (Intergraded DNA Technologies, Coralville, IA). Oligonucleotides synthesized in-house were purified over G-25 sephadex (Amersham Biosciences, Piscataway, NJ). Oligonucleotides were prepared for printing by aliquoting in 384-well plates at a final concentration of 40 μ M in 3X SSC. Printing of the microarrays was done on poly-L Lysine coated slides. Arraying was with a Qarray spotter, (Genetix, Hampshire, UK), with typical spot diameters of 100-160 μ m. Slides were steamed and

UV cross-linked (600kj) using Stratalinker, (Stratagene, La Jolla, CA), and stored for future use (see Figure 3).

Microarrays were validated for their specificity, sensitivity, and functional application. Specificity was determined by hybridization of labeled PCR products to the array. These products only hybridized to their cognate probe on the array. Additionally, labeled mouse RNA was hybridized to the array and little cross-hybridization was observed. Sensitivity was determined by titration experiments to empirically determine the minimal amount of labeled cDNA and genomic DNA standard needed to obtain reproducible data. Functionally, we have performed experiments evaluating the transcriptional changes induced upon exposure of *Yersinia pseudotuberculosis* to increased growth temperature (available at www.biodesign.edu). *Yersinia pestis* microarrays are available through the Western Regional Center of Excellence support (contact thru www.biodesign.edu).



Figure 3 Overall process of microarray construction.

From an annotated genome probes are designed for each ORF. Probes are either synthesized, purified, and quantitated in house, or purchased. They are then arranged in plates prior to mechanical spotting.

Probe design

When I initiated my probe design efforts there were no reports of methods to design long oligonucleotide probe on a genomic scale that took into consideration the specificity of the probe within the genome itself or against another genome. In efforts with Preston Hunter (Center for Biomedical Inventions) we explored the use of what we termed a tuple method where unique probes for each gene were identified through rankings of sequence similarity. The basic idea was to use the frequency of short 3-based pair nucleotides to rapidly identify regions within genes that were less similar to

other genes in the genome. The effectiveness of this method was not functionally tested as a report using a similar, but more powerful method was published (104). This report by Li et al used a suffix array method to identify unique sequences in each gene for probe design, then evaluated the estimated free energy of binding to the target sequence and finally used M-fold to evaluate secondary structure that may prohibit binding. This program, however, did not consider the binding of sequences outside of the target genome. In my case I wanted to preempt as best as possible the binding of mammalian sequences to my designed probes. To do this I used the BLAST program to identify the similarity of my designed probes to mouse RNA sequences using the *Mus musculus* EST database.

The sequence information for *Yersinia pestis* strain CO92 (Colorado 92, a strain isolated in Colorado) was downloaded from the following web page and links from this page

[http://www.ncbi.nlm.nih.gov:80/entrez/utis/qmap.cgi?db=c&form=4&field=ORGN&term=Yersinia+pestis\[ORGN\]](http://www.ncbi.nlm.nih.gov:80/entrez/utis/qmap.cgi?db=c&form=4&field=ORGN&term=Yersinia+pestis[ORGN])

The sequence and annotation for the *Y. pestis* chromosome, and plasmids pPCP1, pMT1, and pCD1 were used.

In order to manage these data all information was compiled into a FileMaker Pro database. This was achieved by first opening all downloaded data in Bbedit a text management software program to insure that the information was in a text tabbed delimited format that could be opened in Excel. The nucleotide and protein sequences that were provided in a fasta format had to be manipulated in this environment to remove unwanted return characters (\r) and insert a tab (\t) in between the gene name and sequence information.

For example;

```
>YPCD1.01 (t)      atggtcacttttgagacagttatggaaattaa.... (r)
```

```
>YPCD1.02
```

Once all of this information was organized in text tab delimited formats they were imported into excel files and organized. These files were then imported into FileMaker Pro databases. All further manipulation was done within these databases.

Location	Strand	Length	PID	Gene	Synonym	Product
271..711	-	147	16120354		YPO0001	putative flavoprotein
804..1265	-	154	16120355	asnC	YPO0002	regulatory protein
1435..2427	+	331	16120356	asnA	YPO0003	aspartate-ammonia ligase 2526..3992 - 489 16120357
					YPO0004	conserved hypothetical protein
3996..5549	-	518	16120358		YPO0005	conserved hypothetical protein
5823..7691	+	623	16120359	kup	YPO0006	potassium transport protein
7896..8315	+	140	16120360		YPO0007	ribose permease
8366..9292	+	309	16120361	rbsK	YPO0008	ribokinase

Table 1 Sample table of genomic information for *Y. pestis* .

Tables of additional information were also included. The synonym was used as the unique identifier and gene ID for all future work (YPO0001 etc).

The following is a summary of many trouble-shooting efforts to get the program to run efficiently.

Important Notes;

- The program has a problem designing long probes(>50 bp) for genomes with a GC content > 57%
- All text MUST be Capitalized for the program to work

- All redundant or highly similar genes must be input as a single gene. This was important for YP do to the large number of transposons and truncated gene products in the genome.

The program is run in Linux using a Solaris OS and the program required three input files, four input values and two output files.

Input file 1:

Sequence file- this file contained the sequence of the genome or transcriptome of the organism. It must be on 1 line without any header or unusual characters.

Input file 2:

Fasta file- this file is a fasta list of all the genes for which probes are to be made.

The file must be in the following format;

```
>YPO0001(\r) TTGACCGATGACCCCGGTTTCAGGCTTCACCACAGTGTGGAACG(\r)
```

```
>YPO0002(\r) CTAATCTCAGCGCTCCGCTGACCCCTCAGCAAAGGGCTTGGCT(\r)
```

Etc.

Note: this file must end with a ">" or else the program will not design a probe for the last gene.

Input file 3:

List file- this file contains a list of the genes and their position within the sequence file. Each gene is on 1 line with the start and stop positions separated by 2 spaces. An example is provided below.

YPO0001 1 1395

YPO0002 1396 2532

YPO0003 2533 2754

YPO0005 2755 3879

YPO0006 3880 5814

All of these files must be saved with UNIX line breaks as generic text files. All characters must also be in capital letters.

The input files were generated for *Y. pestis* and the three plasmids pPCP, pMT1, and pCD1. To do this the last 500 bp of each gene was used for probe design. If the gene was not greater than 500 bp the entire gene was used. This was done to bias the probe selection towards the 3' end of the mRNA transcript. The coding sequence file was generated by concatenating all the orfs 5' to 3' together, such that a single “transcriptome” sequence was generated (Figure 4). This was used as opposed to the entire genome because of the increased sensitivity of the probes for their target and the probability of finding a probe for each gene.

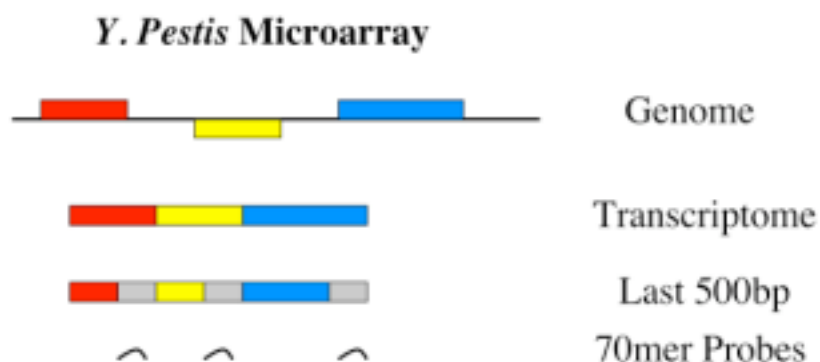


Figure 4 Graphical representation of Genomic data manipulation prior to probe design.

The predicted ORFs of the annotated genome were extracted and concatenated to create a virtual transcriptome. 70mer probes were designed within last 500bp of each ORF.

To further evaluate the quality of the designed probes the probes were run through BLAST using two different sequence databases. The first was a sequence database of all *Y. pestis* ORF's. The second was a sequence database of the Mouse RefSeq database obtained through the NCBI ftp web site.

All of these analyses were done on a PC using the stand-alone BLAST program from NCBI. Specifically the BLASTall program which allows all the probes to be run through the program sequentially. The results of these analyses are summarized in Table 2.

- Gene duplications or Genes with >95%nucleotide identity represented with a single probe (ISE,TNP,etc)
- No probe is expected to hybridize non-specifically based on free energy calculations
- BLAST results of probes against *Y. pestis* transcriptome resulted in no significant scores (E-values >0.001, no more than 17bp match)
- BLAST results of probes against Mouse EST DB resulted in no E scores >0.0006 or 27bp identity

Table 2 Summary of 70mer probe designed to *Y. pestis* genome.

Probe synthesis

Synthesis was performed using either a MerMade II or MerMade V synthesizer from Bioautomation (Plano, TX)(143). Procedures and protocols were run as optimized to a great extent by Bioautomation staff members and Xiang Chen. The processing of oligos is described and is summarized in Figure 3.

Quality of the synthesis was done two-fold by spot-checking the size of the oligo itself on a polyacrilamide/urea gel and all the oligos were evaluated based on their yield determined in the quantitation step (Figure 5). All of the handling of the oligos was done using robotic liquid transfer machines.

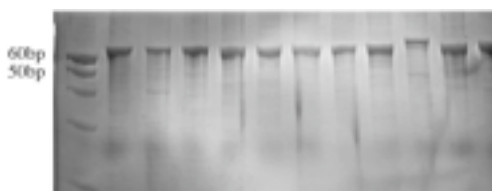


Figure 5 A typical quality control gel of 70mer oligos.

Random samples of synthesized 70mer probes were evaluated on TBE/Urea gels to inspect their quality. The majority of the synthesis product is full-length product with minimal laddering of truncated products.

Purification of Oligos

Purpose of this procedure is the removal of synthesis by-products from the oligos and to decrease the number of truncated synthesis products from the final suspension of product.

Materials used included:

- Sephadex G-25 Superfine resin (Amersham Bioscience 17-0572-02)
- Multiscreen column loader (Millipore MACL09600)
- Multiscreen HV 96 well plates (Millipore MAHVN4550)

Procedure:

1. Resuspend the oligos from the MerMade synthesizer in 150-200ul H₂O
2. Fill the 96-well column loader with G-25 sephadex making sure that all wells are filled and that any excess sephadex is removed.

3. Place the Multiscreen HV filter plate on top of the column loader plate. Invert the two plates such that the sephadex falls into the wells of the filter plate. You may need to tap the loader plate to get all the resin to transfer.
4. After the resin has been transferred add 200ul of H₂O to each well. This will cause the sephadex to swell.
5. Wait 5 min. and then add an additional 50-80ul of H₂O to each well.
6. Let the plate stand for 1 hr to allow the resin to completely swell.
7. Place the resin filled plate on top of an empty collection 96-well plate.
8. Spin down the plates at 2500rpm for 5 min.
9. Take the resin filled plate and place it on a new collection plate (I have used 96-well PCR plates from Applied Biosystems because they fit well together)
10. Add your unpurified re-suspended oligos to the resin filled plate.
11. Spin down the plates at 2500rpm for 5 min.
12. Recover the collection plate and discard the resin filled plate.

Note: I have used the Beckman robot to do the majority of the liquid transfers.

Concentration Determination of Oligos

Purpose of this procedure is to determine the concentration of oligos in a high-throughput fashion using a 96 well plate fluorimeter

Lambert Beer's Law $A = eCI$ where: A = absorbance

e = molar extinction coefficient ($M^{-1}cm^{-1}$)

C = concentration (Molar)

$I = \text{path length (cm)}$

Calculation used for Extinction Coefficient:

$$e = 0.89 * ((\#A's * 15480) + (\#C's * 7340) + (\#G's * 11760) + (\#T's * 8850))$$

Using Costar UV transparent 96-well plates with 200ul of fluid the path length is determined to be 0.52cm (diameter is 0.7cm)

$I = 0.52 \text{ cm}$

Absorbance is determined by subtracting the background from the sample reading. The background has been determined to be ~0.078. The oligo is also diluted 1:200 in water to keep the absorbance reading in the linear range (from 0.2 to 2 OD units).

Mechanical spotting

Mechanical spotting of the microarrays was done with either a Qarray spotter (Genetix) or a SpotArray 72 (Perkin Elmer). The final quality of the slides was similar on both machines. The only difference was in the flexibility of the Qarray software to allow non-adjacent replicate on the array as well as placement of positional makers and controls at the corners of each subarray. The pins employed in these spottings were either Telechem SMP3 pins or Point Technologies split pins. Both sets of pins yielded

spot diameters of 80-140 μM . The substrate employed in my studies is Poly-L-Lysine coated slides. The protocol used for making these was obtained from Pat Brown's web site at Stanford University. All oligos were spotted at a concentration of 40 μM in a 3X SSC buffer. Where there are many buffers used to spot oligos I used 3X SSC for a variety of reasons. Primary, was the prior experience in the lab with this buffer, but also the SSC buffer provides a more predictable spot diameter. In my experience there are many factors involved in the production of consistent quality slides, and changing the buffer is not only labor intensive but also requires a reevaluation of spotting parameter.

The technical considerations for the spotting of the oligos include the pinhead speed, dwell time, spotting time, and spotting over-travel (78). All of these parameters can be modified using the robotic spotter software. My approach to this process is to functionally test each parameter using dye colored buffer. Approaching each variable one at a time and evaluating changes in spot morphology after each modification. I have found that this process is essential whenever spotting buffer is changed, substrate has changed or general environmental conditions have been changed. In general moderate pinhead speeds prevent drying of the oligo during spotting, longer dwell and spotting times produce larger spots, and greater spotting over-travel also creates larger spots. All parameter can be balanced to provide the most consistent spotting across the spotting deck allowing for the greatest margin of error for the uncontrolled variables being the substrate thickness and deck levelness. Where most substrate is manufactured at 1mm there are often minor variations that can affect spotting if the parameters are set

to strictly, particularly spotting over-travel. In conclusion this is a process of functional evaluation.

Microarray validation

General quality

The overall quality of slides was evaluated by the hybridization of labeled analyte followed by scanning at appropriate wavelengths. The production of spots with diameters ranging from 80-120uM was assessed visually along with general slide consistency. A typical example is shown in Figure 6.

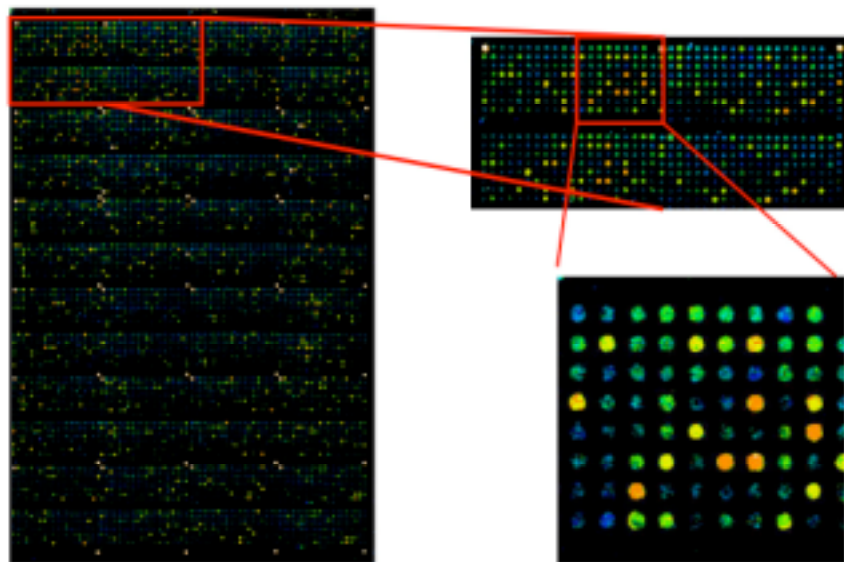


Figure 6 General Quality of mechanically spotted microarrays.

Increasing resolution of a spotted microarray hybridized with genomic DNA (blue) and cDNA (rainbow).

Specificity of *Y. pestis* array

The specificity of the *Y. pestis* array was evaluated by both the hybridization of specifically labeled PCR products of 14 genes, and hybridization of labeled cDNA from mouse RNA. The results of these experiments are summarized in Figure 7.

Specificity

- Labeled PCR gene products only hybridized to their respective probes on the array with no cross-hybridization demonstrating high probe specificity.

- Minimal cross hybridization to labeled mouse cDNA supports the specificity of the probes to *Yersinia sp.* sequences for future in vivo analyses. This cross hybridization was reduced using Genome Directed Primers.

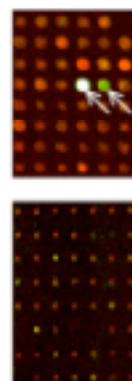


Figure 7 Specificity of *Y. pestis* microarrays

The specificity of the *Y. pestis* microarrays was evaluated both through hybridization of gene specific PCR products and non-specific mouse cDNA. Cy5 labeled gDNA (red) was also hybridized to the array as an internal control.

Reproducibility

An important measure of this resource was its experimental reproducibility.

Figure 8 shows a typical scatter plot of data obtained from 2 replicate experiments. The R-value is 0.94.

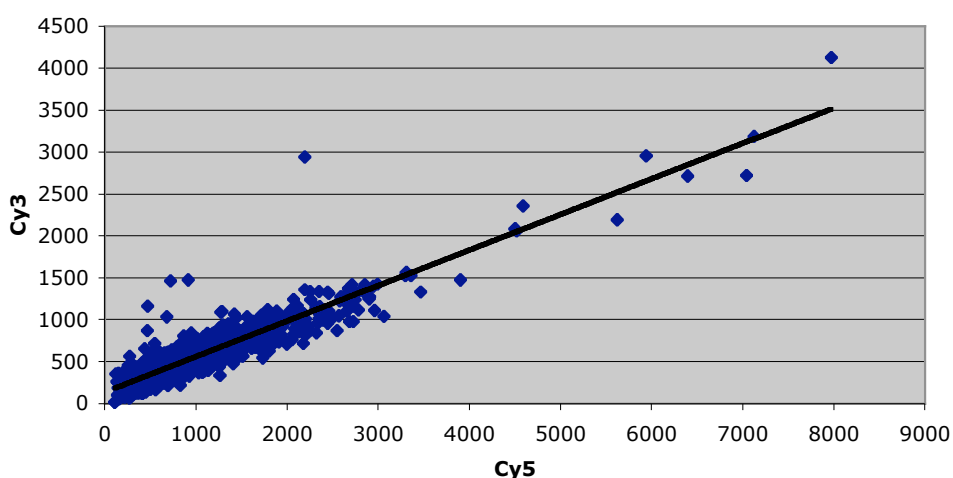


Figure 8 Microarray reproducibility.

The reproducibility of the microarrays was evaluated by comparing hybridization patterns of two arrays. The scale is expressed as genomic normalized intensities (cDNA signal/gDNA signal)

RNA isolation Protocols

Quality evaluation of RNA and cDNA.

Evaluation of RNA and cDNA was performed using the Bioanalyzer 2100, Agilent Technologies(Palo Alto, CA). The RNA 6000 Nano Assay was used for all

RNA analysis, following manufacturer protocols. All cDNA samples were evaluated using the DNA 7500 Assay, following manufacturer protocols.

RNAwiz or Trizol methods for RNA extraction

Harvest cells/ tissues etc per standard protocol

Use appropriate numbers of cells or mg of tissue as described in manufacturers protocol

1. Add 1 ml of Trizol or RNAwiz and homogenize by vortexing sample
2. Incubate sample at 50C for 3 min. and an additional 2 min. at R.T.
3. Add 200 ul chloroform and vortex for 20 sec.
4. Incubate sample at R.T. for 10 min.
5. Centrifuge Sample at >10k xg for 15min. at 4C
6. Prepare a new Rnase free tube by putting 350-500ul Nuclease free OH2
7. After sample is done spinning transfer aqueous phase to the new tube with OH2
8. Mix sample well then add 1ml of Isopropyl alcohol and again mix
9. Allow sample to precipitate for 10 min at R.T.
10. Centrifuge Sample at >10k xg for 15min. at 4C
11. Decant supernatant and add 1 ml 75% EtOH
12. Vortex then spin at >10k xg for 5 min.
13. Carefully remove sup. by aspiration
14. Add 20-50ul TE buffer with Rnasin to resuspend pellet

Analyze sample on Agilent 2100 bioanalyzer for evidence of RNA integrity. This is commonly evaluated by the ratio of the two dominant ribosomal RNA peaks and should be greater than 1.8 (16S/26S).

***Y. pestis* RNA from plate cultures**

1. Streak 4 blood agar plates with *Y. pestis* from frozen stock and grow to confluence at 27°C for 48 hrs
2. At 48 hours, take 2 plates and incubate at 37° C
3. 30 minutes later-process 1 plate from 27° C and 1 plate from 37°C incubators for RNA using Trizol
4. 2 hrs later - process the other plate from 27° C and the other plate from 37°C incubators for RNA using Trizol

Protocol for CFU count and Enrichment of *Yersinia pestis* from mouse lungs

1. Sacrifice animal per established protocols
2. Dissect out all lung tissue and place in dounce homogenizer on ICE
3. Add 500ul of PBS to tissues
4. Homogenize tissue by plunging pestle 15-20 times
5. Remove homogenized tissue with a pipette (use an additional 500ul PBS to wash homogenizer if necessary to collect all the sample.)
6. Using a pipetor or 2 ml pipette determine the volume of the sample
7. Set up a dilution series for plating. Starting with 20ul of sample in 80ul PBS plate 50 ul of this solution and continue two additional 1:10 serial dilutions further. Plate these solutions onto Nutrient agar plates.

8. With the remaining sample pass the solution through a plastic mesh (FACS mesh) into a new tube. This should remove any large debris from the homogenization step.
9. Centrifuge the sample at 1000xg for 5 min at 4C
10. Aspirate off supernatant and discard
11. Immediately add 1ml of trizol reagent to the pellet.
12. Continue with Trizol protocol.

For CFU calculations

Count the colonies on each plate of the dilution series. Only use plates that have counts between 5 and 200 colonies. Multiply the # of colonies by the dilution factor for that plate. Calculate the mg of tissue per ul by dividing total sample weight by the total volume of resuspended sample. Multiply this number by 10 (because 10ul were used in the initial dilution. Divide the number of colonies by this number (mg of homogenized tissue plated). This will give the CFU per mg of tissue. Multiplying this number by the total weight of the tissue will give the total bacterial burden for that animal.

Analyte Labeling Protocols

Generation of Genomic standards for Microarrays.

We employed genomic normalization to compare experiments (169). *Y. pestis* genomic DNA was nick translated and amino-allyl dUTP incorporated. This was done

in a 50 μ l volume containing 15 μ g genomic DNA, 0.05mM dGTP, dCTP, dATP, 0.03mM dTTP, and 0.02mM amino-allyl dUPT, 1X DNA polI buffer, 15U DNA PolI, 0.017U DNaseI, 20 μ g/ul BSA. The reaction was incubated at 15°C for 5-6hrs. Reactions were purified by column filtration and dried down, resuspended in 2X coupling buffer followed by conjugation with 0.1mg Alexa 488 (Molecular Probes). The concentration and dye incorporation was determined by spectrophotometry.

The protocol was modified to incorporate an amino-allyl modified base to improve dye incorporation and allow for flexibility in dye selection. Figure 9

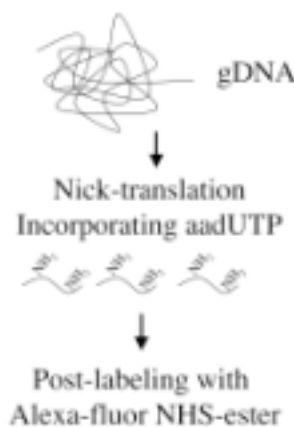


Figure 9 gDNA labeling using amino-allyl modified nucleotide.

Using an aadUTP in the nick-translation reaction allows higher dye incorporation rates and permitted post-labeling with any amine-reactive dye.

gDNA labeling using Nick Translation Protocol.

X	ul	ddH ₂ O to 50ul depending on gDNA concentration
5	ul	dGTPs, dCTP, dATP 0.5mM dTTP 0.3mM aadUTP 0.2mM
5	ul	Nick Translation Buffer (DNA Polymerase I, 10X buffer)

15 ug gDNA
 5 ul Nick Translation enzyme mix*

50 ul Total

1- Incubate at 15 C for 5-6 hrs to get short labeled fragments

2- Add 5 ul of stop buffer (use Dnase I stop buffer)

3- Clean up the probe and hybridize as before

Nick Translation enzyme mix: To each

20 ul DNA Pol I enzyme add
 1 ul Diluted Dnase I enzyme (Ambion)diluted 1:20 total 0.1U
 3 ul DNA Polymerase I, 10X buffer
 6 ul BSA (1 mg/ml)

30 total

Alexa 594	Alexa 488	ug NT gDNA	Actual ng
5ug Ypth cDNA	gDNA	1	54.4
5ug Ypth cDNA	gDNA	0.5	27.2
5ug Ypth cDNA	gDNA	0.25	13.6

Table 3 Typical yields from gDNA labeling reaction.

Using 0.5ug of starting material I typically recovered 25-30ng of labeled gDNA fragments with typical dye incorporation rates of 1 dye molecule every 10 nucleotides. This was sufficient to obtain genomic signal in microarray experiments.

Presently a commercial kit has been used for genomic labeling. The kit from Invitrogen, is called BioPrime Plus. The protocol uses Exo-Klenow to incorporate amino-allyl labeled nucleotides from randomly primed genomic DNA. The process can be improved by initial digestion of the gDNA with a restriction enzyme, however, with bacterial genomic DNA I have not found this to be necessary.

Generation of labeled cDNA

The decision tree for approaching RNA samples is diagrammed.

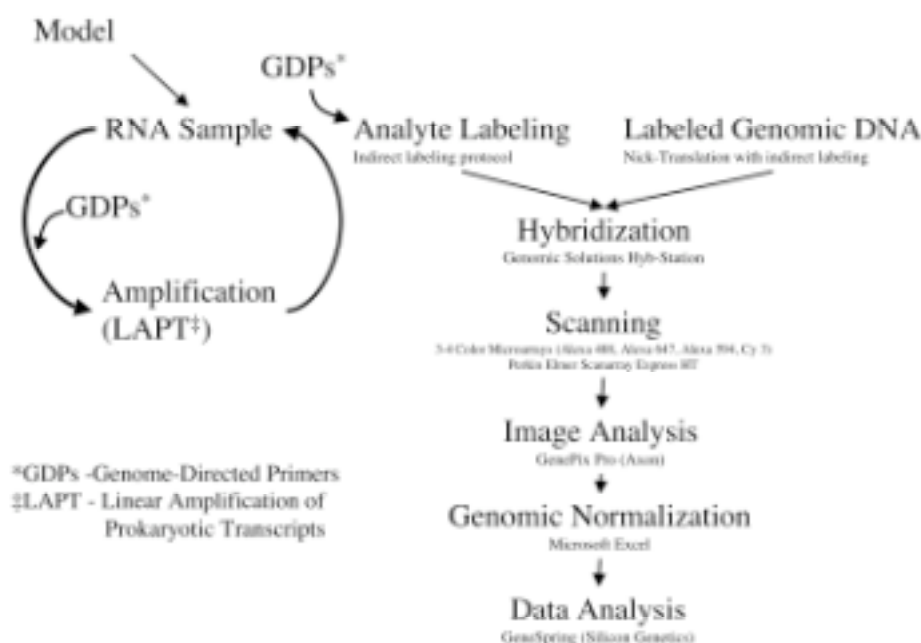


Figure 10 Microarray sample handling for RNA processing through data analysis.

Once RNA is obtained from the study model it is either directly labeled or amplified using LAPT then labeled.

Direct labeling of RNA

Reverse Transcription of Total RNA with Cy3 and Purification for Microarray

Hybridization

Reverse Transcription

1X

18 ul Total RNA (final 10ug)

6 ul Primers GDPs of hexamers @250ng/ul

Mix RNA and Primers for 10 minutes @ 70C

Snap cool on Ice (4C) for 5 minutes

Prepare the following Reaction Mix

8 ul 5X first Stand Buffer

4 ul 0.1M DTT

1 ul 10mM dNTP except dCTP is 2.5mM

1 ul Cy3 dCTP (Amersham)

2 ul SuperScriptII RT

40 ul Final Volume

Combine RNA primer Mix and Reaction Mix

Incubate @42-45C for 1 hr.

Add an additional 1 ul SuperScriptII RT to mix

Incubate an additional 1-2hr

Degradation of RNA

Add 10 ul of 1M NaOH to RT reaction

Incubate @ 65C for 10 minutes

Neutralize Mixture by adding 10 ul 1M HCl

Probe Clean-up

1. Prepare Microcon filter (YM-30) by adding 200ul TE buffer to filter

and spinning @12Kxg for 10 minutes

2. Apply sample to the filter and spin @12Kxg for 10 minutes

3. Wash with 100ul TE buffer and spin @12Kxg for 5 minutes

4. Add 10ul TE to sample (should see colored pellet on filter)

5. Invert filter and place in a clean tube

6. Spin down sample @3.5Kxg for 5 minutes

Preparation of Hybridization Solution

1. Prepare microcon filter as above

2. Combine samples to be hybridized together (i.e. NT gDNA and cDNA)

3. Bring the volume of the sample to 100ul with TE buffer

4. Apply to Microcon filter and spin @12Kxg for 10min

5. Wash filter with an additional 100ul TE buffer and spin as before

6. Add 7.5-10 ul TE buffer to sample (should see a purple pellet on filter)

7. Invert filter into a clean tube

8. Spin down sample @3.5Kxg for 5 minutes

Volume should be ~10ul

For Genomic solutions Hybridization Station add 100ul of

Slide Hyb #3 from Ambion

Indirect labeling of RNA

Modified Fair Play cDNA Labeling Protocol

I. Reverse Transcription

In an RNase free 1.5ml tube:

1. RNA sample, 10-20ug 23ul

2. Primers (GDP or hexamers 250ng/ul) 6ul

use water to bring vol. up if needed

Incubate at 70C for 10min., Chill on Ice

Add to each tube:

10X Stratascript buffer 4ul

20X aa-dNTP mix 1-2ul

0.1M DTT 3ul

RNase Block 40U/ul 1ul

Stratascript RT 1ul

Total vol. 40ul

Incubate at RT for 10 min.

Incubate at 46-48C for 1hr, then add 1 ul SuperScript RT (200U/ul)

Incubate for an additional 1 hr.

II. Degrade RNA

Add 1M NaOH 10ul

Incubate at 70C for 10 min.

Add 1M HCl (to neutralize solution) 10ul

III. cDNA precipitation

3M NaOAc, pH 4.5	8ul
Glycogen, 20ug/ul	2ul
Ice Cold EtOH 95%	200ul

Precipitate at -20C for 1hr or ON

Spin at Max speed at 4C for 15 min.

Decant EtOH

Wash pellet with 500ul 70% EtOH

Spin at Max speed at 4C for 15 min.

Air dry pellet or use Speed Vac (NO heat) for 10-15 min.

DO NOT over-dry pellet

IV. Dye coupling

Add

2X Coupling Buffer	5ul
--------------------	-----

Resuspend pellet by Vortex and/or heating reaction at 37C

Resuspend Monoreactive Cy dye (Amersham) in 45ul DMSO

Add

Resuspended Cy Dye	5ul
--------------------	-----

Allow coupling reaction to occur in the Dark for 30min.

V. cDNA purification (Qiagen PCR clean-up kits)

Add 90ul of H₂O to each tube for a final volume of 100ul

Add 500ul buffer PB

Apply solution to column and spin for 1 min.

Wash column with 750ul buffer PE

Spin for 1 min.

Place column into fresh tube and spin and additional 2 min.

Elute column with 50-80ul of H₂O

Generation of GFP RNA template.

A 1kb transcript encoding the GFP gene was transcribed from a linear fragment of plasmid pIVEX2.3d-GFP, Roche (Indianapolis, IN) using a T7- Megascript kit, Ambion (Austin, Texas) in 40ul reactions with an additional 400U of T7 polymerase for 16hrs. The product was treated with 4U of DNaseI for 30 minutes followed by DnaseI inactivation using *DNA-free*, Ambion (Austin, Texas). The RNA was purified using RNeasy protocol, Qiagen (Valencia, CA). RNA was evaluated using a Bioanalyzer 2100, Agilent (Palo Alto, CA), and quantified using a Nanodrop, Nanodrop Technologies, Inc. (Rockland, DE).

LAPT protocol for RNA amplification

Summary

Reverse transcription and template-switch primer addition was performed using a half reaction (50ul) of the Super SMART™ protocol, Clontech (Palo Alto, CA). The primers used for RT were either one of five gene-specific primers to GFP, Random

hexamers, Invitrogen (Carlsbad, California), or Genome-Directed primers designed towards *Yersinia pestis* CO92. The primers were annealed to the RNA by heating the RNA and primer(s) mix to 65°C for 5 minutes then cooling the reaction to 42°C prior to addition of the template-switching primer and reverse transcription reagents. The template-switch primer was designed and synthesized in-house (12µM, CGAAATTAATACGACTCACTATAGGAGAGTACGCGGG). The first-strand and template-switching reaction was performed at 42°C for 2hrs 15minutes in a thermocycler with a non-heated lid. Second strand cDNA synthesis was done by the addition of 2U RnaseH, Roche (Indianapolis, IN), Advantage 2 polymerase (to 1X final concentration), 10X Advantage 2 polymerase buffer(to 1X final concentration), nucleotides (to 1X final concentration), Clontech (Palo Alto, CA), and water to a final reaction volume of 150µl. The second-strand reaction was incubated in a thermocycler with a heated-lid, with the following protocol; 37°C for 5', 94°C for 2', 65°C for 1', followed by 75°C for 1hr. Double-stranded cDNA was purified using PCR clean-up kits, Qiagen (Valencia, CA), eluted in 80µl nuclease free water and concentrated to 5µl in a speed-vac with no heat.

Detailed LAPT Protocol

Reverse transcriptase-template switch (RT-TS) reaction (per tube):

X ul	RNA
3.5 ul	primer (12 uM) or 4.2ul primer 10uM

X ul Nuclease Free Water (fill to 28.5 ul)

28.5 ul Total Volume

Mix contents and briefly spin in micro-centrifuge.

Heat in thermocycler with NO heated lid at 65°C for 2 minutes, then reduce temperature to 42°C until ready to add remaining reagents.

Add remaining reagents (form Master Mix if applicable):

10 ul First-Strand Buffer

1 ul DTT (100 mM)

5 ul 50X dNTP

2.5 ul Rnase-In

2.5 ul Powerscript RT

3.5 ul Template switch primer (12 uM)

24.5 ul Total volume

28.5 ul + 24.5 ul = 53 ul Total Volume.

The tubes were then incubated at 42°C for 90 minutes in thermocycler with NO heated lid to allow for RT extension.

Double-Strand Synthesis (per tube):

53 ul Total volume from RT-TS Reaction

75 ul Nuclease Free Water

15 ul Advantage PCR Buffer

3 ul	10 mM dNTP mix
1 ul	Rnase H
3 ul	Advantage 2 Polymerase
150 ul	Total Volume

Place tubes in hot-lid thermocycler and Run at 37°C for 10 minutes to digest the mRNA, 94°C for 2 minutes to denature, 65°C for 1 minute for specific priming, and 75°C for 30 min. for extension.

DNA clean-up: dsDNA was then purified using a QIAGEN PCR Purification Kit elute in 100ul

After DNA cleanup, the DNA was spun down in a SpeedVac.

***In vitro* Transcription.**

In vitro transcription, using the double-stranded cDNA as template, was carried out for 16hrs using T7 Megascript kits, Ambion (Austin, TX), in 40µl reactions with an additional 400U of T7 polymerase. RNA was recovered following protocols described for RNeasy kits, Qiagen (Valencia, CA), and eluted in 80µl nuclease free water with 20U of SUPERNase•IN™ (Ambion).

T7 Amplification Reaction

40ul MegaScript Reaction

4ul dATP
4ul dCTP

4ul dGTP
 4ul dTTP
 4ul 10X Reaction Buffer
 2ul Enzyme Mix
 2ul T7 polymerase
 1 Super RNAsin
 X DNA template
 Water to 40ul

Incubate at 37C for 6-16hrs

Amplification of *in vivo* samples

Reverse transcription and template-switch primer addition was performed using a half reaction (50µl) of the Super SMART™ protocol (Clontech, Palo Alto, CA). The primers used for RT were Genome-Directed primers designed towards *Yersinia pestis* CO92. The template-switch primer was designed and synthesized in-house (12µM, CGAAATTAATACGACTCACTATAGGAGAGTACGCGGG). The reaction was performed at 42°C for 2hrs 15minutes. Second strand synthesis was done by the addition of 2U RnaseH (Roche, Indianapolis, IN), and Advantage polymerase and 10X buffer (Clontech, Palo Alto, CA), in a final reaction volume of 150µl. The reaction was incubated at 37°C for 5', 94°C for 2', 65°C for 1', followed by 75°C for 1hr. Double-stranded cDNA was purified using PCR clean-up kits (Qiagen, Valencia, CA), and dried in a speed-vac with no heat.

Microarray Hybridization and Analysis.

Samples were prepared for hybridization by concentrating appropriate samples together using Microcon YM-30 filters (Millipore, Bedford, MA). The sample was eluted in 10 μ l water and 100 μ l Slide Hyb #3 (Ambion), added. Samples were hybridized to slides in a Genomic Solutions Hybridization Station (Harvard Biosciences, Holliston, MA) using the protocol: 65°C for 3hrs, 60°C for 3hrs, 55°C for 12hrs, followed by three washes with wash sol. 1 (2X SSC, 0.1% SDS), three washes with wash sol. 2 (0.1X SSC, 0.1% SDS), and three washes with wash sol. 3 (0.1% SSC).

Slides were subjected to a brief wash in water and dried by centrifugation. Slides were scanned at 10 μ m resolution at the appropriate wavelengths using a Scanarray Express, Perkin Elmer (Wellesley, MA). Images were imported to, and data extracted using GenePix (v3.0) software (Axon Instruments, Union City, CA).

Data analysis.

Using Excel (Microsoft), data was initially thresholded by removing all signals less than two standard deviations above negative control spots. The ratio of the expression intensity over the genomic standard intensity was then calculated. The Z-scores for each experiment was also calculated with Excel using the formula $Z\text{-score for gene } x = (\text{expression value for gene } x - \text{average expression value for the experiment}) /$

(standard deviation of the expression values for the experiment). These data were imported into GeneSpring (Silicon Genetics), software for analysis. Each experiment was normalized to the 50th percentile of each experiment. The cross-gene error model was used based on experimental replicates. All statistical analysis was performed using tools within GeneSpring. Clustering of data was done using standard clustering method unless otherwise noted in the text.

We also utilized the Artemis genome annotation tool (Sanger Center) for visualization of genomic structure, genomic organization, %GC content, and miscellaneous genomic features.

Quantitative Real-Time PCR (QRT-PCR)

Selected genes (n=18) were subjected to QRT-PCR(158). Templates used were 2.5ul of a 1:100 dilution of cDNA samples derived from the reverse transcription of 10ug of RNA in standard 20ul reactions (Stratagene). Each sample was run in triplicate for the specific gene in study along with a standard of 16S ribosomal RNA amplification for normalization. The PCR efficiency for all primer sets was determined using a minimum of 4 logs of diluted *Yersinia pestis* genomic DNA as template, as determined by the equation $E=10^{[-1/\text{slope}]}$. The calculated primer efficiencies were incorporated into the determination of fold change as described by M.W Pfaffl(138). The basic equation for fold change determination is as follows Fold change was determined using the formula $\text{Fold change}=2^{\Delta\Delta c_t}$ where $\Delta\Delta c_t=(C_t\text{target}-$

$C_{16SrRNA})_{time_x} - (C_{target} - C_{16SrRNA})_{time_0}$, with the C_t for each measurement being corrected by the PCR efficiency. Time zero was defined as either the 27°C control for *in vitro* studies or Day 1 post-infection for *in vivo* studies.

Models of *Yersinia pestis* infection

Bacterial growth and RNA preparation.

Yersinia pseudotuberculosis pIII pYV was kindly provided by C. Darby (Stanford, CA). Bacteria were grown on LB/Agar plates at 27°C O.N., and cells collected by addition of 2ml of RNA Protect, Qiagen (Valencia, CA), directly to the plates. RNA isolation was done following protocols described in RNeasy kits Qiagen (Valencia, CA). The isolated RNA was pooled and served as a total bacterial RNA stock.

The models of *Yersinia pestis* infection used in these studies were carried out at the University of New Mexico Health Science Center under the director C. Rick Lyons

Figure 11.

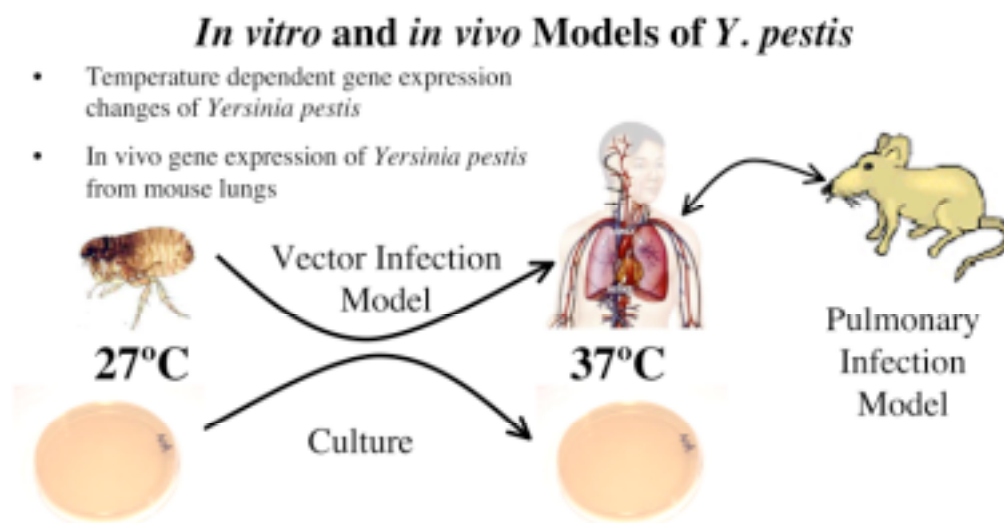


Figure 11 Models of *Y. pestis* infection used in these studies.

Two models were employed in these studies, an *in vitro* model used to mimic temperature-dependent changes commonly encountered during vector-borne transmission to mammals and an *in vivo* model of pulmonary infection.

The *Yersinia pestis* clinical isolate used in these studies.

YPNM1, was obtained from the New Mexico State Diagnostic Lab and was isolated from an infected cat. All work was performed at animal biosafety level 3 using protocols approved by the institutional biosafety committee.

***In vitro* Growth of *Y. pestis* and RNA isolation.**

To evaluate the effect of temperature on transcription the frozen stock of YPNM1 is streaked onto selective agar to provide single colonies. Lawns are then prepared from a single colony and placed at 27°C. At forty-eight hours, some plates were placed at 37° C and other left at 27°C. Bacteria are subsequently harvested at timed

intervals. Bacterial RNA isolation was performed using Trizol (Qiagen, Valencia, CA). The organisms were disrupted using a bead beater and the RNA isolated according to the manufacturer's protocol.

Pulmonary infection model of *Y. pestis* .

Mice were inoculated on day 0 with ~ 5,000 organisms intranasally. This was accomplished by applying 50ul of inoculum directly to the nares of isoflurane-sedated mice, and the mice were allowed to inhale the inoculum. Mice were then monitored closely post inoculation. All mice succumbed to infection between days 3-5 and were euthanized upon signs of overwhelming morbidity. All protocols were approved by the University of New Mexico and UT-Southwestern IACUCs.

CFU determination of infected mouse tissues.

Mice were infected i.n. with *Y. pestis* . Mice were euthanized at 24, 48 and 72 hours after infection. The lung, spleen and liver were harvested, organs homogenized and CFU determined by plating serial dilutions on selective agar and incubating at 27 ° C for 2 days.

Sample collection and RNA isolation from mouse pulmonary infection model.

Organs to be harvested were mechanically disrupted using a bead beater in Trizol. Isolation of RNA was then performed using the manufacturer's protocol. Ten

percent of the isolated RNA was then plated onto selective agar to confirm the absence of viable organisms.

CHAPTER THREE

AMPLIFICATION OF SENSE-STRANDED PROKARYOTIC RNA

Summary

Microarray expression analysis has proven to be a valuable methodology. In eukaryotic systems where RNA is limiting, established protocols for amplification of mRNA which rely on the poly(A) tails are well established. In contrast, the difficulty in amplifying prokaryotic mRNA has limited the application of microarrays to microbiology. Here I present a method for the Linear Amplification of Prokaryotic Transcripts (LAPT) that is relatively efficient and unbiased. The overhang tailing activity of Moloney murine leukemia virus reverse transcriptase is used to add the T7 promoter to cDNAs during reverse transcription. The promoter addition is uncoupled from the initial priming event allowing the promoter to be attached to the 5' end of the RNA transcript. This enables the amplification of sense-stranded RNA that is representative of the complexity and distribution of the original transcript pool. In microarray assays, amplified prokaryotic RNA (10ng total RNA starting material) showed good Spearman correlations to an un-amplified control sample. Using genome-directed primers to bias addition of a T7-promoter to bacterial transcripts allowed amplification of prokaryotic transcripts in the presence of mammalian RNA (at a eukaryotic/prokaryotic RNA ratio of 500 to 1). This technology should facilitate the study of prokaryotic transcriptomes in situations, such as *in vivo* studies or mixed

microbial populations, where the prokaryotic RNA amount is limited and/or the non-target/target RNA ratios is high.

Introduction

While the development of microarray technology has facilitated the survey of gene-expression levels on a global scale (52), a practical limitation of these studies is the amount of RNA that is required to detect the entire transcriptome. This is particularly an issue for *in vivo* pathogenesis studies when relatively few pathogens are present, and the pathogen RNA (prokaryotic) cannot be easily separated from host RNA (eukaryotic). It has been shown that pathogens modify their global transcriptional-profile uniquely *in vivo*(28, 29, 121, 122, 148, 171). These studies, however, relied on either non-natural *in vivo* growth chambers(122, 148), or extensive bacterial enrichment procedures(28, 29, 171)to obtain enough RNA for microarray analysis. Furthermore, the ability to evaluate microbial expression from mixed populations has been limited by the inability to selectively amplify a single microbial transcriptome. This is becoming increasingly important in the desire to study mixed microbial communities. The development of a general method to expand prokaryotic mRNA, particularly in the presence of host or non-target mRNA, would be an important advance in studying pathogenesis and ecology. We report such a method here.

Genome-directed primers (GDPs) enable the selective priming of organism-specific RNAs from a heterologous sample(170). This process generates a minimal set

of primers required to prime all or a set of transcripts from a sequenced genome.

However, sufficient RNA from the organism is still required for downstream analysis. Oftentimes prokaryotic RNA, in the context of a mammalian infection, is limiting. A commercial technology, MicrobEnrich (Ambion, Austin, TX) has been developed to remove mammalian RNA, enriching for prokaryotic RNA. Again this technology is limited by the initial amount of prokaryotic RNA, and may confer bias in the mRNA left (see below). Alternatively, the enrichment of the bacteria from the sample prior to RNA extraction has been utilized. This approach is often precluded by technical limitations. Furthermore, the short half-life of most bacterial RNA makes time a significant variable in sample quality. When mixed microbial populations are evaluated, the inability or time required to separate species, and the lack of any means to separate RNA populations after extraction precludes many studies. Therefore, methods to efficiently amplify prokaryotic mRNA are needed.

Two methods to amplify signals from prokaryotic transcripts have been reported to date. Differential expression using customized amplification libraries (DECAL) (3), utilizes a non-linear amplification strategy (PCR) and is able to detect 4-fold changes in expression levels from 10ng starting material. However, a detection threshold of 4-fold change is limiting and DECAL has not been demonstrated to work using *in vivo* material. Second is a method described by Motley et al., which uses polyA polymerase to add an adenine tail to bacterial transcripts making it amenable to eukaryotic methods that produce anti-sense RNA(122). Where this process enables the amplification of anti-sense bacterial transcripts it does not provide any means to discriminate RNA

populations prior to or during amplification. This method is also used in commercial bacterial amplification systems (Ambion). Again this strategy produces anti-sense RNA and requires enrichment of the bacterial RNA prior to amplification as the addition of the T7-promoter is non-discriminatory. In the widely used amplification protocol for eukaryotic RNA (Eberwine amplification) (180), anti-sense RNA is amplified from original transcripts by priming the RNA with a T7 promoter conjugated to an oligo(dT) primer to prime polyadenylated RNA during the reverse-transcription reaction. The resultant T7-tagged cDNA is used to produce large amounts of anti-sense RNA (aRNA) in an *in vitro* transcription reaction that is processive and independent of the sequence length or composition. The anti-sense RNA product is an amplified “mirror image” of the original RNA sample, and hence requires compensation for strandedness in downstream applications. The efficiency and un-biased nature of T7-promoter driven amplification has been well demonstrated and this technique has been used to perform a variety of sensitive assays, most notably expression analysis (19, 123, 181).

Moloney murine leukemia virus reverse transcriptase (MMLV) leaves a tail of 2-5 cytosine residues after reverse transcribing a RNA template (36). An oligonucleotide containing 3' guanine residues can anneal to these cytosines, allowing the MMLV to effectively switch templates and synthesize the complement strand of the oligonucleotide. This phenomenon, known as template-switching, has been used for a variety of techniques (117, 159, 181, 192)(e.g. 5' RACE, cDNA library construction, Step-Out PCR, and microarray probe generation).

Here I present a method that takes advantage of the template-switching affect of MMLV to add a T7-promoter to the 5' end of RNA transcripts. We call the process Linear Amplification of Prokaryotic Transcripts (LAPT). Application of this process enables amplification of sense-stranded RNA that retains the biological information of the original transcript pool. Furthermore, by coupling LAPT with biased priming enrichment techniques we demonstrate amplifying prokaryotic RNA from a mixed RNA sample. This should enable the *in vivo* study of prokaryotic transcriptomes from any stage or condition, the study of transcriptional profiles of limited samples, and the ability to investigate expression profiles from mixed microbial populations.

Results

Outline of LAPT process.

The general scheme for the LAPT process is outlined (Figure 12 Depiction of the steps in the LAPT process.). In the case of gene-specific priming, a standard reverse primer can be used to initiate reverse transcription. The priming of the 3' ends of a population of RNA transcripts can be accomplished by oligo(dT) primers in the case of polyadenylated transcripts, or random hexamers or genome-directed primers in the case of prokaryotic transcripts. When the MMLV reaches the end of an RNA transcript a template-switching guanine-tailed primer containing a T7-promoter sequence (T7-TS) is added. Enzymatic removal of the template RNA strand and subsequent second strand synthesis results in a linear double-stranded cDNA with a

functional T7-promoter at the 5' end. This modified cDNA serves as a template during *in vitro* transcription reactions using T7 polymerase.

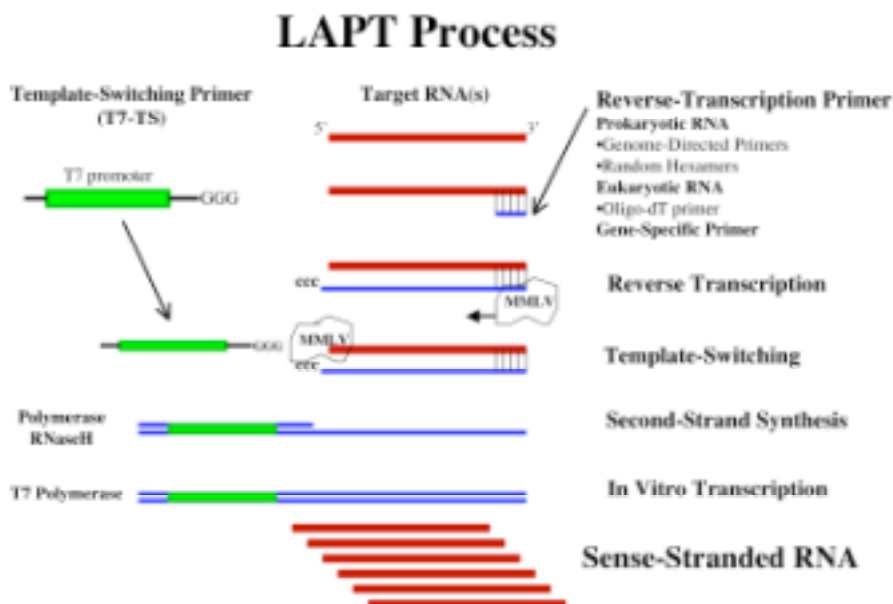


Figure 12 Depiction of the steps in the LAPT process.

Target RNA is primed as in a standard reverse transcription reaction. In the presence of the template-switching primer the MMLV adds the T7 promoter to the nascent cDNA. Second-strand synthesis produces a template for an *in vitro* transcription reaction producing multiple copies of the original target RNA. Unlike the Eberwine process, the priming for reverse-transcription and T7-promoter addition is uncoupled; allowing for a variety of reverse-transcription priming options.

Addition of a functional T7-promoter to the 5' end of GFP RNA.

I first demonstrated that the T7-TS primer was efficiently added to the 5' end of a green-fluorescent protein (GFP) RNA transcript (Figure 13A). This is confirmed by the generation of an additional double-stranded cDNA product that is ~37bp larger, suggesting that the template-switching oligo is being covalently attached in the reaction.

When the template-switching primer is not added in the reaction mix, this band is not seen. Primer addition occurs when the GFP transcript is primed with gene-specific primers (20mers) targeted to various positions throughout the gene or when random hexamers are used to prime the gene. Though this reaction did not go to completion under these conditions, based on peak intensity, 40-48% of the cDNA product produced had the added T7 promoter. This experiment demonstrates that the template switching reaction occurs independently of the initiation of reverse transcription, and is not influenced by the size or location of the initial transcript priming.

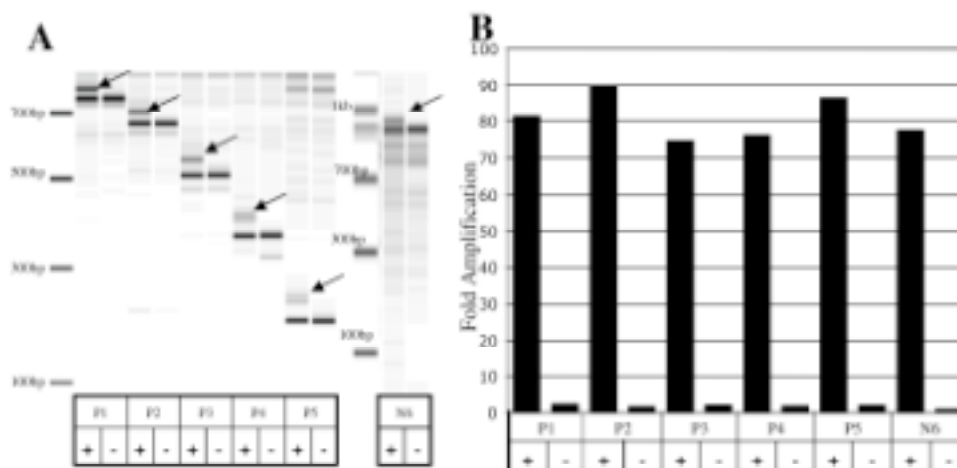


Figure 13 Verification of addition of a functional T7-promoter to the 5' end of GFP transcripts.

(A) Bioanalyzer image after 1 μ g of template GFP RNA was subjected to the Reverse-Transcription reaction with one of 5 priming oligonucleotides (P1-P5) or random hexamers (N6) in the presence or absence of the T7-TS oligo. In the presence of the T7-TS primer a cDNA product approximately 37bp larger is seen (arrows), which is not present when the primer is not added. This shows the T7-TS oligo is being added to the RT product in the reaction. (B) The resultant cDNA products were then purified and used as templates in *in vitro* transcription reactions. The products were quantified and expressed as μ g output RNA/1 μ g input RNA (fold amplification).

To test whether the T7-promoter was functionally active, the GFP-specific cDNAs generated above were used as templates for *in vitro* transcription reactions. Only when the T7-TS primer was used did we see significant (>60 fold as calculated by ug of recovered RNA divided by 1ug initial input RNA used in the reverse transcription reaction) amplification from the original input RNA (Figure 13B). This data demonstrates that the T7-TS primer was necessary for subsequent transcript production, and that the amount of RNA produced is independent of the length of the RNA product or the priming source. In this experiment the fold amplification appeared to be limited only by the 40ul volume of the *in vitro* transcription reaction that was capable of generating 74-89ug of RNA product.

Microarray construction and amplification of complex RNA samples.

I tested the practical application of the LAPT technique using microarrays for *Yersinia pestis*. I had developed microarrays of 70mer long-oligonucleotides covering all genes described in the genome and three associated virulence plasmids (130). Furthermore, we designed and synthesized 294 GDPs specific for all *Y. pestis* genes and ensured that each gene was able to be primed 3' from the location of the cognate 70mer probe. The genetically similar *Yersinia pseudotuberculosis* (*Yptb*) was used to validate these resources (data not shown), and was used as a surrogate for *Y. pestis* in the remainder of this study (81, 141).

In order to test the ability of this process to amplify a complex mixture of RNA I applied the technique to total *Yptb* RNA preparations. Aliquots from 5 μ g to 0.01 μ g RNA were amplified according to the protocol outlined in Figure 12 using the *Y. pestis* GDPs to initiate reverse transcription. The quality of the RNA preparations and amplified products was evaluated using a Bioanalyzer 2100 (Agilent) and showed expected patterns (Figure 14)

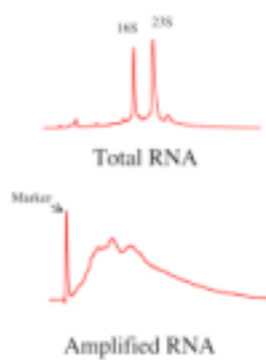


Figure 14 Characteristic bioanalyzer pattern of amplified RNA.

The figures are representative of the RNA quality before and after amplification. The amplified RNA showed a mean size of ~850 bp which is slightly truncated due to the priming site of the genome-directed primers.

The amplification of the RNA appeared to be limited only by the *in vitro* transcription reaction volumes and conditions. As a result of this process an ~3000 fold amplification of prokaryotic RNA was obtained from 10ng of total RNA starting material, yielding 30-50 μ g of amplified sense-stranded RNA, an amount sufficient for multiple microarray experiments or most other downstream techniques (Table 4).

Starting Concentration	Total yield ug	Fold Amplification
5000ng	60.8	12
1000ng	52.5	52
500ng	48.0	96
100ng	49.5	495
50ng	49.0	981
10ng	28.5	2847

Table 4 Typical RNA yields from LAPT.

The table is representative of the total yield and fold-amplification seen when *Yptb* total RNA was subjected to the amplification process.

Maintenance of biological information of a complex RNA sample.

Besides the amplification itself, another important consideration is how well the mRNA representation is maintained. I used global gene expression patterns to evaluate the linearity of the amplification process. RNAs from amplified and un-amplified samples were normalized based on concentration and used to prepare labeled analyte for microarray hybridization. I employed genomic normalization in these studies for benefit in data quality and ability to compare multiple samples in a minimum of experiments (169). The average intensity of each gene was higher in the amplified samples than the un-amplified control. This is illustrated in Figure 15A, where the mean genomic normalized signal intensity (fluorescence signal of the gene/fluorescence signal of genomic control for that gene) is 5.3 where in average of the un-

amplified control it is 2.9. Given that the genomic control signal is normalized to only be 2-3% of the dynamic range of possible signal intensity (as not to interfere with signals from genes), these data suggest that my ability to confidently detect low-level transcripts is improved following amplification. It was observed that as the sample was subjected to increasing amplification, the data centered more closely around the mean expression level for the experiment, suggesting that amplification results in a decrease of the signal distance between any two genes. As a measure of the linearity of the amplification process Figure 15B shows a scatter plot comparing an unamplified and amplified RNA sample. The Pearson correlation value is 0.77 and demonstrates maintenance of biological information.

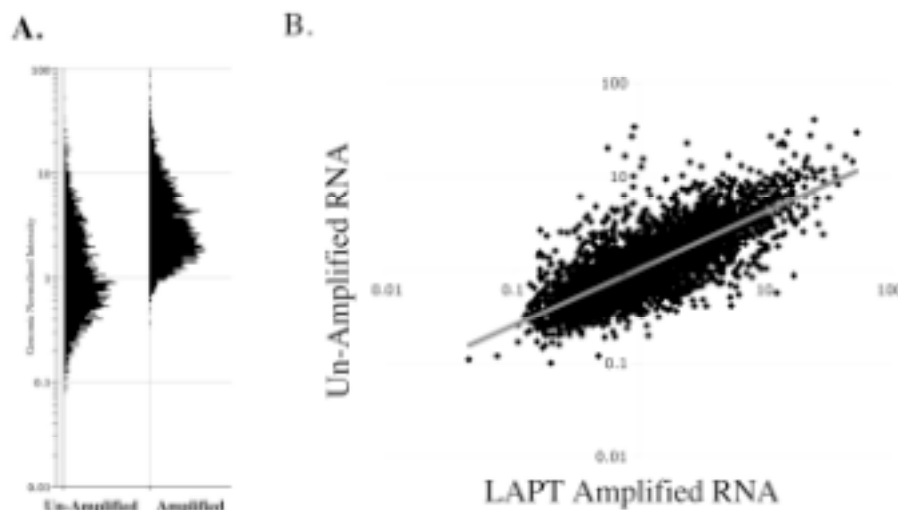


Figure 15 Analysis of microarray signals from un-amplified and LAPT amplified RNA.

(A) Amplification increases the mean Genomic-normalized signal. The mean Genomic-Normalized signal increases after amplification supporting more confident detection of the entire transcriptome. (B) Scatter plot of microarray analysis comparing unamplified control RNA vs. LAPT amplified sense-stranded RNA from 1ug of starting material.

To further test whether there was a bias in the amplification process, I evaluated the correlations between the un-amplified control and multiple amplified RNAs starting from 5ug of initial template down to 10ng of initial RNA template (Table 5). The Spearman correlation, a measure of the maintenance of rank-order between populations, determined that the rank-order was well maintained and suggested that although some of the magnitude in differential expression was reduced, the order of the gene-expression levels was largely maintained. To test this assessment in a different way, ANOVA was performed using the same data set. This determined that only 0.6-2.6% of the genes showed significant differences after amplification, whereas, 81.0-86.0% were similar across the range of amplification (Figure 16). These data indicate that the complexity and composition of the original transcript pool was largely maintained and that RNA subjected to the LAPT process retains most of the biological information observed in the initial population of small starting amounts of RNA.

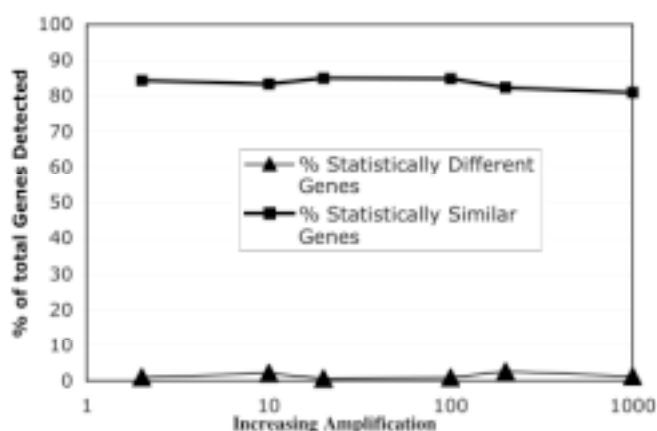


Figure 16 LAPT amplification maintains expression information on microarrays.

The graph shows the percentage of similar and dissimilar genes as determined by ANOVA analysis. The fold amplification represents the amplification required to obtain 10ug of LAPT amplified RNA for array hybridization.

Selective amplification from a mixed sample.

The two common hindrances when studying microbial transcription from mixed RNA samples are the amount of RNA available and the presence of excess non-target RNA. To evaluate the ability of this method to overcome these obstacles, I amplified prokaryotic RNA that was spiked with various amounts of excess mouse RNA (universal mouse reference RNA from Stratagene). Even in the presence of excess mouse RNA, up to 500 to 1, the rank-order of the population is largely maintained (Table 5).

Starting amount of <i>Yptb</i> RNA (ng)	Mouse RNA (ng)	Fold excess Mouse RNA	Spearman Correlation
5000	0	0	0.75
1000	0	0	0.69
500	0	0	0.67
100	5	50	0.65
100	10	100	0.51
50	0	0	0.44
10	0	0	0.45
10	5	500	0.50

Table 5 Spearman correlations of amplified *Yersinia* samples relative to an unamplified control.

The Spearman correlation coefficient, a measure of maintenance of rank-order in the population, for samples amplified from decreasing amounts of starting material in the presence or absence of mouse RNA in the reaction.

These data affirm that the biased priming of the *Yptb* RNA by the GDP's drives the selective addition of the T7-promoter. Furthermore, the biological information of the sample was maintained during amplification. It is probable that non-target RNA is inadvertently being primed and subsequently amplified, but this does not interfere with maintenance of rank-order in part because the hybridization to the microarray selects against the eukaryotic transcripts.

CHAPTER FOUR

EXPRESSION PROFILING OF *YERSINIA PESTIS* DURING MOUSE PULMONARY INFECTION

Summary

Yersinia pestis, the causative agent of plague, can be transmitted by infected-flea bite or inhaled aerosol. Both routes of infection have a high mortality rate, and pneumonic infections of *Yersinia pestis* represent a significant concern as a tool of bioterrorism. Understanding the transcriptional program of this pathogen during pulmonary infection should be valuable in understanding plague pathogenesis, as well as potentially offering insights into new vaccines and therapeutics. Towards this goal we developed a long-oligonucleotide microarray to the plague bacillus and evaluated the expression profiles of *Y. pestis* *in vitro* and in the mouse pulmonary infection model *in vivo*. The *in vitro* analysis compared expression patterns at 27°C versus 37°C. The *in vivo* analysis used intranasal challenge to the mouse lung. By amplifying the *Y. pestis* RNA from individual mouse lungs we were able to map the transcriptional profile of plague at post-infection days 1 to 3. Our data present a very different transcriptional profile between *in vivo* and *in vitro* expression, suggesting *Y. pestis* responds to a variety of host signals during infection. Of note was the number of genes found in genomic regions with altered %GC content that are up-regulated within the mouse lung

environment. These data suggest these regions may provide particularly promising targets for both vaccines and therapeutics.

Introduction

Events of recent years have reemphasized the need for the development of new therapies to biothreat agents. The goal is to develop both vaccines and therapeutics to respond, and potentially deter, the use of dangerous pathogens. A pathogen of considerable concern is *Yersinia pestis*, the causative agent of plague. Classified as a category A agent, *Y. pestis* has a number of characteristics that make it of particular concern (87). Notably, no vaccine for *Y. pestis* is in production, and the previous vaccine formulation may not be protective after pulmonary exposure, the likely route of malicious release (39). Additionally, the high mortality rate of untreated pulmonary plague coupled with its non-descript clinical presentation emphasizes the need for effective prophylaxis. Confounding the need for new therapies for plague is the fact that *Y. pestis* is amenable to genetic manipulation and both natural and non-natural antibiotic resistance in *Yersinia sp.* (41, 63, 73) have been described. We are interested in applying microarray technologies to facilitate development of countermeasures to *Y. pestis* and other pathogens.

The ability of *Y. pestis* to successfully infect and cause disease in multiple hosts is attributed to environmental recognition and coordination of appropriate gene expression (134). As a zoonotic disease, sustaining itself within natural reservoirs, *Y.*

pestis must shuttle between arthropod vectors and mammalian hosts of which humans only play an incidental role (127). These different environments result in utilization of unique systems to survive. Within flea vectors the bacteria must counteract lysozyme-like enzymes, establish protective biofilms, and transition from the midgut to the proventriculus/esophagus where they are able to transmit to mammalian hosts (107). Similarly, survival of *Y. pestis* in the mammalian host, whether acquired via vector or aerosol, requires a variety of systems involved in effector secretion, iron acquisition, capsule synthesis, toxin production, and other described virulence determinants (127). Expression of these determinants has been shown *in vitro* to be regulated by temperature, pH and divalent cation concentration. Recently, expression analysis has been used to map the transcriptional response of *Y. pestis* to individual environmental stresses *in vitro*, such as temperature, antibiotics, and osmotic stress (74-76, 121, 139). Whereas the responses to these environmental stresses have been elucidated using *in vitro* systems, there has not been a comprehensive analysis of their expression within a host. Most previous studies have focused on vector transmission, while the transmission of concern as a biothreat agent is aerosol.

Little is known about the pathogenesis of *Y. pestis* when delivered by aerosol. Plague has been shown to have a higher LD₅₀ when delivered intranasally (I.N.) compared to subcutaneously, both of which can be reduced by prior growth at higher temperatures (134). Furthermore, *Y. pestis* exposed to macrophages are subsequently more resistant to killing by phagocytes (34). Given the limited study of *Y. pestis* in

pulmonary models and the risk this disease poses, it warrants further investigation of plague pathogenesis via this route of infection.

The advantages of *in vivo* models for pathogenesis research are well documented with technologies such as signature-tagged mutagenesis, differential fluorescence induction, and other *in vivo* expression technologies being used for these studies (79, 114, 179). Signature-tagged mutagenesis has been employed identify 16 *Y. pestis* genes required for spleen colonization following subcutaneous infection, which represented >5% of the genes in the screen (56). Along with the success of these techniques is their limitations imposed by their experimental design and technical constraints, as in the mentioned study only 300 mutants were screened. *In vitro* models have the advantage of highly controlled variables, however they often are unable to accurately mimic the diversity of environmental conditions encountered *in vivo*. *Y. pestis* virulence has often been studied *in vitro* by evaluating the affects of different growth temperatures and Ca^{2+} concentration on gene expression and virulence (134). These, while insightful, are unable to dissect the influence of diverse environmental factors or changes in *Y. pestis* physiology after infection.

Microarray technology has enabled the survey of genome-wide transcriptional activity of a pathogen (109). Efforts to apply this technology to pathogen samples from *in vivo* environments have been limited by both a lack of pathogen abundance and the presence of host mRNA in preparations. Despite these limitations there are examples of array technology being utilized to evaluate the expression profiles a pathogen from *in vivo* environments (100, 122, 148, 166, 171). These recent studies have demonstrated

that pathogens utilize unique expression patterns within their hosts. Furthermore, the results of these studies have provided unique insights into the pathogenic mechanisms of these bacteria. For the majority of these studies, non-natural *in vivo* models were employed often with extraction or enrichment procedures. Recent technological advances have facilitated the application of microarrays to *in vivo* pathogen gene analysis(169, 170). We have employed these approaches to directly map the global expression patterns of plague within a mammalian host following pulmonary infection.

Results

Characterization of the arrays using temperature-regulated gene expression *in vitro*.

Before embarking on analysis of *in vivo* expression we first wanted to validate the arrays and protocols we were using. Several groups have now published data on global microarray analysis of gene expression of *Y. pestis* in different *in vitro* contexts (74, 76, 121). We wanted to determine if our arrays and protocols produced results generally in concordance with previous results. This data also served for comparison to the *in vivo* data we obtained.

For our *in vitro* model we evaluated the gene expression changes upon temperature shift from 27°C to 37°C using plate cultures of *Y. pestis* . We selected the

plate model to aid us in focusing on temperature-induced changes as opposed to growth phase-dependent changes to validate our resources, as well as to provide a general transcriptional reference for future *in vivo* expression data. Though at a general stage of low activity, the representation of bacteria at multiple stages of growth on solid media enable the population averaging of those gene changes not modified by temperature. Analysis of 2-fold expression changes from 27°C to either 30 minutes at 37°C (Early temperature dependent gene changes) or 2 hours at 37°C (Late temperature dependent gene changes) demonstrated that 2.4% or 3.1% of the transcriptome was altered respectively (Figure 17). When the analysis was performed to include all genes with 1.5 fold expression changes the percent of the transcriptome modified increased to 8.1% and 14.9% for early and late gene changes respectively. These data show a lowered state of transcriptional activity in these cells, however reinforce previous studies describing temperature as a significant environmental signal affecting transcription in *Yersinia* species. Our analysis of the pattern of gene expression was also largely consistent with previous findings of temperature-induced changes in *Y. pestis* (76, 121). An example of parallel findings includes the general trend of genes being transcriptionally down-regulated in this condition, including a large number of ribosomal and amino acid biosynthetic genes. However, we did note a few differences. For example, many of the *ysc* and *yop* genes were not up-regulated in our system, which may be the consequence of the growth on plates as opposed to liquid log-phase growth employed in other studies.

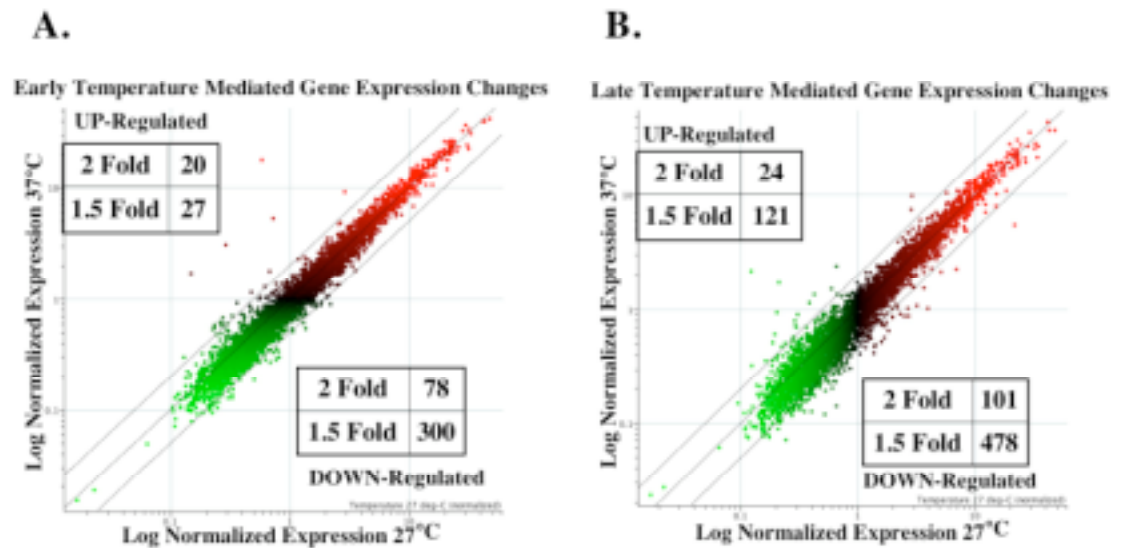


Figure 17 Scatter plots of Early (0.5 hr) and Late (2 hr) gene expression differences.

Scatter plots of Early (0.5 hr) and Late (2 hr) gene expression differences (A and B respectively) from *Y. pestis* when subjected temperature shifts from 27°C to 37°C. The numbers of genes regulated at the 2-fold and 1.5-fold level are given with 2-fold lines displayed on the graph.

To independently test the data from the *in vitro* microarray experiments, quantitative real-time PCR was performed on eight genes found to be regulated 2 fold from 27°C at either time-point after temperature shift, for a total of 16 measurements. Results of this analysis revealed a concordance of 93.8% in the direction of gene expression change. This is in line with other reported similar evaluations (169). The correlation of microarray data and Q-PCR expression values as measured by linear regression resulted in an R-value of 0.67 (Figure 18). Although not an ideal system to evaluate temperature dependant gene expression changes these experiments validated our arrays and protocols, and provide a reference for later comparison of *in vivo* expression data.

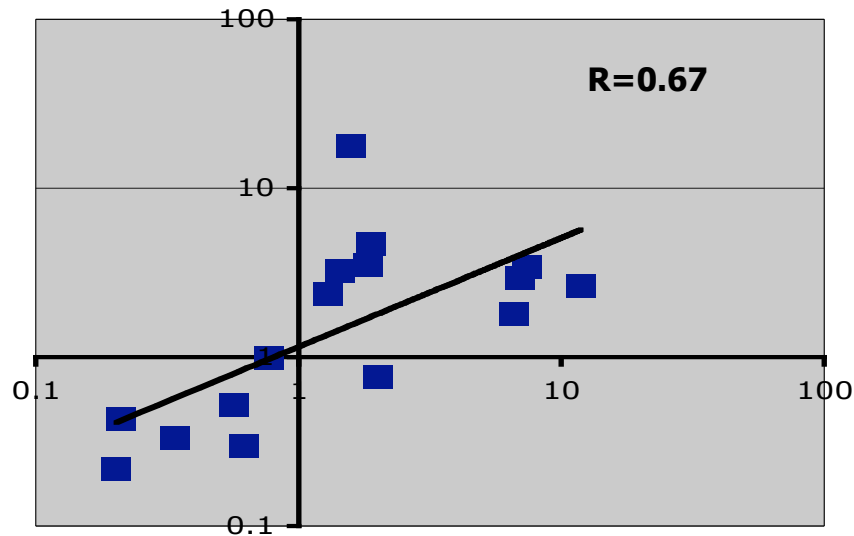


Figure 18 Correlation of RT-PCR and Microarray Data.

Comparison of expression values obtained through quantitative RT-PCR and microarray expression analysis reveals a strong correlation in the direction of gene expression changes and to a lesser degree the magnitude of gene expression ($R=0.67$).

Characterization of pulmonary infection model

The *Y. pestis* pulmonary infection model was established at the University of New Mexico. Clinical isolates of *Y. pestis* (YPNM1) were obtained from the New Mexico State Diagnostic Laboratory. Stock cultures were prepared and used to infect mice. The initial study with *Y. pestis* was to determine whether there was a genetic resistance/susceptibility to pulmonary infection as well as determine the LD_{50} for future experiments. Four different strains of mice were used and, in general, there were only minor differences in the LD_{50} s among these 4 strains of mice (Table 6).

		BALB/c	C3H/HeN	DBA/2	C57BL/6
		LD50	LD50	LD50	LD50
YPNM 1	Pulmonary	600 organisms	375 organisms	275 Organism	375 organisms
YPNM1	Subcutaneous	<2	<2	<2	<2

Table 6 LD₅₀ of YPNM1 in four strains of mice.

The LD₅₀ of *Y. pestis* strain YPNM1 in four strains of mice infected either intranasally or subcutaneously.

The time course to death typically lasted 4-5 days. Of note, the survivors in the lower inoculum dose did not develop any clinical symptoms during the 3 weeks observation period. This suggests that survival was not dependent upon the development of an acquired immune response but rather dependent on whether the infectious inoculum was large enough to overcome the threshold for establishing a viable infection.

Since there were no significant differences in the LD₅₀ among the different strains of mice we chose the BALB/c mouse to further characterize the model. Mice were inoculated on day 0 with ~ 5,000 organisms intranasally and are then monitored closely for morbidity and mortality. Using this protocol we defined the time to death for BALB /c mice as all mice succumbed to infection between days 3-5 (Figure 19A).

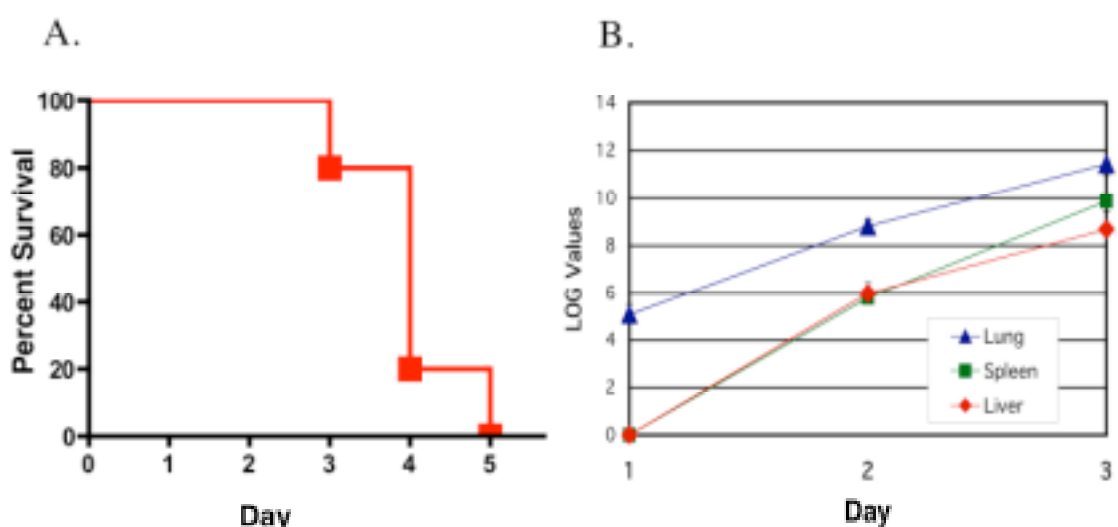


Figure 19 Survival curve and dissemination of YPMN1 in mice.

(A) Survival curve of Balb/c mice infected intranasally with 5000 cfu of *Y. pestis* strain YPMN1. (B) Bacterial burden in lung, spleen and liver at days 1,2, and 3 post infection (Lung (▲), Spleen (■), and Liver(●)).

We investigated the kinetics of the pulmonary infections by determining the CFU of *Y. pestis* in various tissues post-infection. Mice were infected i.n. with *Y. pestis* and were euthanized at 24, 48 and 72 hours after infection. The lung, spleen and liver were harvested, organs homogenized and CFU determined (Figure 19B). Dissemination does not occur until after 24 hours post inoculation, which is also the case for subcutaneous

infections (112). The peak bacterial burden in the lungs reaches 10^{12} at day 3 post infection, which are similar to a subcutaneous model in *Rattus norvegicus* (160).

In general there is little inflammation in the lung until after the *Y. pestis* has disseminated from the lung. Basically the bacilli grows unchecked in the lung until after the organism has escaped, suggesting that there may be virulence factors carried by the organism that inhibit early inflammatory responses to the pathogen in the lung. This is supported by the observation that neutrophils do not infiltrate the lung until 48 hours post-infection when the infection has already disseminated to other organs (data not shown). Recent studies show the preference of *Y. pestis* for the immune cells in mice and point to a direct neutrophil and macrophage inactivation through targeting of these cell types preferentially during infection. Further, although these studies were not pulmonary infections, the persistence of plague bacilli in macrophages later in infection may suggest that these cells are the vehicles which *Y. pestis* uses to escape the lung environment and cause systemic disease. This model is now in standard use in the Western Regional Center of Excellence Small Animal Model Core.

Amplification enables *in vivo* expression profiling.

In developing the mouse model of aerosol plague infection it was determined that infecting with 5000 cfu was sufficient for initiating fulminate pneumonic and subsequent systemic disease which resulted in death 3-5 days after infection. Given the bacterial burden at day 1 post-infection was 10^5 bacteria, it was determined that there

was insufficient amounts of bacterial RNA for microarray experiments. Furthermore, my current inability to separate the bacteria from the host tissue made enrichment of the bacteria prior to isolation not possible. To address this issue in a manner similar to previous studies, I used 10 animals at early time-points and pooled the resultant RNA material for each time-point. Enrichment of the bacterial RNA after total extraction was performed using commercial oligo-capture methods (MicrobEnrich, Ambion). With these pooling and enrichment steps only 53.5% and 83.4% of the genes in the genome were found to be expressed at days 1 and 2 respectively (Table 7). Given previous studies that describe higher percentages of the genome utilized at any one time (169) and that most of the genome is transcriptionally active *in vitro*, we felt that the data at days 1 and 2 post-infection may be incomplete in their representation of the transcriptional profile. To address this I employed a method of bacterial mRNA amplification.

In Vitro			In Vivo				
Temp. Shift Data Set			Pooled Data Set		Amplified Data Set		
	#Genes Detected	% Genome Detected		#Genes Detected	% Genome Detected	#Genes Detected	% Genome Detected
27°C	3910	97.4	Day 1	2147	53.5	3951	98.4
37°C 0.5 hr	3716	92.5	Day 2	3350	83.4	3956	98.5
37°C 2 hr	3708	92.3	Day 3	3471	86.4	3956	98.5

Table 7 Amplification allows detection of more of the transcriptome.

Describes the number of genes and percentage of the genome detected in both *in vitro* and *in vivo* experiments. Of note is the increase in the percentage of the detected transcriptome following amplification from *in vivo* samples.

Bacterial mRNA amplification utilized the template-switching mechanism of viral reverse transcriptase to add a T7-promoter to the 5' end of a transcript. This addition only occurs when a RNA transcript is primed at its 3' end for standard initiation of reverse transcription. Thus when this method is used in conjunction with genome-directed primers (170) bacterial transcripts can be selectively amplified from a mixture of bacterial and eukaryotic transcripts. Furthermore, the sequence specific probes on the microarray provide a filter for specific detection of bacterial transcripts. One possibility for the decrease detection of *Y. pestis* transcripts in pooled samples is the loss of transcripts as a result of RNA enrichment procedures. To avoid this possibility, I amplified plague transcripts directly from individual samples without additional processing.

The resultant amplified RNA from individual mouse samples were used in microarray hybridizations. The percent of the genome detected in mRNA (98.5%) is more representative than the corresponding pooled samples (53.5%-86.4%). To further investigate if the decrease in genes detected by the pooling procedure was due to the low abundance of the genes not detected or due to an inherent bias in the purification system I compared the data collected from post-infection day 1 animals. The data collected by pooling was normally distributed among the amplified data, suggesting that there was a bias in the bacterial mRNA enrichment process resulting in the loss of these genes. This could be attributed to binding of bacterial RNA to the capture oligos provided or the binding of bacterial RNA to eukaryotic RNA that was removed. This

data suggests that un-manipulated RNA samples subjected to *Y. pestis* directed mRNA amplification provide the most complete global expression profile.

To test the amplified *in vivo* expression profiles I performed quantitative real-time PCR of amplified RNA samples. Using day 1 as a reference time-point, 10 genes were selected which demonstrated a 2-fold change relative to day 1 at either day 2 or day 3, for a total of 20 data points. Comparison of the microarray and Q-PCR demonstrated a concordant gene expression vector in 80% of the genes (Figure 20).

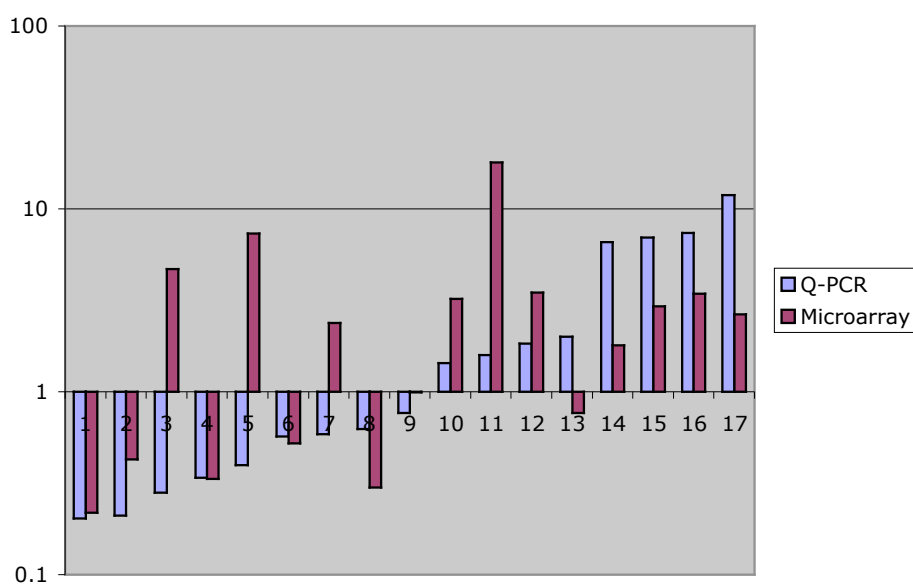


Figure 20 Validation of *in vivo* array expression data with QRT-PCR

The bar graph represents the expression values obtained from microarray experiments and quantitative RT-PCR. There is 80% concordance in the direction of gene expression change.

Global comparison of *in vitro* and *in vivo* expression profiles

Initial global analysis of the *in vitro* and *in vivo* data revealed significant differences in expression patterns. Clustering all *in vivo* and *in vitro* data revealed two distinct condition clusters corresponding to the source of the sample (Figure 21). Furthermore, when these data are clustered by the gene expression values there are distinct patterns of discordant gene expression. Surprisingly, the *in vitro* samples grown at 37°C did not cluster any closer to the *in vivo* samples than did those grown at 27°C.

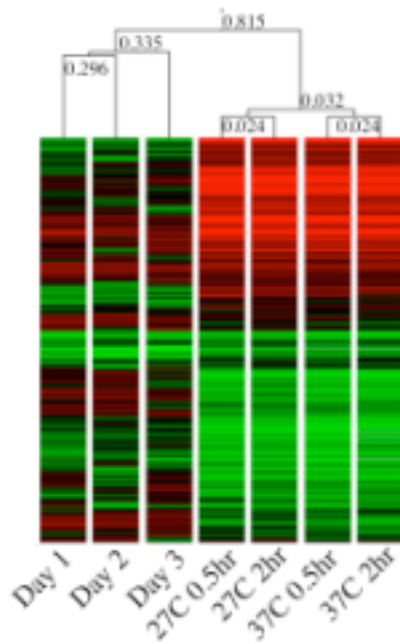


Figure 21 Cluster analysis of *in vivo* and *in vitro* expression patterns of *Y. pestis* .

Clustering of expression patterns from *in vivo* and *in vitro* expression analysis. The Pearson separation ratios are given at corresponding branch points.

A confounding variable in the comparisons of these two data sets is the compression of the gene expression data following amplification. I have noted that following amplification the signal distance between any two genes is reduced, however, the rank-order of these genes are maintained in the population. To minimize these effects on the comparison of these data sets, I evaluated the Spearman correlation between the *in vitro* and *in vivo* data sets. Using the 27°C time-point as a reference the *in vivo* time-points showed a correlation value of 0.2, 0.16, and 0.23 for post-infection days 1-3 respectively. Further, analysis proceeded using the Z-score associated with each gene (169). This measurement allows the expression of each gene relative to its deviation from the mean expression value for an experiment. This enabled the comparison of these data sets in subsequent analysis.

By ANOVA I identified 942 genes with significant differences between *in vivo* and *in vitro* environments. This corresponds to 23% of the *Y. pestis* transcriptome. These genes were equally distributed between all the genetic elements excluding the smallest plasmid pPCP which had no genes significantly changed between environments. Among the 942 genes were the *caf* operon (sans *cafIR*), 7 of 23 members of the *ysc* operon, *lcrV*, *yopE*, K and H. These genes had previously been noted as related to infection (42, 134).

Regulation of *Y. pestis* gene expression during pulmonary infection

I investigated the regulation of gene expression of *Y. pestis* during the time course of infection. Using ANOVA I did not identify any genes with significant expression differences between days 1 and 2. However, 68 genes were identified that demonstrated significant differences from either days 1 to 3, or days 2 to 3 (Figure 22). This represents a small portion of the transcriptome relative to the differences observed between *in vitro* and *in vivo* conditions. My analysis did not reveal any members of the classically described virulence operons of *Y. pestis* such as *caf*, *ysc*, *hms*, or *yop* genes. This suggested to me that although many of these genes were identified when comparing the *in vitro* and *in vivo* environments, that their temporal regulation in the lung has been set by or prior to day 1 post-infection. Most of the genes that appeared to be temporally regulated *in vivo* suggest annotated functions more closely associated with carbohydrate or nutrient metabolism, transport, and acquisition, along with some heat-shock proteins.

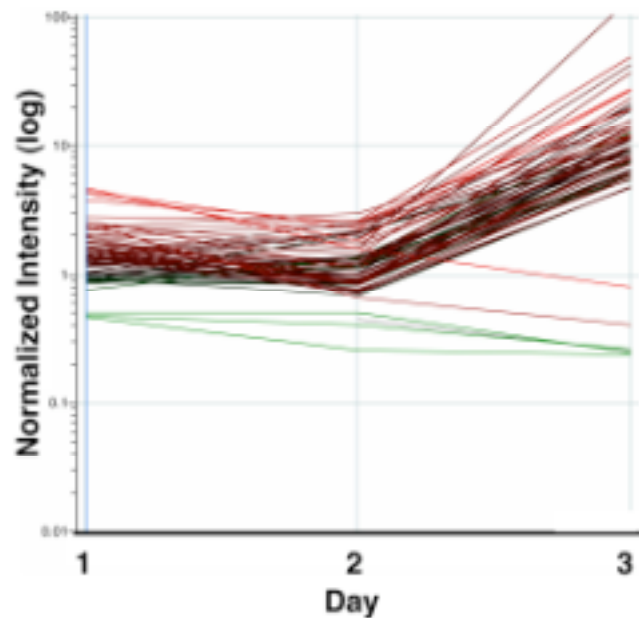


Figure 22 Temporal *Y. pestis* gene expression changes from mouse lungs

Most changes in gene expression during *in vivo* growth in the mouse lung occur between days 2 and 3. The 62 genes identified through ANOVA to be differentially regulated *in vivo* are graphed over the time course studied.

Preferential *in vivo* expression of chromosomal pathogenicity island (PI) and PI-like regions.

I investigated the expression patterns of genomic regions with altered %GC content to test the hypothesis that these pathogenicity-island (PI) or PI-like regions provide functions required for *in vivo* growth. 19 genomic regions were identified in the published annotation of the CO92 genome that demonstrated features of pathogenicity islands, including altered %GC (130). These regions were individually

analyzed by comparing the expression levels of the genes within each region from both the *in vitro* and *in vivo* studies. Again using Z-score analysis to compare the *in vivo* and *in vitro* expression patterns we observed a distinct tendency for enrichment of genes preferentially expressed *in vivo* in these islands. An example of a predicted pathogenicity island where a subset of the genes are preferentially expressed *in vivo* is illustrated in Figure 23.

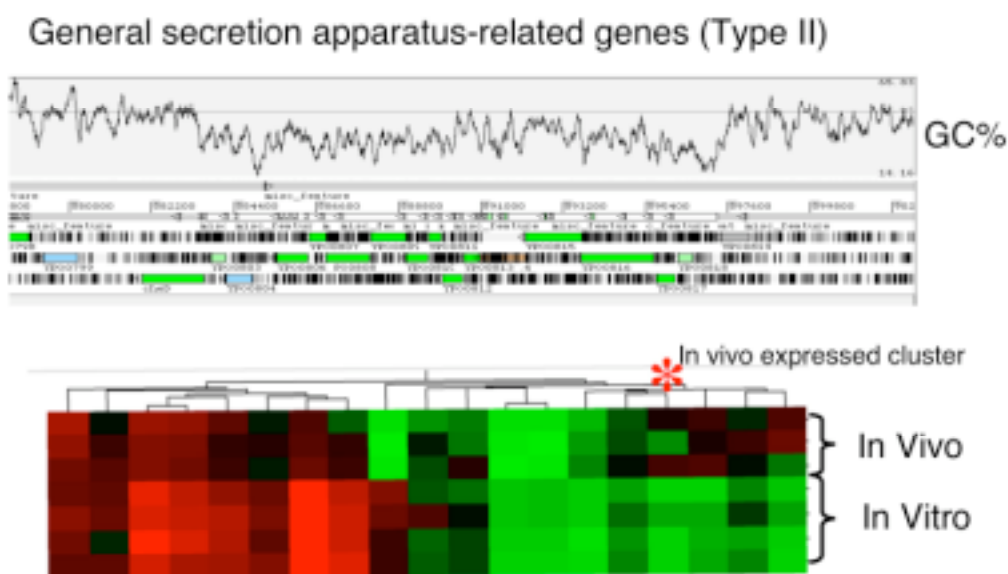
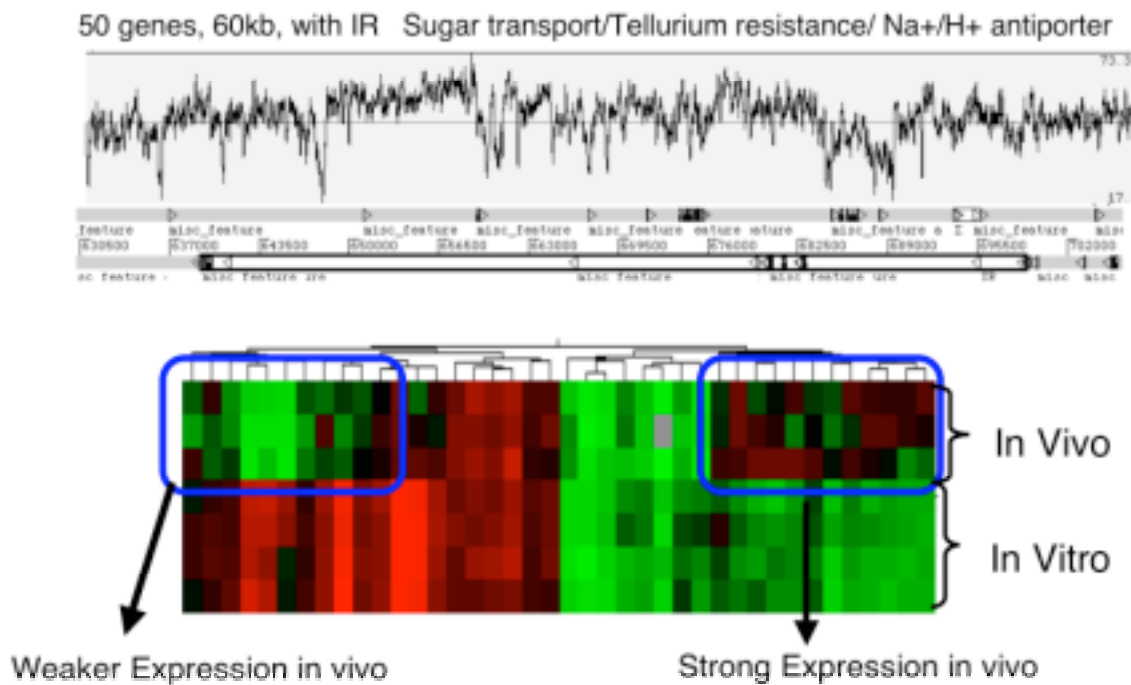


Figure 23 In vivo expression of a gene cluster within a predicted pathogenicity island.

The *in vivo* and *in vitro* conditions cluster independently (horizontally) and reveal six genes that are more prominently expressed *in vivo* (vertical cluster denoted by red asterisk mark). The genomic island is predicted to encode for a type II secretion apparatus and may be utilized for secretion or adherence.



A large annotated genomic island shows two distinct regions (highlighted in blue boxes) that are both repressed *in vivo* or up-regulated *in vivo*.

assumptions that protective antigens must be expressed *in vivo* as well as localized to the outer-membrane (Figure 25).

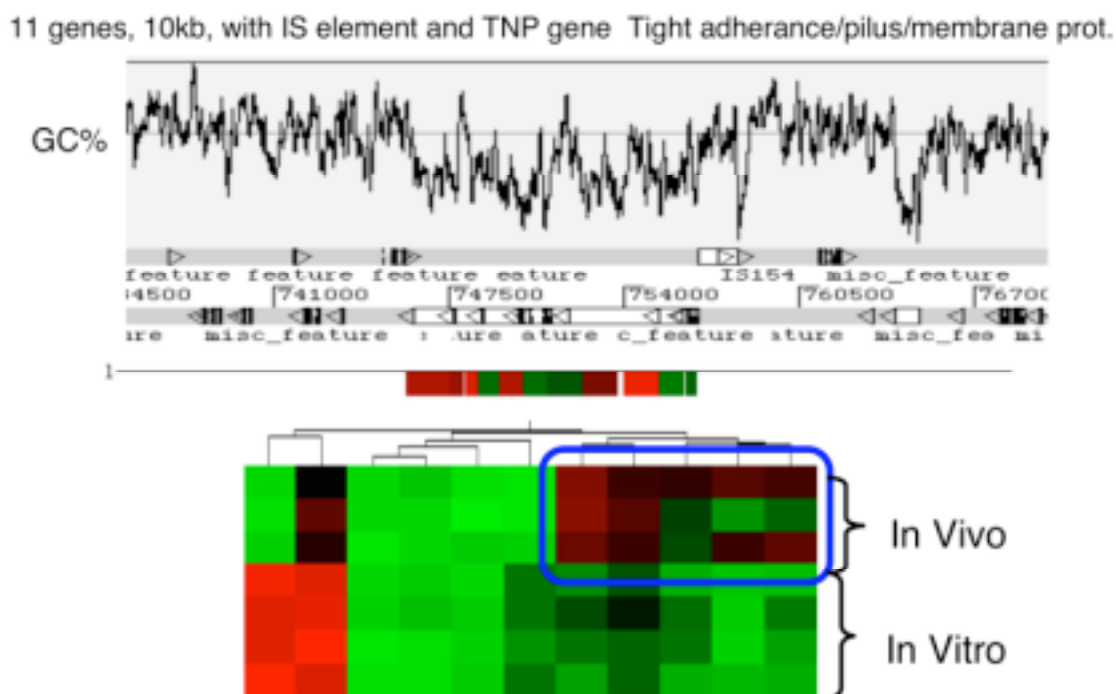


Figure 25 Pathogenicity island with predicted outer-membrane proteins expressed *in vivo*.

The cluster of five genes highlighted in the blue box encode for predicted outer-membrane proteins and are more prominently expressed *in vivo* relative to the *in vitro* model.

A region identified in this study, not predicted to be a pathogenicity island during the original annotation, was discovered through visual scanning of %GC plots and genomic features. This region encoded genes utilized in outer-membrane synthesis and is upstream of an invert-repeat sequence. The region is shown (Figure 26) where in addition to our data we have cross referenced genes identified in a signature-tagged

mutagenesis screen performed in *Y. pseudotuberculosis* (93). Of the 26 genes identified though this screen, using an intraperitoneal infection model, 5 of these genes are located within my region of interest. This finding further supports the use of altered %GC regions within the genome during infection within mice.

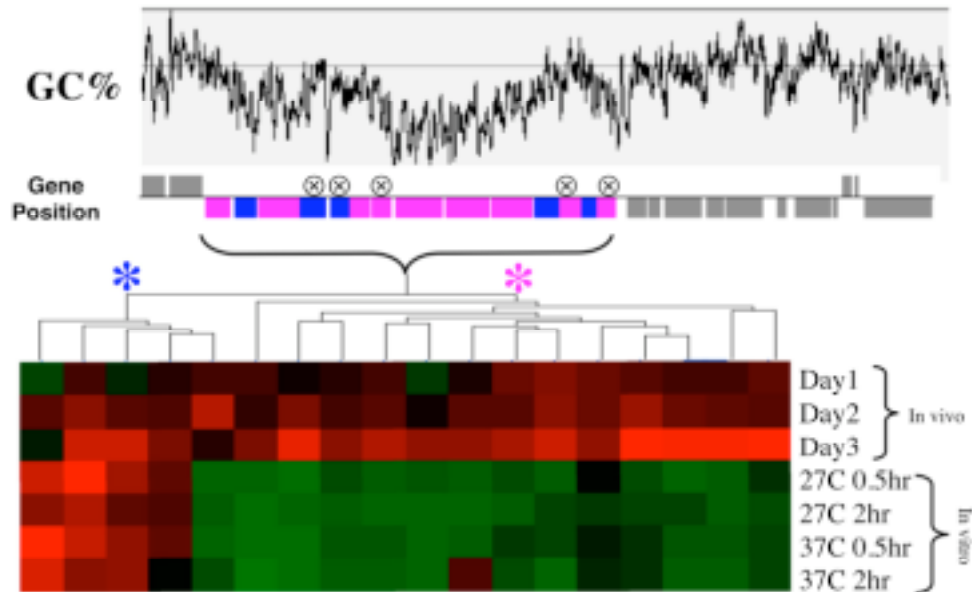


Figure 26 *In vivo* expressed genomic island of *Y. pestis* .

Identification of an *in vivo* regulated PI-like genomic region

The %G/C plot and gene structure is shown for a newly identified possible pathogenicity island. The expression levels for the genes during both *in vitro* and *in vivo* conditions clustered and form two distinct groups (horizontal brackets) related to either *in vitro* or *in vivo* growth. The two primary vertical clusters (tree shown on top of cluster) represent either constitutively expressed genes in either growth conditions (indicated by the purple asterisk) or genes preferentially expressed *in vivo* (indicated by the blue asterisk). Genes previously identified through a Signature-Tagged Mutagenesis screen in an intra-peritoneal infection model of *Y. pseudotuberculosis* are identified by the crossed-out circles (⊗).

CHAPTER FIVE

Discussion

LAPT Discussion

Here I describe a technique that can be applied to amplify small amounts of prokaryotic mRNA in mixed samples. I show that the LAPT procedure amplifies sense-stranded RNA and have applied this technique to directly amplify prokaryotic RNA. I have used this process to selectively amplify the prokaryotic RNA population out of a complex mixture of prokaryotic and eukaryotic RNA. This RNA was used on microarrays to demonstrate that the biological information in the original sample was largely kept in tact.

Recently, GDPs and genomic normalization were applied to map the transcriptional profile of *Mycobacterium tuberculosis* (*Mtb*) from the lungs of mice(171). However, it is worth noting that the earliest time-point evaluated was 7 days post-infection and required the pooling of 50-100 animals at each time-point. LAPT should enable similar analysis with an advantage of examining samples at time-points closer to the initiation of infection with fewer animals. LAPT should also reduce, if not eliminate, the need for the pooling of multiple experimental samples. Furthermore, pathogens not amenable to enrichment procedures should be approachable by LAPT. Of potential benefit is the ability to study expression profiles of pathogens in the context

of the un-manipulated host environment without artificially increasing the number of pathogens present. This benefit was demonstrated in the *in vivo* expression study of *Yersinia pestis* where sample manipulation negatively affected the ability to measure global transcriptional activity.

Similarly to GDPs, which selectively prime prokaryotic RNA, using this method on a mixed RNA source for microarray studies is particularly useful. The directed priming of prokaryotic transcripts by the GDPs biases the T7 promoter addition to targeted prokaryotic transcripts. Additionally, the specificity conferred by the sequence specific probes on the microarray favors detection of prokaryotic signals. Therefore, even in the event of spurious priming by the GDPs of non-target RNAs, the amplification and labeling of these sequences should not be detected. Hence, combining LAPT with microarray analysis provides two filters for specific detection of transcripts and should enable the expression analysis of samples that has previously not been possible.

Combining LAPT and expression analysis to study microbial pathogenesis in the context of a host should provide advantages over current methodologies. Methods such as signature-tagged mutagenesis (STM)(79), *in vivo* expression technology (IVET) (114), or differential fluorescence induction (DFI)(179) all provide insights into gene expression *in vivo*, but are limited by the experimental biases introduced. This in turn limits the breadth of questions that can be addressed by these methods. All three of these examples require *ex vivo* genetic manipulation of the pathogen, and can strongly bias results depending on the temporal expression of a gene *in vivo*. Global gene

expression analysis by microarrays presents the best opportunity to analyze *in vivo* gene activity without introducing non-natural selection bias. LAPT should address the primary limitation of microarray analysis for *in vivo* pathogenesis studies, RNA abundance.

An essential quality for any amplification process is the preservation of the biological information of the original sample. The data presented here suggests that this amplification process has a relatively small affect on the biological information of the original sample. Microarrays are used to identify patterns and changes in gene-expression that are biologically relevant. This is commonly done using various cluster methods and Z-score analysis (52, 174). The use of Spearman correlation to evaluate the maintenance of biological information was used with these downstream analyses in mind. Based on a scale from -1 to 1 , any positive Spearman correlation suggests maintenance of rank-order. Whereas, the focus on rank-order may preclude quantitative assessment of individual gene changes using the array data, the global evaluation will be valid and the concordance of gene change for any one individual gene is largely maintained as demonstrated by the QRT-PCR results. Furthermore, any validation of quantitative gene expression changes requires the use of an alternate method, such as northern blots or QRT-PCR. LAPT does not interfere with this process. Additionally this amplification technique enables multiple replicates from limited amounts of starting material to improve statistical power.

The ability to add the T7-promoter to the 5' end of RNA is independent of the position or size of the primer utilized by the reverse transcriptase. Shown using the

single GFP RNA, this demonstrates the primary advantage of LAPT over other amplification strategies where the means of amplification (e.g. T7 promoter addition) is coupled to the priming of the RNA. Uncoupling these events provides a number of benefits. First, the issue of poor RNA priming due to modification of the reverse transcriptase primers is not present. Second, LAPT does not restrict the modification to the 3' end of the RNA. Finally, LAPT does not require a uniform priming location, such as a poly(A) tail. These features enabled us to amplify a mixture of heterologous bacterial RNA.

This technique should be applicable to any pathogen or microbe. As a tool, microarrays have enabled countless studies, however, continue to require independent validation (northern blot or QRT-PCR). LAPT expands the utility of this research tool and it may allow the analysis of pathogen gene expression at any time during the infection process from most tissues. Thus LAPT facilitates investigation of pathogenesis as well as provide a means to study expression analysis of mixed microbial populations.

Future directions

As a molecular biological technique LAPT has the potential to be utilized in a variety of systems. Outside of *in vivo* expression studies for which I applied the technology, LAPT may prove to be an ideal technology to begin approaching global expression studies of mixed microbial communities. These could be environmental

samples (soil communities), or in the context of a mammalian host such as the gut microbiota. Where genome-directed primers may not provide the level of directed priming required for more complex mixtures, the use of longer gene specific primers for large numbers of genes may overcome this limitation. Furthermore it may be possible to design redundant probes for gene families within a microbial community to evaluate community-wide metabolic gene expression. Although, amplification may not be required for these studies it may be useful to rapidly amplify a subset of genes out of a large sample in order to reduce the background noise from other gene products.

The LAPT mechanism may be a valuable tool as an *in vitro* transcription/translation tool to study proteomics. Generally, cloning of a gene into an expression vector is required prior to *in vitro* transcription translation. The ability of LAPT to either specifically or globally add a promoter to the 5' end of RNA enables the potential to use this to produce RNA for translation in the absence of cloning. This was done while evaluating the protocol using the GFP gene. LAPT-amplified RNA from this gene was subjected to *in vitro* translation and the resulting mixture produced functional GFP protein. This may be used on a total RNA sample from a cell population to produce a synthetic proteome *in vitro*. The advantage of this approach to proteomic studies instead of a direct analysis of the protein fraction would be the ability to apply the other molecular biological techniques available to nucleic acids that are not available to proteins. One such technique would be subtractive hybridization to normalize the abundance of RNA sequences in a sample. This may overcome a fundamental problem in proteomics, being the vast differences in protein abundance

leading to the inability to analyze (by mass-spectrometry for example) low abundant proteins in complex samples.

LAPT may also prove to be a useful screening tool. The use of small RNAs to knock down the transcriptional activity of the cognate gene has been established as a useful technique. It has further been described as a valuable mechanism employed by cells to regulate gene expression. The application of LAPT to amplify the population of small RNAs from a cell under stress, such as viral infection, may enable a better understanding of the diversity of this mechanism.

Another future direction for the improvement of this technology is the screening of MMLV mutants (through gene-shuffling or random mutagenesis) for variants with improved template-switching activity. Additionally a systematic study to discover and/or eliminate any bias in this process would be useful. My data analysis did not identify any clear template-switching bias, based on copy-number, transcript size or %GC content. However, this may be explored and potentially translated into improved performance of this technology.

In general the template-switching mechanism can be used to add any number of sequences to the 5' end of a RNA sequence during reverse transcription. The coupling of this fact with commonly used lab sequences such as the T7 promoter described is only one of many that could be utilized. Others include alternate promoter sequences or unique sequence tags. It is my hope that this technique will continue to be used to study microbial communities in natural environments. Science has much to gain from the

study of bacteria outside of culture flasks and Petri dishes, and LAPT should enable many of these studies.

***In vivo* Expression Analysis Discussion**

In this study I describe for the first time a microarray analysis of the gene expression of *Y. pestis* in the pulmonary model. Further I demonstrate that amplification of the RNA from individual samples is able to provide sufficient material for multiple microarray experiments. This technique combined with both the preferential priming of pathogen transcripts using genome-directed primers and the specificity of the oligonucleotide probes on the microarray has enabled this study. The *in vitro* model of temperature dependant changes in *Y. pestis* demonstrated the level of transcriptional regulation mediated by a single environmental factor. Using cultures grown on plates containing bacteria in many stages of growth, I validated our arrays and protocols and found that simply shifting temperature from 27°C to 37°C altered gene expression in 2.4 to 3.1%, of the genome. A pulmonary model of *Y. pestis* infection in mouse was developed. This allowed analysis of gene expression by microarrays *in vivo*. Conditions were determined that permitted amplification of pathogen mRNA at early phases without loss of representation of genes. I find that ~23% of the genome has significantly different expression between *in vitro* and *in vivo*. Interestingly, there is little change in gene expression between days 1 and 2 post infection. Many of the genes preferentially expressed *in vivo* map to islands of base composition differences.

The absence of *sigE* up-regulation in response to increase temperature suggests an alternate role for *sigE* in *Y. pestis* when grown on plates. The only sigma factor up-regulated by temperature was a *Y. pestis* protein with high similarity to *E. coli rpoD*, described as a general sigma factor. This suggests either a unique role of *rpoD* in *Y. pestis* or the requirement of another environment signal to stimulate alternative sigma factor activation.

Given that the culture media contained Ca^{2+} many of the described *ysc* and *yop* genes showed little transcriptional regulation. This reinforces that a Ca^{2+} depleted environment and not only temperature are required for their expression (134). This known phenomenon highlights the ability of *Y. pestis* to assimilate multiple environmental signals in coordinating gene expression.

The transcriptional profiles *in vitro* and *in vivo* are substantially different. Surprisingly there was no closer correlation of the 37°C *in vitro* versus the 27°C data to the *in vivo* data sets. This may have been in part due to the short time-course of temperature shift relative to the initial lung sample at day 1. Additionally, temperature alone may induce a transcriptional program of bacterial survival in culture as opposed to a pathogenic transcriptional program. Although the compared data sets were derived from un-amplified or amplified *in vitro* or *in vivo* samples respectively, the consistency of change in the rank-order of gene expression supports these conclusions. Given that the expression data from all samples demonstrated characteristics of normal distributions we applied Z-score analysis to assist in direct comparisons of the data sets.

Of note was the number of genomic regions with altered %GC that demonstrated an overall greater expression level in the lung compared to the *in vitro* studies (either 27°C or 37°C). The finding that 5 of the 26 genes identified through a signature-tagged mutagenesis screen in *Y. pseudotuberculosis* appeared within one of these regions is suggestive that these may represent a general genomic motif for *in vivo* expression. Furthermore, the encoding of genes involved in capsular synthesis within this genomic region supports the hypothesis that these regions play specific roles in adapting to host environments.

The relatively small amount of significant gene expression change from day 1 to day 2 post-infection was unexpected. This observance can be explained in part if most changes are elicited early on as the pathogen interacts with the new host environment. It may be that by 24 and 48 hours post infection the expression of *Y. pestis* in the lung has stabilized. Alternatively, there may be multiple niches within the lung itself with varying levels of bacterial growth, CO₂ tension, host-defenses, and other environmental signals. The collection of the entire lung tissue for analysis may have diluted significant alterations in the gene expression of *Y. pestis* within different niches. The observance of many genes increasing their expression levels at Day 3 post-infection that are related to bacterial growth, division and nutrient acquisition may be a reflection of the large increase in the number of bacterium. Given that the entire lung was used in the preparation of the RNA for these experiments and the observed dissemination at these time points, a portion of these bacteria may have been located in the blood and not been confined to the lung.

In conclusion this study demonstrates a protocol for monitoring the expression profile of a pathogen *in vivo* even at relatively low levels in the host. The host lung in primary pulmonary *Y. pestis* infection is the initial site that the pathogen must escape prior to dissemination and disease progression. There are many environmental factors within the host lung that influence the pathogens' expression profile. These transcriptional alterations may not represent specific pathogen virulence mechanisms, but may serve as antigens. These antigens may provide vaccine targets; conversely, during infection of a naïve individual these antigens may play a role in misdirecting the host immune response to targets that may not be present in the disseminated pathogen. Vaccination with the proteins corresponding to the genes, such as the large number of genes expressed from altered %GC regions, found in the lung environment may provide an effective means to prevent or slow down the dissemination of *Y. pestis* to secondary sites of infection during primary pulmonary exposure to the plague. This work certainly calls for further technological advances that would allow looking at gene expression immediately after infection.

Future Directions

In these studies lung was the only tissue evaluated although *Yersinia pestis* disseminates to other tissues by day two post infection. Evaluation of the plague expression profile within these organs may yield unique insights into its pathogenesis. Of greater interest may be expression within specific immune cell types that may help

better understand the initial immune evasion of *Y. pestis* , specifically in the lungs. This could be done *in vitro* using cultured immune cells or from primary cells following a cell enrichment protocol such as FACS or MACS. As suggested by my results, the dynamic transcriptional changes that occur happen early in infection. Analyses of these time-points were preempted by the number of bacteria used to infect the mice and the limitations of the LAPT amplification protocol. Adjusting the infection inoculum or pooling of multiple samples may allow the temporal resolution of the transcriptional changes I observed.

There has been a long-standing desire to correlate the *in vivo* expression profile of a pathogen to the identification of vaccine targets. It is unclear if approaching this task by designing hypothesis based on transcriptional data is the best approach, as global screens for vaccines are developed. This would make the retrospective analysis of temporal expression profiles for protective antigens a simpler task to ask if patterns are present.

APPENDIX I

***Bacillus anthracis* Microarrays**

Objective summary

Similar to the objective and efforts described for the *Yersinia pestis* microarray, the *Bacillus anthracis* microarray resource was created for similar purposes. This resource has been used in a variety of projects and continues to be a valuable scientific resource. The microarray has been used in comparative genomic studies in *Bacillus sp.*, *in vivo* expression studies of *Bacillus anthracis* in a mouse model of infection, functional annotation of the *B. anthracis* genome, and global expression patterns of *B. thuringiensis* during sporulation.

Another objective of this project was the application of the process of microarray design, synthesis and validation as set forth in the *Y. pestis* project. This genome presented an additional challenge in that the annotation of the *B. anthracis* genome had not been made available. In summary this projects goal was to rapidly develop a pathogen microarray resource in the absence of an annotated genome.

Introduction

Bacillus anthracis is a gram positive bacillus, and is the causative agent of anthrax (144). The diseases of *B. anthracis* infection include a cutaneous,

gastrointestinal, and a pulmonary etiology (106). Of greatest concern is the pulmonary disease form due to its high mortality rate. Anthrax is often referred to as wool-sorters disease due to the frequency of pulmonary anthrax infections in those working on animal hides contaminated with the spores of the pathogen. The rapid pathology of the pulmonary form of disease is initiated when *Bacillus anthracis* spores enter the lung where they subsequently germinate to vegetative bacilli. They then reach the mediastinal lymph nodes where a combination of virulence mechanisms results in fatal disease if medical intervention is not taken quickly. Virulence mechanisms include an antiphagocytic poly-glutamic acid capsule and the secretion of two toxins (EF –edema factor, and LF – lethal factor) along with a commonly used transporter (PA –protective antigen)(144). Together these virulence mechanisms result in the rapid progression of systemic disease, vascular breakdown, immune cell destruction, and ultimately death of the host (106).

Use of this pathogen as a biological weapon has a long history with references to potential use in the Bible, and as recently as 2001 in the United States. Similar to the goals of the *Yersinia pestis* array, we aimed to generate a resource that would allow use to better understand *B. anthracis* pathogenesis and serve as a means to develop novel immunological therapies.

Construction of a Microarray for the Anthrax bacillus, *Bacillus anthracis*

The purpose of this project is to develop an oligonucleotide microarray for all genes of *Bacillus anthracis* (BA). This resource will be used to evaluate global gene expression changes of the bacteria.

Genome annotation

The initial source of the Sequence data was the TIGR unfinished genomes database. Subsequently, the genome and annotation was published and posted to the NCBI public databases

The annotation of the chromosome was not available at the initiation of this project. In order to get a preliminary annotation for the chromosome the Glimmer program was used (155).

This program was run on the complete chromosome sequence from TIGR. The program identified 6021 possible genes. The gene sequences were extracted using the extract sequence program included in Glimmer. Each predicted gene was given an identifier based on the glimmer output file. For example for glimmer predicted gene 1 was given the name BA1. All of the information from the glimmer program such as gene location, direction and length were all imported into a FileMaker Pro database.

The subsequent publication of the annotation revealed that I had annotated 821 genes that were not predicted in the published annotation(144). As I had made the gene

finding parameters loose as to not miss any potential genes this was not unexpected. I replaced my gene naming nomenclature with that of the published annotation and for the remaining 821 glimmer predicted genes not found in the published annotation, these were retained in the analysis and prefixed with “glimmer-“ for future analyses.

The two plasmids of BA have been published and annotated pX01 and pX02 (124).

Information was obtained through the NCBI web site:

http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/eub_p.html

Bacillus anthracis virulence plasmid PX01, complete sequence :Accession: [NC 001496](#)

Bacillus anthracis plasmid pX02, complete sequence : Accession: [NC 002146](#)

Probe Design

We used a program ProbeSelect to design gene specific probes (104). This program was requested and received via e-mail.

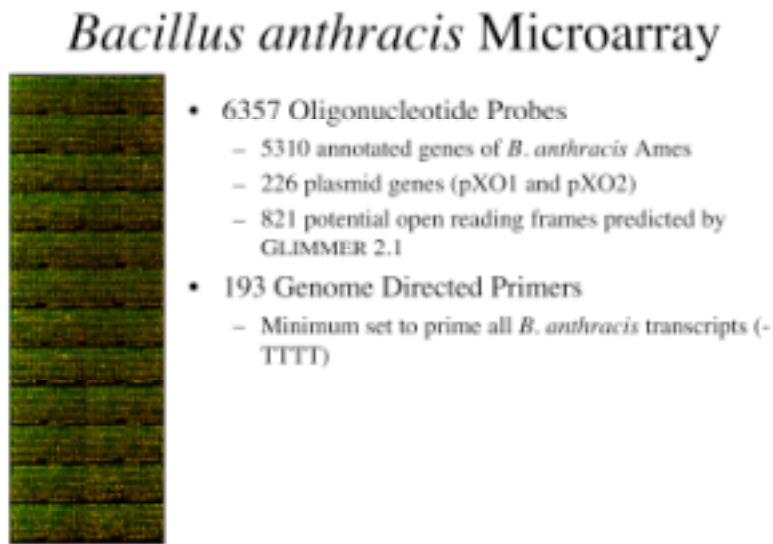


Figure 27 Summary of *B. anthracis* microarray features.

An image of a typical *B. anthracis* microarray is shown, along with the number of 70mer probes.

Probe Synthesis

Long-oligonucleotides were either synthesized or purchased. Handling and processing of the oligonucleotides was done in the same manner as described for the *Y. pestis* microarray (see Figure 3 chapter 2 Material and Methods).

Microarray validation

The specificity *B. anthracis* microarrays were validated by evaluating the hybridization of labeled gene specific PCR products to an array demonstrated in Figure 28 Specificity of *B. anthracis* array., and through hybridization of total labeled genomic DNA shown in Figure 28.

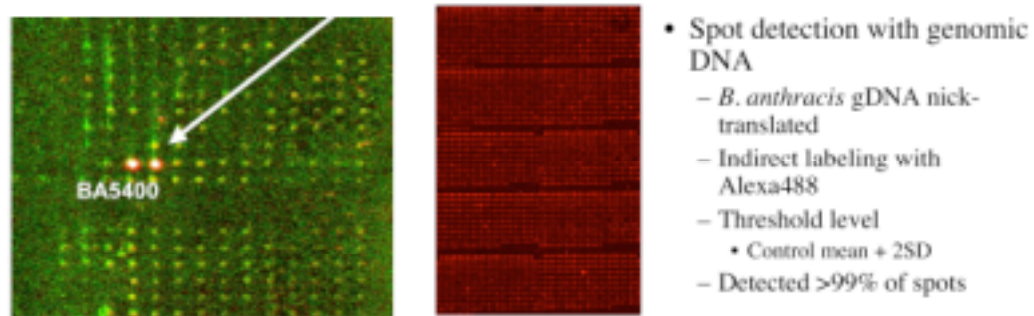


Figure 28 Specificity of *B. anthracis* array.

The specificity of the *B. anthracis* array was evaluated using labeled *B. anthracis* PCR gene products and labeled gDNA.

The microarrays were also tested for their ability to provide reproducible hybridization signals by comparisons of two replicate experiments where RNA from *Bacillus thuringiensis* was used as the analyte (Figure 29).

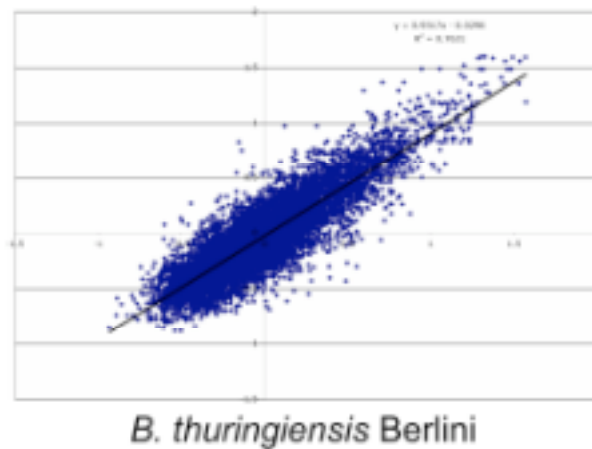


Figure 29 Reproducibility of *B. anthracis* arrays using a closely related *Bacillus* species.

RNA from *Bacillus thuringiensis* was used to evaluate the reproducibility of the *B. anthracis* arrays. The R value for this experiment was 0.87.

Summary

My initial efforts to use these microarrays to study the expression profile of *B. anthracis* from the lungs of infected mice were not successful. Even with the addition of amplification the samples we had available were not able to provide clean and reproducible expression patterns. The reason for this was multi-fold. First and foremost, bacillus infection via the pulmonary route does not result in a bacteria pneumonia, but rather migrates to the mediastinal lymph nodes where vegetative growth occurs. The collection of the lung tissue for RNA isolation did not provide enough bacterial RNA for analysis. Other reasons included the availability of appropriate control samples, which was later addressed by others using this resource.

Following completion and validation of this resource, it has been employed in a variety of studies. Initially, the microarrays were used by a fellow graduate student to investigate the expression changes of *B. subtilis* during growth and sporulation. The purpose of this work was two-fold and was primarily to establish protocols for RNA extraction and labeling in *Bacillus sp.*, and also to provide baseline data for future work. The results of the experiments conducted highlighted the need for mechanical disruption of the bacterial pellet to obtain efficient RNA extraction.

Later the resource was transferred to a post-doctoral fellow who conducted expression studies in an avirulent *Bacillus anthracis sub. sp.* looking into both *in vivo* expression patterns as well as expression from un-annotated regions of the genome. The initial annotation I performed and used to generate the array provided over 800 un-annotated possible open reading frames. These probes provided data on these un-

annotated regions and were classified as either intergenic ORFs, Orfs found on the opposite strand of annotated CDS, or short over-lapping putative ORFs with annotated genes. Expression data from these spots provided leads to explore novel bacillus transcripts.

Finally, this microarray has become a standard resource in the Western Regional Center for Excellence in Biothreat Research. The *B. anthracis* microarrays have been used by multiple labs for a variety of expression studies.

APPENDIX II

Functional Annotation through hybridization of transcribed products

Introduction

Currently, most genomic sequence is annotated through bioinformatic means with no consideration of whether the predicted gene products are actually functionally present (23, 24, 47, 111, 155). Bioinformatic approaches often must make judgment calls based on our current understanding of biology making the results inherently biased. Current methods are also prone to miss RNA sequences that are not translated but play a significant role in the biology of an organism due to their focus on identifying classical gene structures. The vast majority of usefulness from genome sequence data is derived from its annotation. We proposed a new definition of functional annotation, as being the direct detection of transcriptional products using a comprehensive genomic screening method. Functional annotation would reduce the time needed to obtain this information and take an unbiased approach to this problem. The goal of functional annotation is the rapid annotation of genomes and consequent discovery of genes and transcribed RNA species.

This idea has been indirectly explored initially by Selinger, et al (161), who were able to detect antisense transcription from a majority of *E. coli* genes using a high-

density “genome array”. Furthermore, they were able to show examples of mapping transcriptional start and stop sites. Their approach was biased and interpreted based on computational annotation, and this method was not used to identify novel functional transcripts. Later, Tjaden et al(176) used a similar approach to identify 1102 additional transcripts found within the intergenic regions of the *E. coli* genome. A powerful demonstration of a functional assay for identifying transcribed regions of the genome. From this study they proposed 317 new transcripts with potential function as either ORF’s, sRNA, or other uncharacterized regulatory RNAs. This approach was not used as a primary annotation tool but focused on intergenic regions of the computationally annotated genome. The use of a functional assay has not been employed to annotate a prokaryotic genome to date. In contrast, the most comprehensive use of an experimental method to drive the annotation process was done in late 2003 using the plant, *Arabidopsis* genome(189). Additional studies using experimental approaches to annotate portions of the human genome have been reported(92, 149, 162). While many of these reports were not available when this project was initiated, my goal was to set out to define a system for the systematic annotation of a genome using a high-density microarrays as my functional assay. This was prompted by the idea that most high-throughput vaccine discovery methods rely exclusively on computational annotated sequence data(172), and may inherently be missing valuable and potentially immunogenic transcribed/translated regions of the genome.

The goal of my approach to functional annotation is to functionally discover genes and transcribed RNA species through the hybridization and detection of RNA or

complementary DNA derived from the target organism to designed nucleotide probe pairs. Each probe pair will represent a defined region of the genome and will be used to detect analyte derived from either strand of the DNA. Using genome sequence data this technology will be able to distinguish which areas of the genome are being used as templates for RNA synthesis.

Methods and results

Basic concept

To test this idea we adapted the methods of expression profiling arrays to provide us with qualitatively different information. As opposed to identifying the level of expression of a defined genomic sequence or gene, we sought out to identify any activity of all genomic regions. To achieve this we used a tiling approach, where probes representing both strands of the genome are made across a contiguous span of sequence at defined intervals. The span of sequence can be the entire genome or a region of interest and the intervals can be adjusted to the desired resolution.

Transcribed RNA from the organism of study, grown under a variety of conditions, would be used as a template to generate labeled analyte. This labeled analyte would then be used to hybridize and subsequently detect the cognate probe or probes that correspond to the RNAs original strand and position within the genome. The basic outline of the process is described in Figure 30.

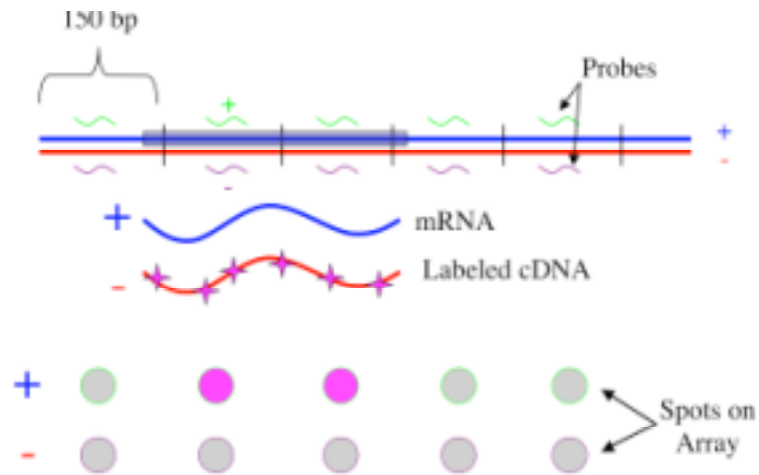


Figure 30 Diagram of the hybridization based annotation method.

Strand specific probes are used to detect transcribed RNA following reverse-transcription and labeling.

Initial Experiments in Functional Annotation

To evaluate the essence of the functional annotation concept a preliminary experiment was done using spotted oligonucleotides. A 25kb region of the *Mycobacterium tuberculosis* genome was selected randomly. This region was divided into equal 150bp segments. The internal 50 bases of each segment were selected to serve as a probe for that region of the chromosome. These probes were synthesized and subsequently spotted on poly-L Lysine coated slides at 40uM concentration in 3X SSC. These slides were then hybridized to Cy3 labeled cDNA from *Mtb* total RNA derived from 14 day cultures.

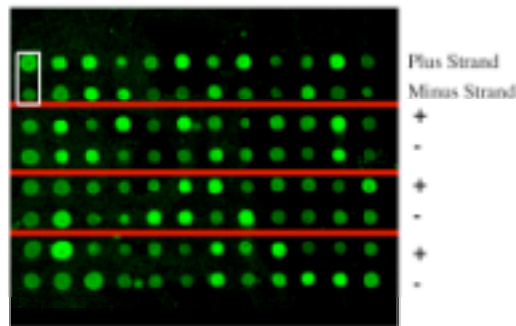


Figure 31 Example of the *Mtb* 25kb array hybridized with labeled cDNA from H37Rv grown to log phase.

The ratio of the positive and negative strand was used to correct for experimental noise and determine the strand origin of the transcript.

The data was analyzed by taking the log ratio of the positive strand probe signal over the negative strand signal. In doing this we hoped to reduce the hybridization noise associated with both background signals as well as hybridization efficiencies associated with probe properties. Therefore, if the log ratio was positive we concluded there was greater signal coming from the positive strand, and conversely if the log ratio was negative. This approach did assume that transcription was generally limited to one strand. The results of this analysis is graphed below, and showed a greater than 80% concordance with the published genome annotation

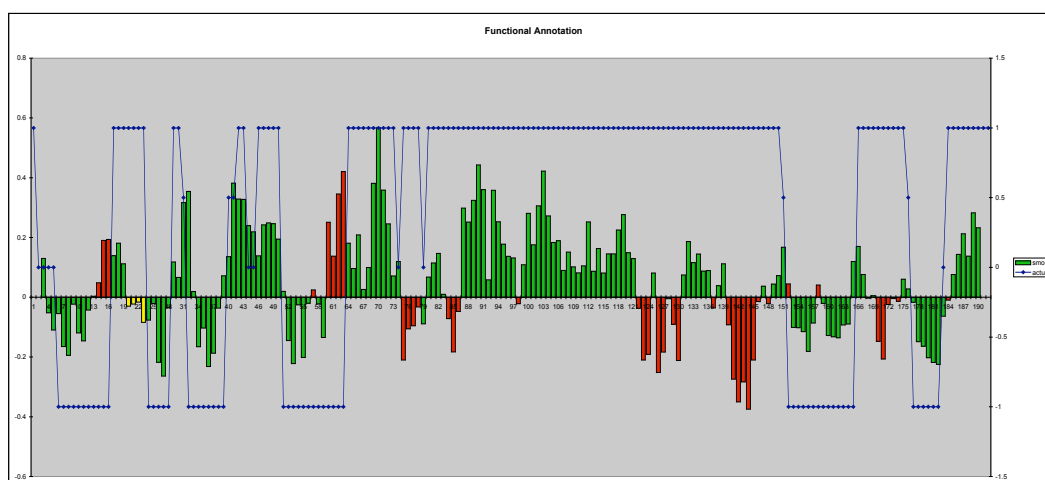


Figure 32 Graph of positive strand/negative strand probe intensities.

Each bar represents a 150bp region of the *M. tb* genome with positive strand transcription appearing >0 and negative strand transcription <0 . The blue line represents the gene location according to the published annotation. Green bars are regions where the functional and published annotation is in concordance and red bars are regions where there are discrepancies.

Conclusions of this work were that the basic concept and approach of this technology was sound. The practicality of this approach to screening an entire genome is very limiting using spotted oligonucleotides. Therefore we sought out to adopt this technology using the digital optical chemistry platform (16, 64, 164).

D.O.C. platform for functional annotation

Digital optical chemistry was developed to rapidly generate custom high-density oligonucleotides directly on a substrate using light-directed synthesis (16, 64). This platform was advantageous for this project due to its ability to rapidly produce a high-density array that was able to span the entire sequenced genome. The outline of the process described in Figure 33.

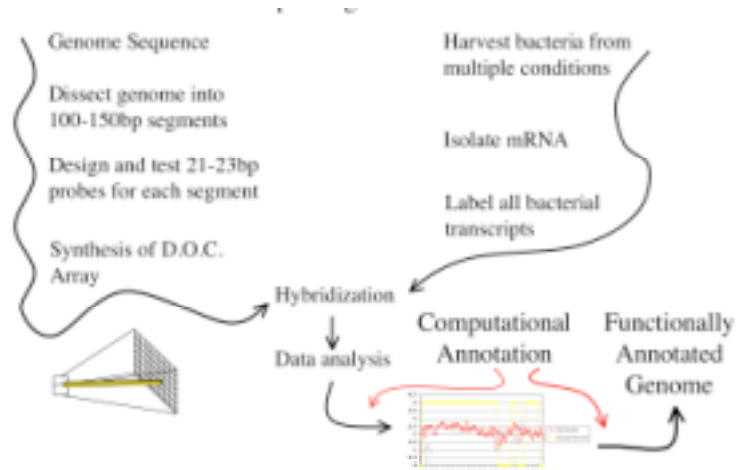


Figure 33 General schema for the functional annotation of microbial genomes using D.O.C. high-density arrays.

For subsequent experiments we moved from the GC rich genome of *M. tuberculosis* to *Y. pestis* that has a GC content closer to 50% and presented fewer probe design problems. In addition material was available for this system, and it is a pathogen of present concern.

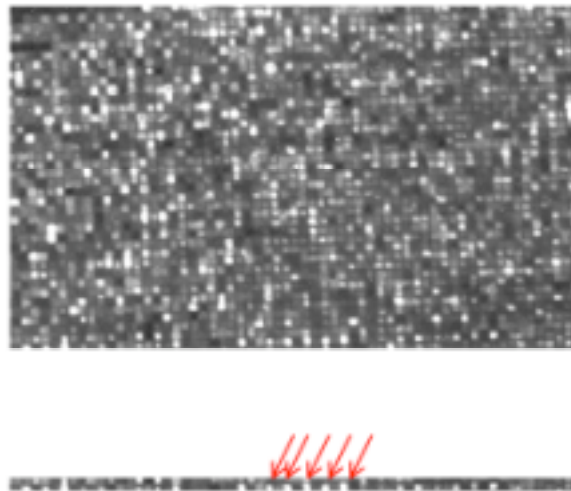


Figure 34 Signal patterns of an annotation array hybridized.

The alternating positive/negative strand probes produces an alternating signal pattern in adjacent probe sets, specific for a transcribed gene (red arrows in inset).

Focused evaluation of a functionally annotated genomic region in *Y. pestis*

I attempted to validate a region of the *Y. pestis* genome that I had functionally annotated. I had compared my empirical data with the published annotation and identified a region where I had both agreement and disagreement (Figure 35). I classified the disagreements into regions of expression on the opposite strand or expression from regions where no annotated gene was found. I then proceeded to test these using northern blots. This experiment failed multiple times resulting from technical issues and was not completed.

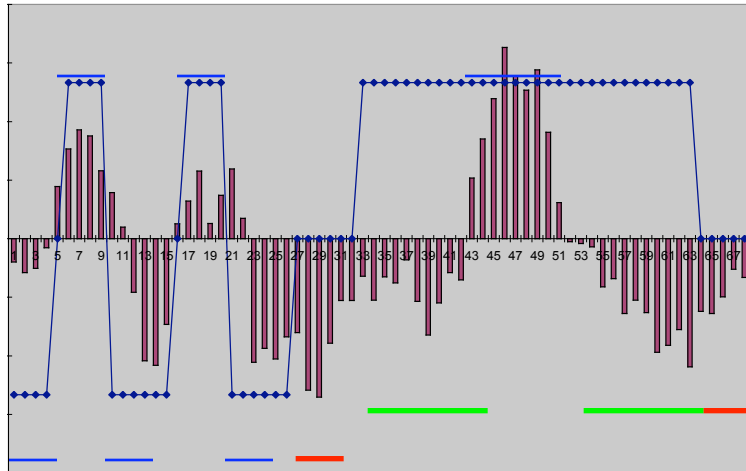


Figure 35 Genomic region of *Y. pestis* used for validation of Functional annotation.

The data shown corresponds to a region of the *Y. pestis* genome evaluated on the functional annotation chip. The maroon bars show the strand direction of detected transcription. The dotted blue-line represents the gene location of the published genome. Blue bars represent concordant observation where green bars represent opposite strand expression relative to the published annotation. The red bars highlight regions where strand specific expression was observed in a region annotated not to have any CDS.

Resolving power of functional annotation

Though now investigated in published reports I wanted to evaluate my ability to identify the transcriptional start and stop sites using functional annotation arrays. Additionally, it was important to establish the density of probe coverage to accurately identify transcriptional units. This was to be used to determine the numbers of probes needed in future studies. I used the ribosomal operon to address this question and the results are depicted in Figure 36 and Figure 37. These results demonstrated that, whereas a higher probe density might allow for more detailed mapping of transcriptional start and stop points, the variability of probe physical characteristics makes interpretation more challenging. In addition, higher density coverage requires increasing the number of probes. The practical approach became to identify quality probes at lower coverage, screen the transcriptome using these probes, and follow-up the analysis with more detailed mapping of transcriptional start and stop sites.

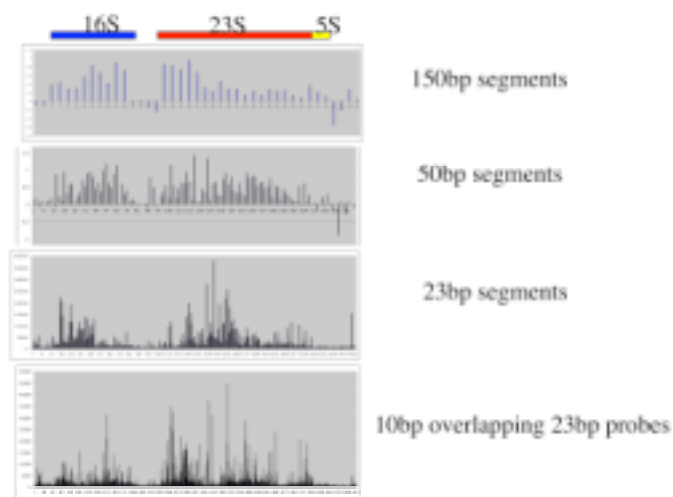


Figure 36 Resolving power of the arrays was tested using the ribosomal operon.

The resolution increased, however in the case of adjacent 23bp probes and 10bp overlapping probes the absence of an opposite strand reference made the background hybridization levels difficult to distinguish due to the heterogeneity of the probe hybridization efficiency.

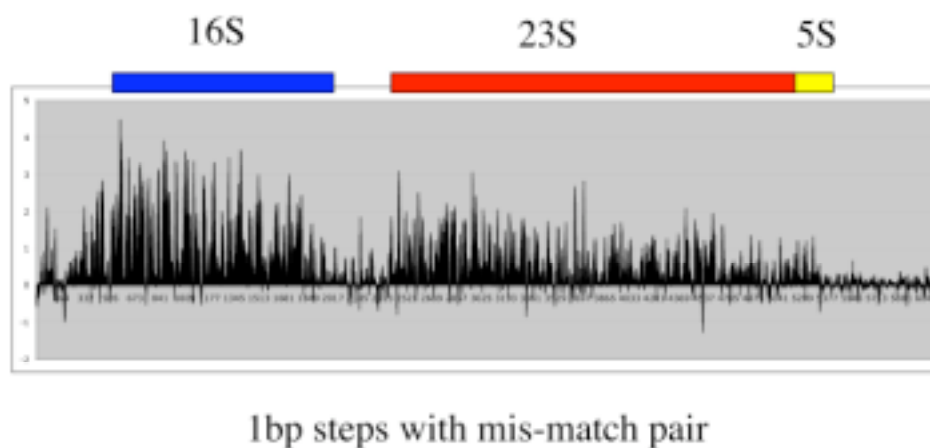


Figure 37 Single base-pair resolution of the ribosomal operon is enhanced with mis-matched probe normalization.

Functional evaluation of probe hybridization efficiency

In previous studies I relied on bioinformatics selection of optimal probes for hybridization (110). It became evident that even using algorithms, such as M-fold, the ability to accurately predict large numbers of probes that work efficiently was inadequate. I approached the probe design and hybridization efficiency problem through functional testing. Using genomic DNA hybridized to an array consisting of the top three probes bioinformatically selected, I showed that there was no bioinformatics bias in the most functional probe (Figure 38).

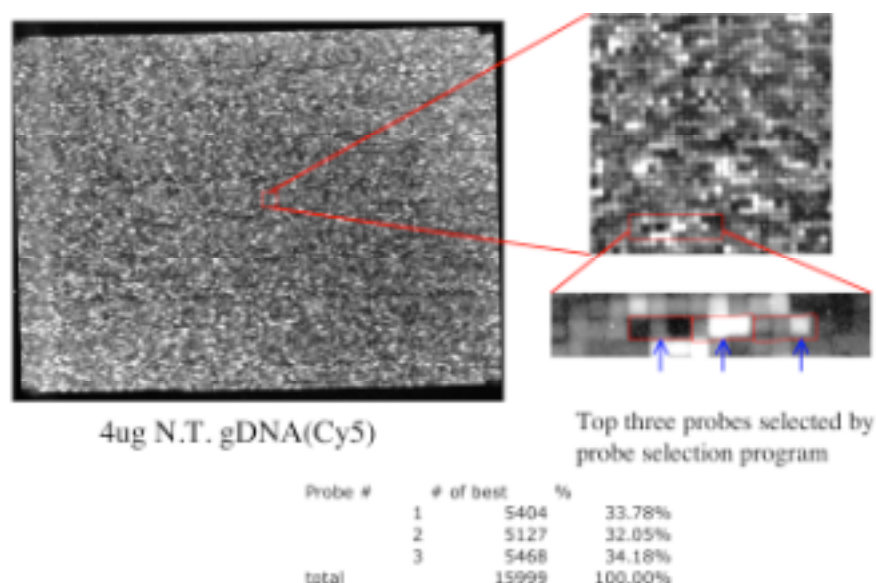


Figure 38 Functional evaluation of probes.

Three probes for each 150bp region of the *Y. pestis* genome were selected using M-fold and synthesized using D.O.C.. The hybridization of labeled genomic DNA was used to identify if computational methods accurately predicted the probe with highest hybridization signal. The results demonstrated that the current computation predictions were insufficient to identify functional probes and directly screening probe efficiency will lead to higher signal-to-noise ratios in future experiments.

Discussion

Functional annotation solves two significant problems. First, is the speed at which genome sequence data can be annotated, and secondly the accuracy of this annotation to include all transcribed genomic regions. Currently, genomic sequence is annotated through bioinformatics and educated guesses, which are both based on our current understanding of Biology. The fundamental idea behind functional annotation was that I did not assume our current understanding of transcription is complete. In essence the goal was to allow the biology to describe what parts of a genome are being transcribed. Functional annotation also was unique in that it took a non-computational approach to annotation and sought to obtain information through hybridization events.

I executed all aspects of this project, sans the synthesis of the D.O.C. microarrays.

APPENDIX III

Bioinformatic Studies

Identification of potential cross-protective antigens

Summary

An ideal application for bioinformatics would be to rapidly identify protective vaccines against multiple pathogens including biological warfare agents. The expansion of genome sequence data and bioinformatic tools have enabled this possibility to be explored. By comparing the genomes of multiple pathogens, the regions of identity of an immunologically recognizable size can be identified. Experimentally, we can evaluate their ability to confer protection against infection. *Yersinia pestis* and *Bacillus anthracis*, the causative agents of the plague and anthrax respectively, are potential biological warfare agents. Despite this commonality the biology and pathogenesis of these organisms are very different. Bioinformatic analysis revealed 2167 peptides, from 8 to 36 AA long, which are found in both pathogens. Evaluations of genomic distribution of these regions demonstrate diverse representation of the proteome. Evaluation of physical characteristics associated with immunogenicity did not reveal any obvious biases. However, alignments of genes containing identical peptide regions suggest two distinct groups of peptides; those from classical homologs

and those from chance or non-classical phylogenetic relationships. This may provide a valuable tool for the rapid generation of vaccines directed towards multiple biological warfare agents by targeting common epitopes.

Methods and Results

Genome sequences for *Yersinia pestis* and *Bacillus anthracis* were obtained from the Sanger Center and Institute for Genomic Research, respectively. Annotations were made using Glimmer software(155). Genome comparisons were done using two versions of the MUMmer program(48). The MUMmer program aligns genomes through the identification of identical regions called maximal unique matches (MUMs). These MUMs were used for our analysis. All other manipulations were done using Bbedit (text editor), Microsoft Excel, FileMaker Pro, and in house software. The basic strategy is diagramed below in Figure 39.

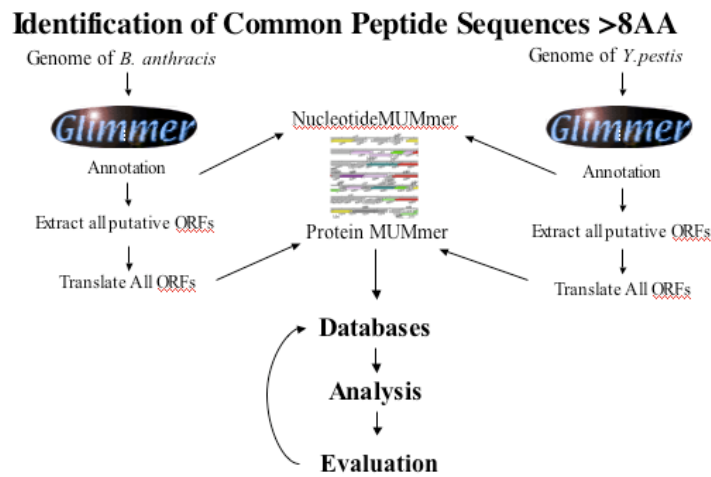


Figure 39 Bioinformatic strategy for identifying identical peptide sequences between multiple genomes.

After identification of the 2167 peptides I asked if these peptides originated from evolutionarily conserved proteins. To test this I used Clustal W to align the two proteins from which the common peptide was identified. Clustal W analysis suggests that the majority of gene pairs are homologous (defined by >20% identity). However, 25% do not, and appear to be delineated by an inflection point (red arrow) in Figure 40. Suggesting the MUMs found within these genes are derived from chance or non-classical phylogenetic relationships

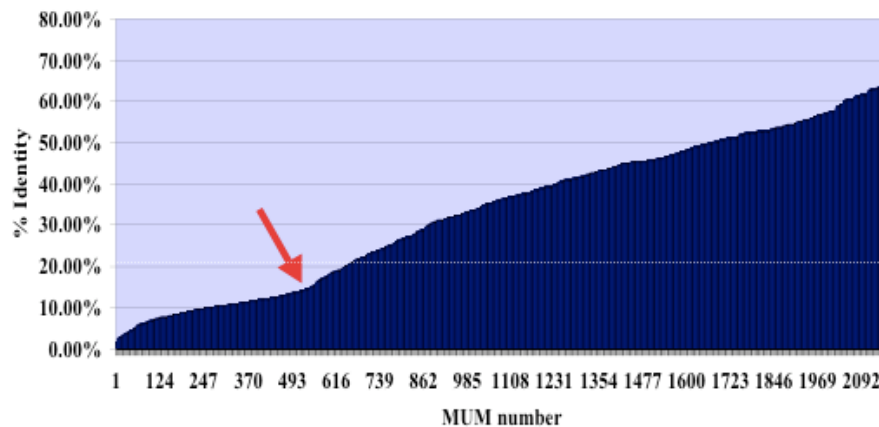


Figure 40 Clustal W alignments of gene pairs with common peptides.

Following alignment of the gene pairs from *Y. pestis* and *B. anthracis* containing identical peptides the percent amino acid identity was calculated, sorted and graphed.

Results Summary

Further analysis of the common peptides was performed to identify any functional or biochemical bias in the peptides themselves or the gene pairs in which they were identified. These analyses in did not reveal any discernable patterns.

Interestingly when three genomes were compared over 1300 common peptides were identified between the three genomes. Given that the functional test for this project was the ability of a common peptide to protect against both pathogens I proceeded to construct LEEs containing the common epitopes. At this time the project was turned over to another graduate student, who I assisted in supervising the completion of the project. The results of many animal experiments concluded that multiple antibody responses to these epitopes were difficult to obtain and thus the pool size had to be reduced. Although one pool of animals showed some level of protection against *Y. pestis* (sub-coetaneous challenged), this result was not reproducible and may have been due to variations in the infectious dose received by the animals. Ultimately, it was

determined that the animal models and pathogens used to evaluate this idea was sub-optimal. Since, the concept of a cross-protective antigen remains to be thoroughly evaluated I have prepared a proposal to better test this hypothesis.

Proposal for the identification of a single-antigen and cross-protective vaccine

Statement

The current trend in vaccine development is to use subunit vaccines because of their superior safety and fewer side-effects relative to whole-pathogen vaccines. However, subunit vaccines present a different hurdle to success: the identification of effective antigens. This has been typically addressed by propagating and then surveying the pathogen for highly abundant or immunogenic proteins. More recently bioinformatics approaches have been applied which attempt to predict protective antigens (70, 71). While these have met with some success, they have failed against many important human diseases. In several cases in which protective antigens have been identified, vaccine development has favored the formulation of combination vaccines as a means of making vaccination more convenient and therefore increasing patient compliance (the MMR combination vaccine as an example). This has introduced the hurdles of co-formulating proteins with biochemically diverse properties and raises concerns of possible epitope competition. This proposal is aimed at solving the

problems of both antigen identification and co-administration. This will be done by combining bioinformatics and high-throughput vaccine screening to identify a single antigen that is able to cross-protect against more than one pathogen.

Evidence that these cross-protective antigens exist comes from expression library immunization (ELI)(18). ELI takes advantage of genetic immunization to rapidly screen thousands of a pathogen's coding sequences to identify protective antigens. During an ELI screen to identify protective antigens toward *Legionella pneumophila*, a library of genetic vaccines from two other intracellular pathogens, *Salmonella typhimurium* and *Chlamydia psittaci*, were used as negative controls. However, the protection experiment revealed that animals inoculated with library of *S. typhimurium* sequences, but not those inoculated with the *Chlamydia* library, survived a challenge of *L. pneumophila*. These data suggested that some specific antigen from one pathogen was able to protect against a different pathogen.

Presently, there are almost 200 bacterial genomes sequenced with many more in progress. These databases describe the level of sequence conservation within homologous genes as well as the constraints biological constraints on coding sequences. For example, there are $\sim 10^{11}$ possible permutations of an 8 amino acid (aa) peptide (the size required for MHC class I presentation). A typical bacterial genome encodes $\sim 10^6$ possible 8-mers. Therefore, if sequence space was random the probability of any two 8aa peptides being the same between two bacteria is very low (about 10^{-10}). However, our informatics analyses of these databases have shown this to be untrue. We have instead found over 3,000 identical peptides ranging in length from 8-38aa between *Y.*

pestis and *B. anthracis* coding sequences. Many of these peptides occur between proteins in each pathogen with little to no sequence similarity, indicating that the identities are not the consequence of homologous evolution. We propose taking advantage of sequence information, bioinformatics tools, genetic immunization (173) and high-throughput vaccine screening to systematically search for cross-protective epitopes (Figure 41).

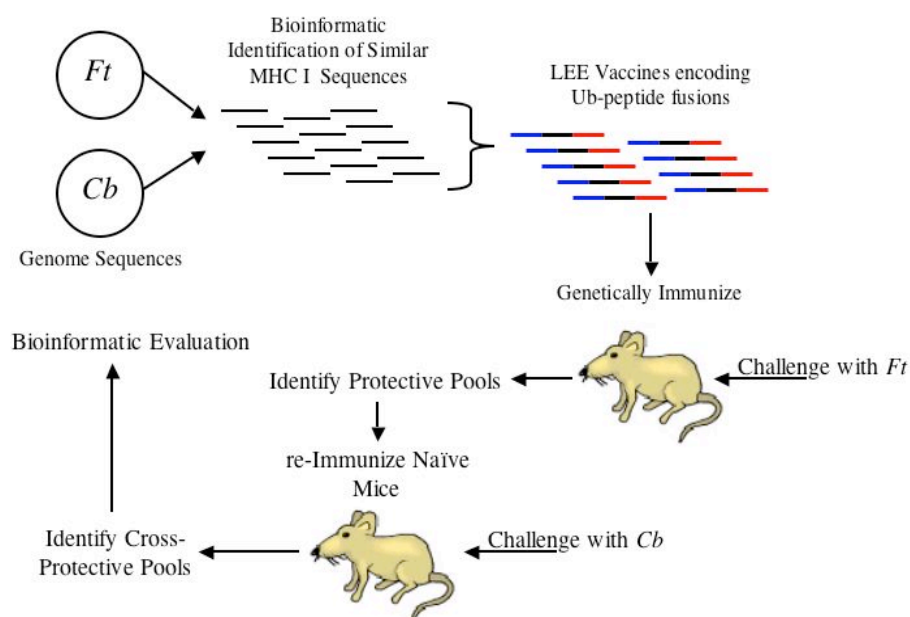


Figure 41 Proposed experimental approach to identify cross-protective epitopes.

Bioinformatics

The sequence of both of the pathogens has or is being elucidated. I have previously used bioinformatics tools to identify all identical predicted protein regions greater than 8aa. Based on previous results we would predict ~1000 non-redundant identical amino acid sequences. Given that both *F. tularensis* and *C. burnetii* are

intracellular pathogens, we will focus on the identification of MHC class I binding epitopes. All possible 8 to 11 aa peptides of each pathogen will be evaluated for predicted MHC class I binding using publicly available binding matrices and programs. Only those peptides with values above a threshold set for potential binding will be used for cross-genomic comparisons. We will identify all peptides present in both genomes that share 6 of the 8 amino acid identity. We can loosen or tighten this criteria to provide us with ~500 similar MHC class I binding peptides. We will initially use the *F. tularensis* sequence as our parent sequence for vaccine construction.

Vaccine construction

Taking advantage of the lab's expertise in genetic immunization, ELI, and linear expression elements (LEE) (167) for making screening library, we have developed high-throughput methods for synthetically producing ORFs and delivering them *in vivo* for expression without ever constructing plasmids. LEE protocols both significantly speed production of expression constructs and avoid the inherent biases and contamination problems of bacterial cloning. Oligonucleotides sufficient to assemble 500 peptide sequences will be synthesized. These oligos will be 5' flanked by a stretch of 30 bases that will correspond to the last 30 bases of the ubiquitin (UB) gene. The 3' end of each peptide-encoding oligo will correspond to the first 30 bases of the growth hormone (GH) terminator sequence. Promoter-UB and GH terminator linear elements will be separately generated in large batches by PCR(35, 167). Covalent attached of the PCR-built expression elements and the peptide-encoding ORFs will be accomplished by

a two-step overlapping PCR protocol. These constructs will be competent to express each peptide, as a ubiquitin fusion, in mice.

The Challenge – Protection Experiment

While we ultimately need vaccines that are fully protective, ELI screens are designed to obtain partial protection because diluted pools of antigens are delivered with adjuvant or optimized formats. Therefore we will design the challenge-protection assays to measure partial protection and to calculate quantitative protection scores. The readout of protection will be survival; therefore, extended time to death can be a significant measure of a potentially useful peptide. We will initially screen the 500 peptides in the *F. tularensis* model. We have previously demonstrated that large numbers of T-cell antigens can be co-delivered and successfully raise T-cell responses leading to protection. We will plan to pool up to 25 library LEEs into an inoculum (20 groups of 25 LEEs). Animals will be immunized by gene gun delivery at weeks 0, 4, and 10. Challenges with *F. tularensis* will be conducted at week 12, and mouse deaths will be monitored. Positively scoring inocula will be prepared again as LEEs for gene gun administration and used to inoculate fresh groups of mice. These mice will be challenged with *C. burnetii* and evaluated in a disease model of splenomegaly. If any one of these tularemia-protecting pools also confers significant protection against Q-fever disease then it will be considered cross-protective. If no protective pools are identified then we will return to the bioinformatics analysis and loosen our criteria to identify additional sequences, and possibly reduce the complexity of the pools to 10

LEEs per group in another round of testing. On the other hand, positive results would imply that one or more constituents of the pool(s) are cross-protective epitopes. We believe that these data would provide us the necessary preliminary results to win external funding. The initial screen could then be conducted on a larger scale, and cross-protective pools would be reduced to their valuable components. Identification of a set of empirically discovered cross-protective epitopes would enable us to search for correlated bioinformatic values. If successful, predictive parameters might be used in the future for the identification of more cross-protective epitopes.

I executed the bioinformatics aspects of this project. I also provided training and critical input in execution of the epitope delivery system constructed by Hoang Huygn. The proposal was originally written by myself and subsequently edited and reviewed by Dr. Kathryn Sykes.

PEPSEL - Rapid screening of sequence space for the identification of vaccine targets

Introduction

The basic premise of this project is to rapidly screen the sequence space (nucleotide, protein, etc.) of a pathogen for regions in that space that are most unlike the host organism. The identification of these regions can be used to reduce the amount of a pathogen's genome for vaccine target identification. This idea is based on a paradigm in vaccinology that states an antigen must be foreign to be immunogenic. This method will rapidly identify regions of potential high immunogenicity.

The basic protocol for this process involves three basic steps. First is to analyze the sequence space of a host organism (Human, Mouse, etc). This analysis will involve the reduction of sequence space into small segments of all potential combinations and permutations of the analyzed space (nucleotides, amino acids). This database will include the relative frequency of occurrence for each possible segment. Second, the sequences of a pathogen will be analyzed using the values assigned in the host database. Thirdly, regions of the pathogen's sequences (of length n) that have the lowest sum of scores will be targeted for vaccine design.

Intellectual property disclosure

The strategy and application of this project was disclosed to the intellectual property committee of the University of Texas Southwestern Medical School. The summary description and initial results that were provided are presented below.

The essence of this invention is to use genome sequence information to, on a genome wide scale, identify areas of a pathogens' protein(s) that are most unlike a host's proteins. It is thought that if a protein is foreign, or unlike a host protein, it will be more immunogenic. This idea is reinforced by the negative selection of cross reactive T cell receptors in the vertebrate thymus. This approach is novel because it permits the rapid quantification of a pathogens' proteome for protein, protein segments, or regions that are most unlike a host proteome. In order to achieve this goal two resources must be available. First is a collection of protein sequences of a host organism such as mouse or human. These are publicly available through the NCBI in the form of GENBANK or RefSEQ databases. Second, the annotated genome of a pathogen must be available. Again, many of these are available publicly.

The method for this analysis is described below and summarized in Figure 42.

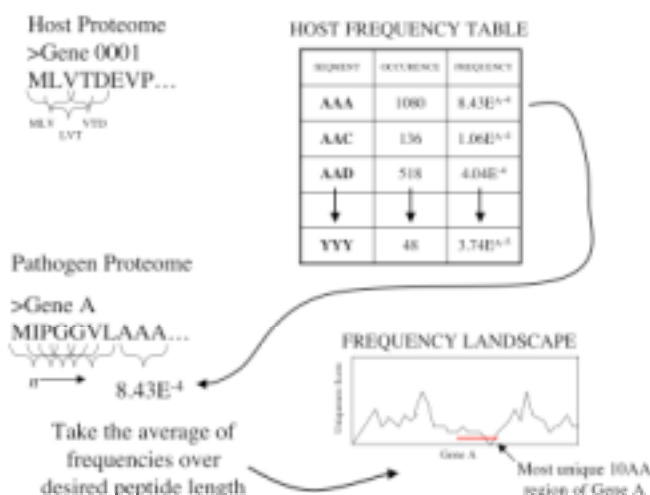


Figure 42 Summary of PEPSEL process.

The host organisms' genome is represented as a collection of protein sequences. An amino acid "window" size is selected (for example 3). A table is generated that represents all possible combinations of amino acids within that window size. For 3AA $= 20^3 = 8000$. Next all the protein sequences of the host organism are scanned using this window.

Each time an amino acid combination is encountered it is tabulated as one occurrence in the table. For example the sequence MIDIPIDI would look like this

MID 1

IDI 2

DIP 1

IPI 1

PID 1

This process is completed for all protein sequences of a host genome

Next the table is converted from a table of occurrences to a table of frequencies by generating a ratio for each AA combination of the number of specific occurrences over the total number of occurrences within a genome. For example in our previous example if the genome consisted only of this 8 AA protein the frequency of the tripeptide "IDI" would be 0.333

The next step is to utilize these frequencies while looking at pathogens' protein sequence. To do this the sequence of a pathogen's protein(s) is analyzed by breaking

the sequence into all the tripeptides found in that protein and assigning them the frequency value derived from the host proteome.

Next a landscape is generated by summing the host frequencies across the pathogen protein using a window of a length of AA's (~10AA). From this one can identify areas of the pathogen's proteins that are most unlike the host by identifying regions that have low frequency sums.

Results and Application

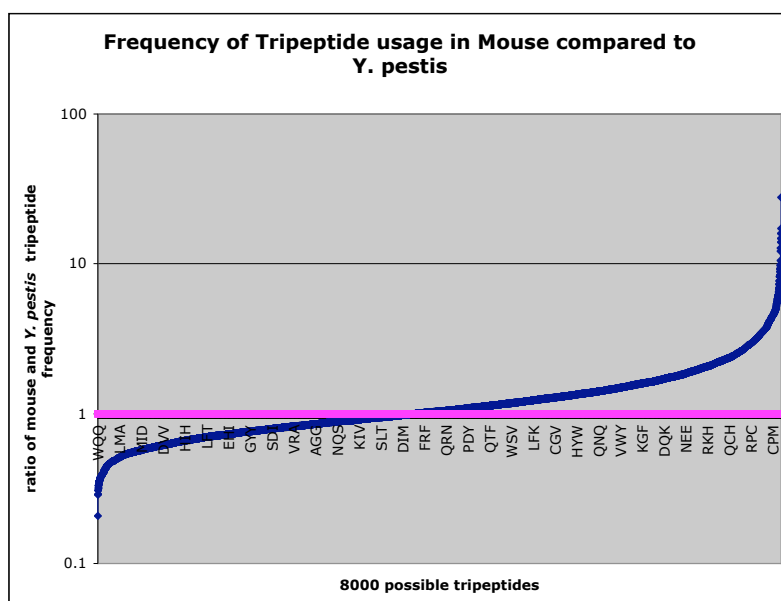


Figure 43 Comparison of tripeptide frequency in mouse compares to the gram-negative pathogen *Y. pestis* .

The blue line represents the ratio of the number of individual tripeptide frequencies identified in mouse relative to *Y. pestis* . The ratio of all 8000 tripeptides was sorted and graphed. The purple line represents the theoretical line of equal tripeptide frequency. Those tripeptides on to the left of the intersection of the two lines represent tripeptides used more frequently in mouse proteins and conversely those to the right of the intersection are used more frequently in *Y. pestis* proteins.

An application was developed to integrate this measurement in selection of antigenic regions of proteins for antibody production. This was used not only for identification of target epitopes in foreign sequences, but also in identification of epitopes within self proteins (mouse proteins immunogenic in mice) to generate antibodies using a high-throughput antibody production process (35). The University of Texas at Southwestern Medical Center did not pursue the intellectual property development of this process.

I initiated the ideas, approaches, and various tests involved in this project, and Dr. Ross Chambers provided critical comments during these steps. The program was written by Dr. Chun-Hui Bu.

PathoGene: A Pathogen Coding Sequence Discovery and Analysis Resource

Summary

PathoGene is a web-based resource that streamlines the process of predicting genes in microorganisms and designs polymerase chain reaction (PCR) primers for amplification to facilitate sequence analysis and experimentation. PathoGene currently supports primer design for every complete microbial, viral, and fungal genome as cataloged in GenBank by the National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>). The resulting primers can then be subjected to a stand-alone BLAST system called PathoBLAST in which the predicted PCR product and/or primers can be compared against the genome of interest or a similar genome to find related genes or estimate primer quality.

Introduction

Since the 9/11 terrorist attacks and subsequent Anthrax attacks, there has been a growing determination to prepare for future bioterror strikes or other public health crises such as Severe Acute Respiratory Syndrome (SARS) or West Nile virus. On September 4, 2003, Health and Human Services Secretary Tommy G. Thompson announced grants to establish ten Regional Centers of Excellence for Biodefense and Emerging Infectious Diseases Research (RCE).

The RCE program provides a coordinated and comprehensive mechanism to support the interdisciplinary research that will lead to new and improved therapies,

vaccines, diagnostics, and other tools. We have created a resource, PathoGene, which provides one capability important to researchers involved in this work and general microbiology. PathoGene is a web-accessible, up-to-date, quick, and validated tool for the identification of coding sequences (CDSs) and the design of computed reagents (primers sequences, amplicons, etc.).

To study a microbial CDS, its sequence must first be found within the genome. Its flanking sequences, on which PCR primers are to be designed, must also be found. Although PCR primer selecting software tools are available, generation of the input files is often cumbersome if performed manually. This is made more difficult due to strict formatting requirements, potential user length constraints, and number of CDSs to be studied. PathoGene was designed to assist microbiologists accomplish the above tasks in a fraction of the time. It is a web-based computer program (CGI) that integrates public databases and bioinformatics tools to completely automate PCR primer design process. At the time of writing, PathoGene supports primer design for 518782 CDSs in 160 microbial organisms from 238 genomic annotations, 1063 viruses from 1518 annotations, and 30 fungi from 69 annotations. PathoGene along with documentation is available as a free service of the Region VI RCE Computational Biology Group at <http://pathogene.swmed.edu>. Its efficacy has been validated in the lab via PCR amplification of a test set of CDSs from gram-positive *Bacillus anthracis* and gram-negative *Yersinia pestis*.

Materials and Methods

Computational Resources

The PathoGene source code was written in the Perl scripting language (v.5.8.0) and uses a standard CGI module. All WWW interfaces were created using standard HTML (v.4.01) and embedded Javascript. PathoGene and its pathogen database currently reside on a Linux (kernel v.2.4.19) web server. The database is continuously updated by a suite of local tools as new sequences and annotations become available in GenBank (22).

Validation Resources

PathoGene was tested and validated using 24 randomly chosen CDSs from *Bacillus anthracis* (Ames strain) and *Yersinia pestis*. *Yersinia pestis* gDNA was isolated from a cultured clinical isolate. The gDNA of *Bacillus anthracis* and *Yersinia pestis* was provided by C. Rick Lyons (University of New Mexico). The PCR primers were designed using the optimal conditions (157) of 800 bp product size, 23 bp primer size, 60°C primer melting temperature, and 50% GC content. The primers were obtained commercially from Qiagen Inc. (Valencia, CA, USA). Primer resuspension and redistribution were performed using a TECAN (Maennedorf, Switzerland) Genesis Robotic Sample Processor 200 to a final concentration of 10uM. PCR was carried out in the MJ Research (Waltham, MA, USA) PTC-225 Peltier Thermal Cycler in 20ul reactions (1X Taq buffer with MgCl₂, 0.2M dNTPs, 2U Taq polymerase, 0.4uM each

primer, 10-20 ng gDNA) using a standard cycling protocol (95°C for 5 min.; 30 cycles of 94°C for 45 sec., 55°C for 45 sec., and 72°C for 90 sec., followed by 5 min. at 72°C).

Sequence Analysis and Algorithm

PathoGene takes an organism selected by the user from a drop-down list, a user-supplied FASTA formatted sequence, or an uploaded FASTA formatted file as input. Users may also specify a particular CDS to design primers and choose the method in which CDSs are to be found. The organism input portion of the PathoGene user interface is depicted in Figure 44. Once an organism is chosen, PathoGene can find CDSs either by parsing the complete genomic annotation (Annotation Method) or by executing Glimmer (v.2.13) (47, 155) to predict CDSs given options defined by the user and the genomic sequence prepared by our script (Glimmer Method). If a FASTA formatted sequence or file is submitted, PathoGene will automatically execute the Glimmer Method to predict CDS locations. However, it should also be noted that if a small sequence (~ 2-3 kbp) is provided, a sufficient training model may not be generated to allow Glimmer to find the remaining genes properly. If a fungal species is chosen, PathoGene will automatically default to the Annotation Method because Glimmer is not compatible with eukaryotic genomes.

PathoGene™

A CDS Finding and Primer Design Tool for Microorganisms

Primer DB	PathoBLAST	About	Disclaimer	Home	Contact	Help
-----------	------------	-------	------------	------	---------	------

Input

Choose organism:

☒ Bacteria:

☐ Virus:

☐ Fungi:

☐ NIH Categorized Agents:

☐ CDC Categorized Agents:

Choose locating method:

☒ GenBank Annotation

☐ TIGR Glimmer v2.13 [glimmer options:](#)

Find CDS by:

☒ CDS Number CDS:

☐ CDS Name (Annotation only)

☐ CDS Locus Tag (Annotation only)

OR

Enter a FASTA format sequence:

> [optional comments] [carriage return or enter]
[DNA sequence with A, C, G, T, M, N, and/or W]

or upload a FASTA format file (PLAIN-TEXT only):

Figure 44 Organism input portion of PathoGene user interface

Once the location of the target CDS(s) is known, PathoGene will extract its sequence and the user-defined flanking and buffer sequences from the genome. Primer design will be avoided within the CDS's adjacent user-defined buffer regions to facilitate sequencing. Primer3 (v0.9) (152) is then used to design PCR primers within the flanking sequences of the CDS. Primer picking parameters such as product and primer sizes, GC content, and melting temperatures can be adjusted by the user. In addition, Real-Time PCR (RT-PCR) primers can also be made, but the product size is restricted between 80 – 200 bp. Generation of internal primers (for in-frame cloning) is also possible but not

recommended as the optimal primer picking parameters are severely relaxed. Results are then reported to the user. The above process is generalized in Figure 45.

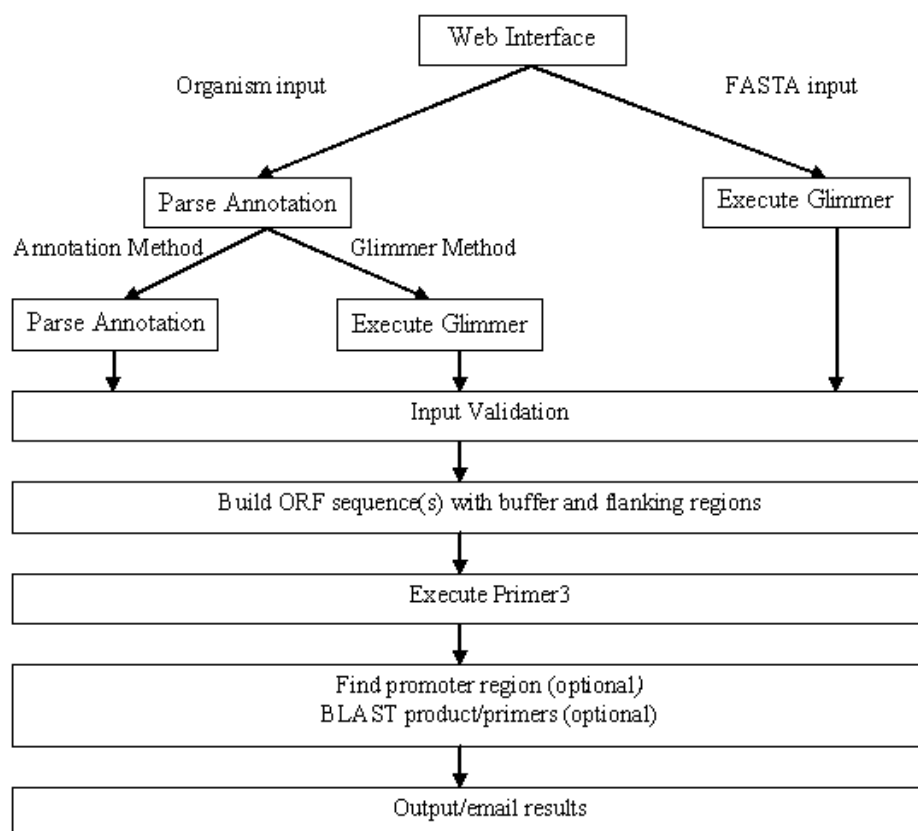


Figure 45 Flowchart describing PathoGene processing

In the event that a CDS is large, it can be broken into 500 bp segments. Primers are designed for each segment, which can then be amplified individually. The products can be recombined into a single CDS by molecular methods (e.g. overlap PCR) or computationally via sequence analysis. By default, the separated products will have a maximum overlap of 250 bp but this can vary depending on the chosen flanking and buffer sequence sizes. PathoGene also provides the option to BLAST (5, 6, 65, 113, 191)

the PCR product and/or primers against the target genome or another user-defined microorganism using the PathoBLAST resource (<http://rce.swmed.edu/genomes/>). This can determine the uniqueness of the predicted PCR product and the theoretical specificity of the primers. PathoBLAST is an independently developed, stand-alone BLAST tool that shares the same microorganism library as PathoGene.

The “promoter region,” defined as the closest, upstream, same-strand, non-coding region between the target CDS and the next upstream CDS with a maximum size of 2 kbp, can be found and have primers designed for it. PathoGene can also find the 2 kbp upstream region of each CDS. This information can be valuable to researchers and can easily be introduced into promoter motif algorithms and tools (14).

Primer DB is a complementary database to the PathoGene resource. It was designed to serve as a repository of validated primers generated by PathoGene. With this tool, researchers can submit and retrieve primers and the parameters under which they were made.

Output Format

PathoGene generates a separate results page once primary processing (prior to Primer3 execution) is complete. This page will display general information for the selected organism such its accession number, the total number of CDSs found, and the size of the genome. More specific information about the target CDS will then be provided such as its name, locus tag, PathoGene number (position which the CDS appears in the

annotation or Glimmer prediction), and its start and stop bases. If the CDS is found on the negative strand, this will also be noted. PathoGene will then output a series of sequences:

- a) The sequence of the CDS.
- b) The “promoter region” and/or 2 kbp upstream sequences if requested.
- c) The sequence of the CDS along with flanking and buffer regions attached.

Upon completion of Primer3, a hyperlink to the generated primer sets will be created. Hyperlinks are also created to PathoBLAST if the user chooses to BLAST the product and/or primers.

Validation Method

To confirm PathoGene as a rapid automated tool for primer design, we performed a comparison between PathoGene and manual CDS processing. We processed 24 randomly chosen CDSs from *Bacillus anthracis* and *Yersinia pestis*, and those larger than 500 bp were split into 500 bp segments. This yielded 30 segments for each organism and primers were individually generated for each segment. Using PathoGene, the Annotation and Glimmer Methods completely processed a CDS in an average of 6 seconds and 45 seconds, respectively. Manual processing was done using Microsoft Development Environment (v.6.0), Glimmer, Primer3, the genomic annotation, the accompanying FASTA formatted sequence, and find, cut, and paste operations. Per CDS, it took an average of 19 minutes and 27 minutes to emulate the Annotation and Glimmer Methods, respectively. Manual processing utilized the same primer picking parameters for Primer3 as PathoGene. Although PathoGene and manual processing yielded identical results, the

significant decrease in processing time and the standardization of the approach, not prone to user errors, demonstrates one value of the PathoGene resource.

To validate the generated primers, PCR was performed using the three best-fit primer pairs designed by Primer3 for the 30 segments in each organism. *Bacillus anthracis* primers had a 98.9% success rate. All primers gave a single product of appropriate size except for one pair. *Yersinia pestis* primers had a 96.67% success rate. Three primer pairs failed to give any product, however they were all designed to amplify the viral transposase gene *tnp*. There are numerous viral transposases and their presence and location are known to vary between strains of *Yersinia pestis*. The strain used in the PCR is a clinical isolate, thus its genomic sequence may vary from the annotated CO92 strain used to design the primers. Because all three primer pairs failed to generate a product, it is highly probable that the gene *tnp* is either not present in the clinical isolate or in a different genomic text within the isolate. All other primers gave a single product of the appropriate size. Therefore the success rate may be higher than what is reported. The fact that the PathoGene primers, designed using a sequenced pathogen, were able to successfully amplify homologous genes from a clinical isolate lends confidence to the utility of the program for practical applications.

Results and Discussion

PathoGene can facilitate microorganism investigations by automating most of the steps prior to PCR amplification. It has been shown to give identical results to manual processing but is on average 100 times faster. Furthermore, PathoGene reduces the

potential of human error. We have also shown that the program is capable of rapidly selecting PCR primers that generate single products from either the target strain or similar non-sequenced strain. This can facilitate the development of PCR-based diagnostics or genotyping tools. Furthermore, the program can be used to design primer sets for the generation of PCR-based amplicons for microarrays regardless of its annotation status. This will accelerate the study of pathogens for which detailed annotations have not been completed. While other studies and computer programs (140, 188) have been published relating to these computational tasks for microarray design, none offer the flexibility of PathoGene in terms of annotation status or the ability to BLAST the results for quality and specificity determination. Also, many do not have the capability of genome-wide primer design nor are the primer design parameters optimized for pathogen genomes. The PathoGene output can be applied to almost any PCR-based study of microorganisms.

Future upgrades include the addition of a motif finding function for the “promoter region” sequences. Also, a pre-computed primer database for every CDS in every organism under default parameters is in the early stages of development. This database, PathoGeneDB, will offer users instantaneous results for primer queries. Although the running time of PathoGene is negligible, the database will help reduce computational costs when a large volume of simultaneous queries are received or when computationally expensive requests are made such as primer generation for every CDS in an organism. Additional gene identification programs will also be added to provide the user with a variety of options.

We believe PathoGene is a rapid and versatile tool that can facilitate primer design and accelerate the study of microbial genomes. The immense size of PathoGene's organism library and diverse functionality provide researchers with a wealth of capabilities. The ability to examine individual and proprietary sequences can aid annotation and genomic projects. Through the synergy of these benefits, we believe this resource will facilitate work leading to new advances in biodefense.

I contributed critical input, and design in the planning of this work. The computer software was written by Kar-wai Ng. I performed the PCR validation of the output, and was involved in writing the manuscript.

APPENDIX IV

Biosignatures of infection

Serum Protein Signature of Pre-Symptomatic Influenza Infection

In my opinion the primary problem to be addressed in the use of *Y. pestis* as a biothreat is currently not treatment or prophylaxis, but the ability to detect and diagnose a potential release of this pathogen. Currently, the clinical goal standards for detection include culture or direct antigen detection in blood. The time required for the former precludes specific therapy and more importantly containment, while the later requires the infection to have progressed to a significant stage in its pathogenesis. Both of these methods do not provide information about an individuals' response or prognosis to the infection. Furthermore, both of these tests would provide information only after an individual was symptomatic. The significant need in this field is the presymptomatic detection of *Y. pestis* exposure.

The focus of this study is the detection of presymptomatic serum proteome changes in response to infection. Initially this work is not focused on the biology responsible for these changes in the sera, however, these data may be developed in this manner. I have selected the use of a commercially available antibody microarray to

assess the proteomic changes in the sera of influenza-infected animals. These data are being used to address fundamental and technical issues regarding our question. I have already established reproducible patterns of presymptomatic serum proteomic changes in response to viral infection. Work has continued to validate these data using auxiliary technologies and better characterize their nature.

Dynamic and quantitative measures of informative biological changes. Where information is defined through the interrogation of complex and temporal data normalized to an individual.

BioMarkers	BioSignatures
1 data point	> 1 data point
Based on Population Average	Based on an Individuals Average
Often Post-Symtomatic	Potentially Pre-Symtomatic
Diagnostic?	Diagnostic and Prognostic
Static measurement	Dynamic measurements

Table 8 Differences between BioMarkers and BioSignatures.

It is important to highlight the differences between biosignatures from traditional biomarkers. These differences are summarized in Table 8. A classical example of a biomarker is the prostate cancer marker PSA (prostate specific antigen). Where the presence of this protein in the serum is not diagnostic for prostate cancer, changes in the level of this protein may be informative in the progression of a cancerous lesion. Furthermore, the distribution of population averages for PSA levels is broad and makes a

standard for interpretation of these data difficult. The aim of a biosignature, is to integrate the information content of many proteins to develop a higher degree of diagnostic and prognostic information specific for an individual.

Methods

I took advantage of a commercially available antibody microarray made by Clontech (BD biosciences), to test an influenza infection model for presymptomatic proteomic changes in the serum. Balb/c mice were infected with 10^5 pfu of influenza strain PR8 intranasally. Serum samples were collected from animals at 24, 48, and 72 hours post infection. Serum was obtained by cardiac-puncture followed by coagulation and centrifugation of the collected blood. These samples were aliquoted and frozen at -80°C . The experimental outline is described below in Figure 46.

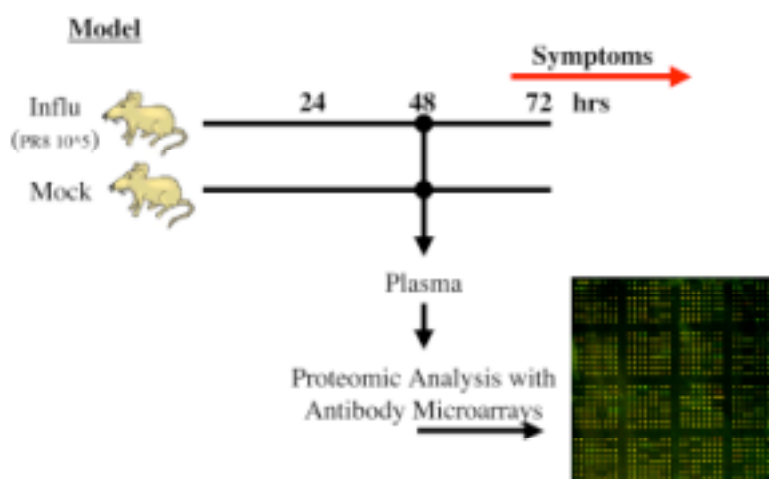


Figure 46 Experimental approach to presymptomatic serum proteomic changes following influenza A infection.

My goal in this experiment was to select a presymptomatic time-point that would provide the greatest opportunity to identify preclinical changes in the plasma proteome using antibody microarrays.

To reduce the loss of information total sera proteins were used in these experiments (albumin was not depleted). The purification and labeling of these proteins was done according to the manufacturer's suggested protocols. In all experiments the large-scale preparations (using PD-10 columns) were used as I determined that this was able to remove free dye more effectively than the small-scale protocols that utilized spin columns for protein purification and labeling clean-up.

Hybridization and experimental set-up followed manufacturer's suggested procedure. The slides were then scanned using a scanarray express (Perkin Elmer), and the signals were balanced by adjusting laser power and PMT levels.

Results

I focused on the 48-hour time-point as this was still pre-symptomatic, yet later in infection to maximize my chances of identifying serum proteomic changes. Three replicate experiments were completed using serum from animals 48 hours post-infection. These data were analyzed using GeneSpring software (Silicon Genetics), and each sample represented two slides with dye-swapped labeling. Data that had a coefficient of variance less than 30% for the three experiments was analyzed in a scatter plot.

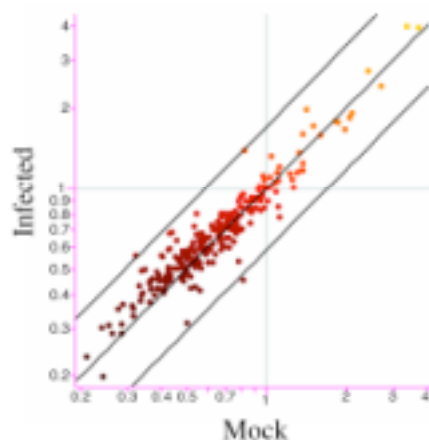


Figure 47 Scatter plot of serum protein changes 48hrs post-infection with influenza A.

Three independent experiments were conducted and filtered to retain only highly reproducible data. The 2-fold change lines are graphed.

Using the binding pattern of these antibodies towards total plasma proteins 48hr post-infection, we were able to identify 18 proteins with significant changes between

the naïve and pre-symptomatic infected animals Table 9. 14 of these antibodies demonstrated an increase in substrate binding following infection. Interestingly, four antibodies showed a decrease in substrate binding. Possible mechanisms for infection specific down-regulation of serum proteins are; one, the down-regulation of secretion/production of the serum protein, two, the increase in specific degradation of the protein in the serum, or thirdly through the uptake of the serum protein by host cells. Although these experiments are unable to distinguish any of these possibilities, the observation raises an interesting line of investigation regarding mechanisms of protein depletion within the serum following infection.



Table 9 Proteins found to have the greatest change in the serum following influenza infection

Many of the proteins identified are not thought to be plasma proteins and yet are detected and show significant changes following infection. Those increased following infection (red) outnumbered those identified as decreasing (blue) with the transcription factor HIF1 α (circled) showing the greatest increase in the plasma proteome.

Given that a majority of the antibodies on the microarray were specific for cellular proteins, I was interested in determining if the protein being captured was the

antigen the antibody was originally generated towards. In the one case I was able to test using a second monoclonal antibody specific for G-protein family member, I was able to show binding of the antibody to a spot on the antibody array specific for a G-protein. This work suggested that, in this case, it did appear that this cellular protein was present in the plasma proteome. This has yet to be confirmed using western blots or mass-spectrometry (MS) analysis as the abundance in sera is below the detection limits of western blotting and the diversity of plasma proteins (and large amounts of albumin) have make clear MS identification difficult.

Summary

Detection and discrimination of pre-symptomatic infectious diseases will be a valuable medical diagnostic tool. Towards this goal I demonstrate a distinct pattern, or signature, of host proteins identified in serum able to discriminate pre-symptomatic influenza infection in a mouse model. I employed a commercial antibody microarray with ~500 monoclonal antibodies towards a variety of cellular antigens. Using the binding pattern of these antibodies towards total plasma proteins 48hr post-infection, I was able to identify 18 proteins with significant changes between the naïve and pre-symptomatic infected animals Table 9. Additional efforts are bringing into focus important aspects of presymptomatic detection of infection. The feasibility of biomarker versus biosignature discovery for the detection of presymptomatic infection is being addressed. My efforts in this project are helping to understand the resources, both practical and computational, needed to define a quantitative biosignature of infection.

Immunosignatures - High-throughput Patterning of T-cell Repertoires

Statement

In this proposal we seek to develop a novel method for high-throughput assessment of T-cell repertoires. In the context of advancing understanding of immune cell responses and development, current procedures are limited in the lack of detailed information they provide. A process to acquire large volumes of data regarding the status of the immune cell repertoire would benefit both basic science and clinical diagnosis.

The immune system's genome is unique because it is continually in a dynamic state of genomic rearrangement. This dynamism is focused primarily in two genomic loci; the B-cell receptor (BCR) genes, and the T-cell receptor (TCR) genes. The unique rearrangement of these gene products in their respective compartments and cell populations provide the specificity associated with appropriate immune responses. Understanding the composition of immune cell populations is vital in the study of basic immunology, clinical immunology, vaccinology, and autoimmune diseases. Presently, this work is advancing primarily through the sequencing of genes from individual cells. There are an estimated 10^{12} lymphocytes in the body with 10^7 of these circulating in the blood at any given time. With the potential diversity of $\alpha\beta$ TCR being 10^{15} and for $\gamma\delta$ TCR being 10^{18} it becomes clear that assessment of the TCR diversity by PCR or sequencing is limiting. Understanding the composition of the circulating lymphocytes will provide unique insights into many fundamental scientific questions.

Research Impact

If successful this work will provide a novel means to both assess and study immune cell repertoires. An additional tool to assist in the study of autoimmune disorders, chronic diseases, and immune dysfunctions will greatly influence both basic and clinical science efforts working to understand these problems. Furthermore, this method could serve as a platform for clinical diagnostic tools. However, our desire will be to define our system to ultimately address basic scientific questions that have been unapproachable with current methods. Specifically 1) What is the diversity of TCRs generated in response to an antigen? 2) Do individuals arrive at the same, similar, or different immune solutions to the same antigen? 3) How does antigen re-exposure sculpt the memory response?

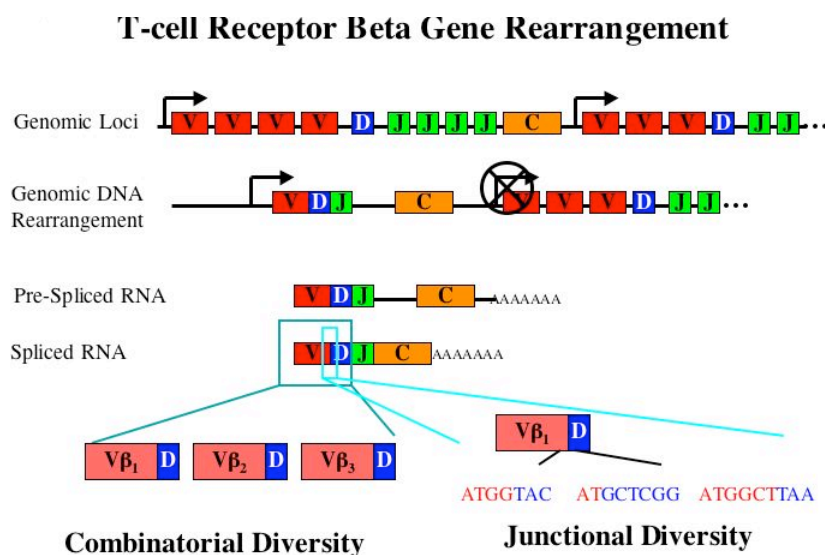


Figure 48 Generation of TCR diversity.

Research Plan

The premise of this work involves the assessment of combinatorial and junctional diversity of TCR beta chain proteins. We will approach this problem in three aims, the results of which will define the practicality of our approach. The generation of TCR beta chain diversity is described in Figure 48.

The majority of T cells contain TCRs with alpha and beta heterodimers. Although antigen specificity is driven by both subunits.

We will also take advantage of our experience with microarrays. Specifically we will employ the Digital-Optical Chemistry synthesis platform developed at UTSW and our recent published demonstration of a universal microarray. Alternatively, the availability of custom commercial arrays may supplement this work

Aim 1: Define the Ovalbumin specific TCR beta chain cDNA hybridization pattern to a Universal Microarray.

We will take advantage of a clonal T-cell line specific for the recognition of an epitope found within Ovalbumin. The cDNA specific for this gene will be amplified, isolated and sequenced. It will then serve as a template for analyte preparation prior to microarray hybridization using a unique linear amplification technique (LAPT) discussed earlier. The experimental outline is depicted in Figure 49.

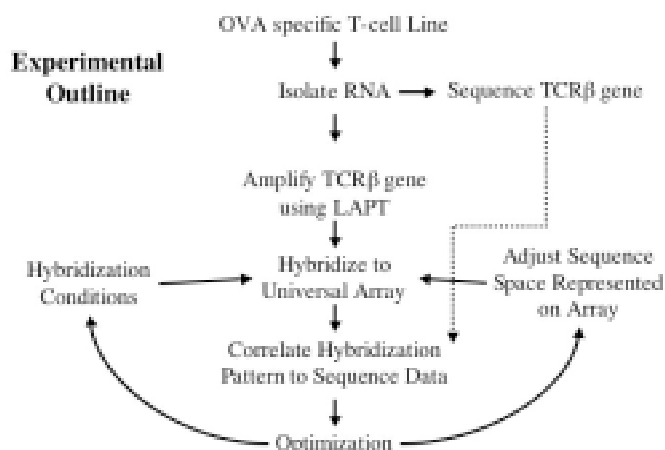


Figure 49 Outline of experimental approach to immunosignatures.

Knowledge of the specific analyte applied to the array and the flexibility of the microarray platform will enable us to define hybridization conditions. Additionally, using computational tools, the data will facilitate the expansion of informative sequence space represented on the array. The results of this aim will allow us to define the appropriate substrate and conditions for the detection of a single TCR beta gene.

Aim 2: Differential detection of two specific TCR beta genes.

This aim will parallel the approach described in aim 1 with the addition of another TCR beta gene during hybridization. With these experiments we will further characterize hybridization conditions in a more complex system as well as refine the borders of informative sequence space. The latter will be of particular importance as significant portions of the TCR beta gene is comprised of conserved regions. In our efforts to assay the combinatorial and junctional diversity this sequence space will be

less informative. Titration experiments with these two genes will assist in establishing detection limits and relative abundance in our system.

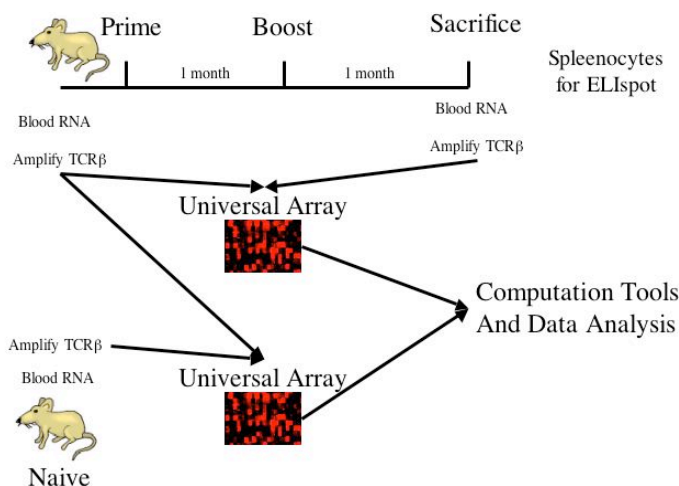


Figure 50 Proposed system for validation of TCR-based immunosignatures

Aim 3: Detection of specific TCR beta gene rearrangements in mice following immunization with ovalbumin.

Again we will take advantage of the current knowledge regarding the OVA immunogen. The experimental approach for this aim is outlined in Figure 50.

The proof of principle experiment for our proposed technology platform will be the specific detection of TCR beta gene rearrangements following immunization. We will evaluate this at two levels.

First, RNA isolated from blood of a naïve mouse will be collected and the TCR beta genes amplified using a conserved TCR beta constant region primer to drive linear amplification. The mouse will then be immunized with OVA in complete Freund's

adjuvant and boosted 4 weeks later. Three days after immunization the mouse will be sacrificed and blood and spleenocytes collected. The development of specific OVA T cell responses will be demonstrated using ELISpot. RNA isolated from the blood will be used to generate a second collection of TCR beta gene cDNAs. The pre-immune and post-immune TCR beta genes will be differentially labeled and co-hybridized to an array. These data will be mined for the appearance of unique TCR beta rearrangements induced by immunization with OVA. Additionally, we will study the data for the appearance of the specific TCR beta gene rearrangement defined in aim1.

Second, the pattern of TCR beta gene rearrangements from a different non-immunized mouse will be compared to the immunized animal. This will serve as an important control for the possible development of non-OVA specific TCR beta gene rearrangements in the blood. Furthermore, it will serve as a baseline measurement of the variation of TCR beta genes found within inbred mice.

Results of these studies will help define the feasibility and sensitivity of this approach to assay immune gene repertoires. Furthermore, these results will provide substantial data to direct future studies which could take advantage of this method and would be more amenable to external funding.

APPENDIX V

Toward A Minimal Small Molecule Microarray For Protein Fingerprinting: The Utility of Promiscuous Protein-Binding Agents

Abstract

We have demonstrated that hybridization of a protein to a microarray comprised of thousands of different peptoids provides a unique pattern that can be taken as a “molecular fingerprint” of that protein. The pattern is the result of a quantitative assessment of the binding of the protein to each feature on the array. For a number of reasons it would be desirable to be able to conduct such experiments on spatially defined arrays containing fewer features. Thus, we have mined the data sets obtained for hybridization of GST, MBP and Ubiquitin to 7680 feature arrays for clues as to which subset of peptoids would be of the greatest utility in protein fingerprinting. We show here that one route to a much simpler, but effective, protein fingerprinting array is to employ peptoids that are relatively non-specific ligands for a variety of proteins. We demonstrate that quantitative analysis of 65 such features is more than sufficient to uniquely identify the fingerprint of a fourth, previously untested, protein, AAT. Thus, while non-specific protein-binding molecules are of little interest in pharmacology and chemical genetics, the data reported here indicate that they will be useful tools for analytical work.

Introduction

There is increasing interest in the discovery of new biomarkers for disease diagnosis. However, it is becoming increasingly difficult to identify single proteins that serve this purpose(8). Thus, several investigators have turned towards protein profiling or pattern recognition approaches(136, 137). In this approach, one monitors the levels of many different proteins (or other biomolecules) simultaneously and attempts to correlate the pattern observed with phenotype. This has most commonly been done using mass spectrometry-based methods, particularly the ProteinChip® / SELDI technology(58, 132). While this approach has been controversial, there have been several reports of interesting correlations between proteomic patterns and disease state progression, particularly in the cancer arena.

We have recently initiated a program to determine if protein profiling methodologies can be adapted to other analytical formats, in particular that of a small molecule microarray (SMM), that might have advantages in cost and reproducibility. One could imagine carrying out protein profiling in an array format by hybridizing a complex protein mixture to a microarray. The huge number of protein-small molecule binding events would be expected to form a pattern that might be of diagnostic value. A broad view of protein binding could be obtained by fluorescently labeling the proteins prior to hybridization. However, chemical labeling of protein mixtures is problematic in many ways(98) and could be difficult to conduct reproducibly. It would also result in abundant proteins dominating the pattern observed and these are often not the most interesting factors from a functional standpoint. An alternative would be to employ

selective techniques to visualize the binding pattern of some fraction of the proteins in the sample. For example, one could employ activity based labeling to fluorescently tag proteases in the sample specifically(17).

Alternatively, a labeled antibody capable of recognizing many different proteins in the sample, for example one that recognizes a post-translational modification such as a phosphotyrosine, could be employed to visualize a binding pattern.

The first step in addressing the feasibility of protein profiling on SMMs would be to ask if a unique and reproducible binding pattern is produced when an array is exposed to a single protein. Towards this goal, we showed recently that when a fluorescently labeled protein is hybridized to an array of 7680 peptoids (163, 193) a unique protein “fingerprint” is produced (146). This was derived from a quantitative measurement of the fluorescence intensity at each feature on the array, as determined using a standard array scanner of the type employed to analyze DNA microarrays. We also showed that useful fingerprints could be derived for a native protein in a complex extract. This was done by visualizing the binding pattern of the protein of interest selectively through subsequent hybridization of a specific antibody and a labeled secondary antibody. Similar results were obtained by Carlson and coworkers, who employed arrays with features created from co-spotting mixtures of activated carboxylates onto amine-functionalized slides (R. Carlson, personal communication). Earlier, Mihara and co-workers demonstrated that liquid arrays of designed peptides carrying environmentally sensitive fluorescent tags were also capable of protein fingerprinting(168, 178) (also see (15)for a more specialized liquid array-based protein

fingerprinting approach). Thus, these initial studies have made clear that it is technically feasible to measure unique protein fingerprints in a microarray format and have set the stage for more complex profiling experiments of the type described above.

A long-term limitation of the peptoid arrays that we employed in these studies is that the molecular structure of the peptoid that constitutes a feature is unknown. This is because the peptoids spotted onto the arrays were derived from libraries made by split and pool solid phase synthesis. Individual beads in the library were segregated into the wells of a microtiter plate and then cleaved from the beads. The resultant DMSO solutions were spotted robotically onto a chemically modified glass slide. Thus, while hundreds of slides can be made from a single library synthesis, once a library is exhausted, it would be almost impossible to recreate the same array, a significant problem if the goal is to use these arrays to visualize patterns of diagnostic utility. Edman degradation (4) or mass spectrometry-based sequencing could be employed to determine the sequence of any peptoid and it would be possible, albeit tedious, to do so for the entire library. But even with this information, it would be highly impractical to independently resynthesize thousands of individual peptoids to create spatially defined microarrays. For this reason it was of interest to us to explore the feasibility of constructing simpler arrays comprised of fewer peptoids. In particular, we wished to address the question of whether one could fingerprint proteins with less than 100 features, a number that would enable the creation of spatially defined arrays by independent parallel synthesis of individual features.

We report here a facile strategy for simplifying the arrays. Through an analysis

of three large data sets obtained from different protein hybridizations, we found that 65 of the 7680 features on the array retained high levels of all three proteins, but to different extents. We therefore investigated the hypothesis that these 65 “promiscuous peptoids” would be capable of providing a distinctive protein fingerprint for almost any protein through a quantitative analysis of the relative level of binding of that protein to each feature. We show that this is indeed the case, providing a simple strategy for the construction of highly simplified protein fingerprinting arrays.

Results

Identification of Promiscuous Peptoids In A Large Library.

We have examined the large data sets produced by hybridizing three different proteins: Glutathione-S-Transferase (GST), Maltose-Binding Protein (MBP) and Ubiquitin (Ub) (in each case 500 nM labeled protein in a 100-fold excess of bacterial proteins obtained from an *E. coli* extract) to 7680 feature peptoid arrays (146). The general structure of the library, which was constructed using a microwave-assisted(125) “sub-monomer” synthesis(193) and the amines employed to make it are shown in Figure 51.



Figure 51 Peptoid library construction.

(A) shows the general structure of the library and (B) shows the sub-monomers used in the synthesis of the library.

The raw images of these hybridization experiments are shown in the following Figure 52.

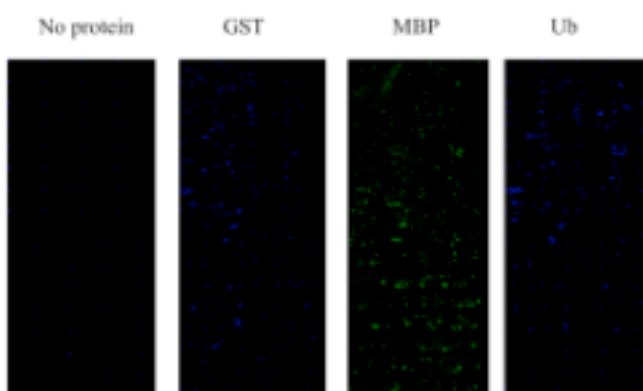


Figure 52 Raw images of peptoid library microarrays after hybridization with three labeled proteins.

Since the goal of this effort is to identify the peptoid features most useful for protein fingerprinting, we first determined the features in each experiment that reproducibly

provided a signal at least 10 standard deviations above the average signal from spots containing only DMSO (peptoid solvent), which we defined as background. 265 of the 7680 features (3.5%) satisfied this criterion in the GST hybridization experiment, 599 (7.8%) in the MBP hybridization experiment and 181 (2.4%) in the Ubiquitin experiment. Clearly, there are a significant number of features on the array that bind these three proteins with sufficient affinity that one could imagine building a simpler array with far less than 7680 features.

As shown in the Venn diagram in Figure 53, many of the highly fluorescent features represented peptoids that were quite specific for one of the three proteins, 75 (1.0%) for GST, 395 (5.1%) for MBP and 32 (0.4%) for Ubiquitin. A smaller number of features displayed high levels of fluorescence in two, but not all three of the hybridization experiments. Finally, 65 features (0.8% of the array) displayed a level of fluorescence above the threshold in all three experiments (Figure 53).

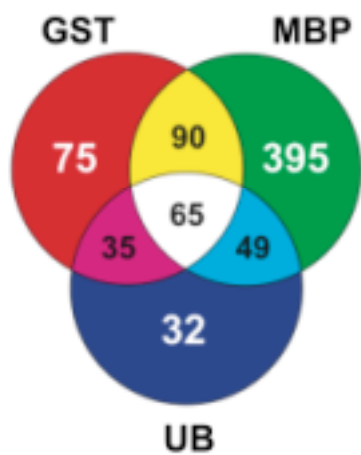


Figure 53 Venn diagram to identify promiscuous binding peptoids from three unrelated proteins.

The existence of what we will term the “promiscuous” features intrigued us. While these features bound GST, MBP and Ub strongly, there were nonetheless easily distinguishable differences between their intensities in the three data sets. We decided to explore the possibility that many or all of these 65 peptoids would bind to almost any protein hybridized to the array with the hope that the quantitative pattern of binding to this much simpler array would provide a trustworthy protein fingerprint.

The Majority of the Promiscuous Peptoids Bind The Previously Unstudied Protein AAT.

To investigate this point, we carried out another hybridization experiment with the protein AAT, which we had not studied previously and therefore had no prior

knowledge of how well it would bind to any of the peptoids on the array. The AAT hybridization experiment was carried out under conditions identical to those employed for the other proteins in order to facilitate comparisons between the data sets (500 nM fluorescently labeled protein in the presence of a 100-fold excess of *E. coli* proteins). To further facilitate comparison, we carried out the hybridization on the full 7680-feature array rather than re-spotting only the 65 promiscuous peptoids, and then focused on only the features of interest in the analysis.

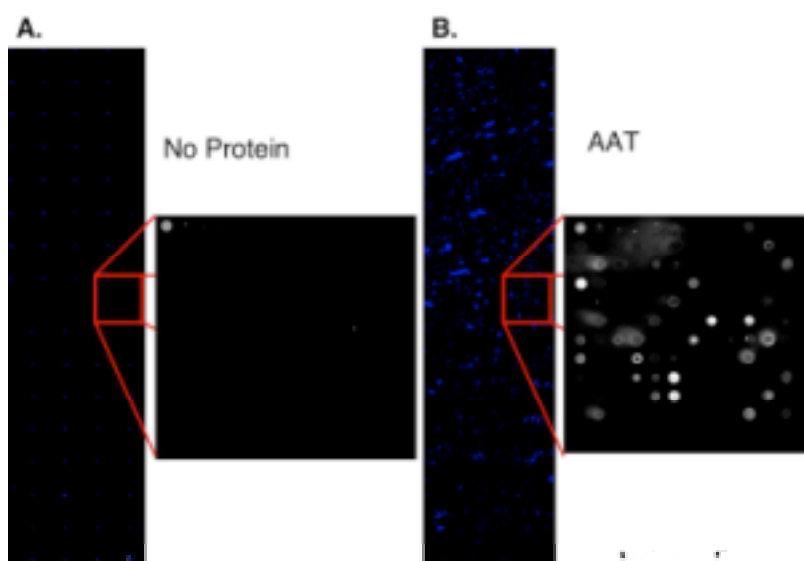


Figure 54 Hybridization of AAT to the peptoid array.

Hybridization pattern of AAT to the peptoid array compared to a no-protein control. The control spot is seen in the upper left hand corner of the inset images.

The full hybridization pattern of AAT on the array is shown in Figure 54. To assess the reproducibility of the AAT data, two completely independent experiments were performed and the two data sets compared on a scatter plot (Figure 56A). The

reproducibility was very good, comparable to that observed with the other proteins (spearman correlation coefficient of 0.74). As might be expected, the variance between the data sets is greater at lower intensities, consistent with a higher signal to noise ratio for proteins binding to lower affinity peptoids. As the signals increase, the variance between experiments decreases (see Figure 55).

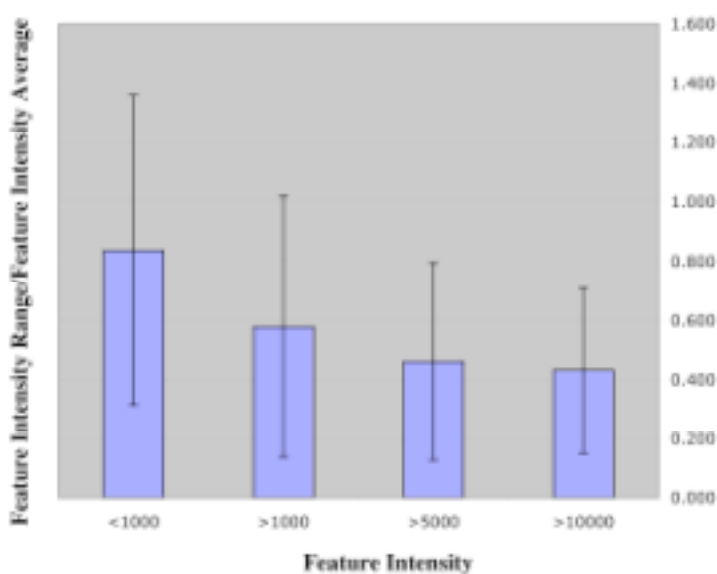


Figure 55 Inverse correlation of variance with peptoid signal intensity.

The highest intensity features are consistent between arrays with lower variance in absolute intensity. Figure 56B-D shows scatter plots that compare the AAT hybridization data sets with the other three. The large clusters of points well off the diagonal are indicative of the two highly divergent binding patterns of different proteins. In these plots, all of the features that exhibit fluorescence above the average

negative control, however small, are included. The shaded rectangle indicates the intensities that are less than or equal to two standard deviations above the background.

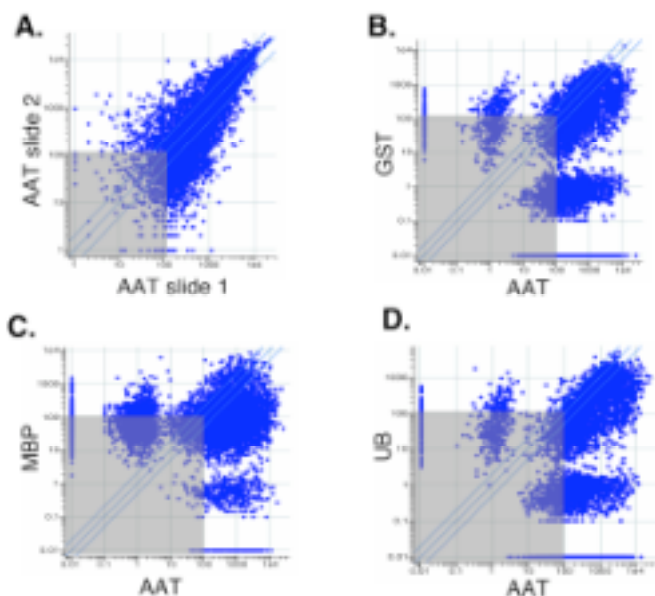


Figure 56 Scatter plots of peptoid array data for the protein AAT.

The scatter plots of peptoid array data is given for AAT compared to (A) a replicate AAT hybridization, along with comparisons to the three proteins GST, MBP, and UB, (B),(C),(D) respectively.

The Promiscuous Peptoids Support Protein Fingerprinting.

To address the question framed above, we then focused on the 65 promiscuous features and quantified the intensities at these positions in all of the arrays. A color code was assigned to represent the feature intensities and these were assigned a color barcode to facilitate visualization of the patterns. As shown in Figure 57, this color barcode provides a clear distinction between AAT and the other proteins. Of the 65 promiscuous features analyzed, the fluorescence intensity in the AAT hybridization was greater than

10 standard deviations above the average negative control for 45 of them. Of note, an additional 12 of the 65 promiscuous peptoids showed binding intensities well above the slide background for the AAT protein, but fell below the 10-fold cut-off. To the best of our knowledge, GST, MBP, Ubiquitin and AAT are not related in any way. Thus, these data strongly support the idea that a majority of these promiscuous peptoids will function as ligands for most proteins hybridized to the array.

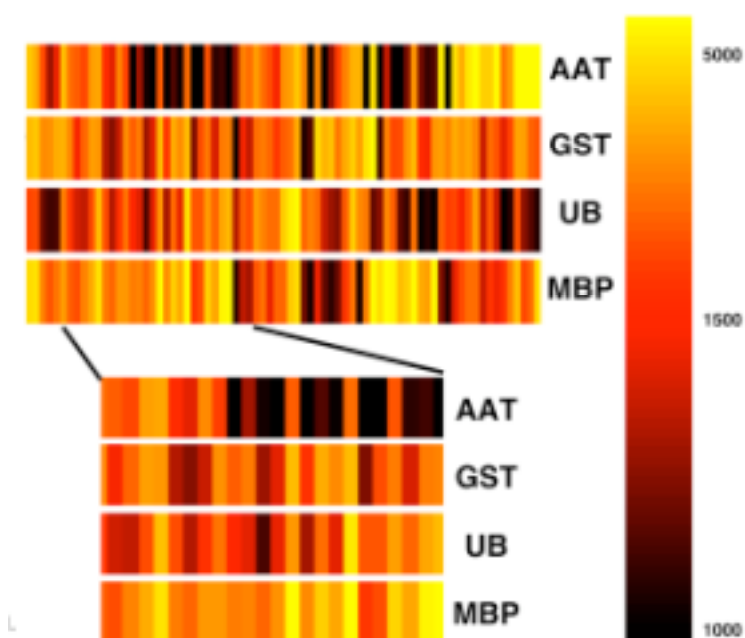


Figure 57 Graphical representation of peptoid array data.

The colored bar-code demonstrates unique binding patterns for all four proteins. A blow-up of one of the regions is provided for greater detail.

To further address if these 65 peptoids are able to provide reproducible “protein fingerprints” a blinded experiment was conducted. Three data sets, prepared from one or more of the 4 proteins employed in this study, were analyzed without knowledge of

the protein identity. Figure 58 shows the “barcode” of the 4 proteins in addition to the three unknown samples. The samples were clustered based on their Spearman rank order of the 65 peptoid binding intensities. Note that this approach corrects for differences in the intrinsic intensities observed in any single hybridization experiment, which could vary based on a variety of factors. The rank-order was used to place the emphasis on relative binding pattern of the 65 peptoids within an individual sample. The tree organization allowed us to correctly identify the unknown samples, shown by the close relatedness of the first two blinded samples to the AAT pattern, and the third blinded sample to the Ub pattern.

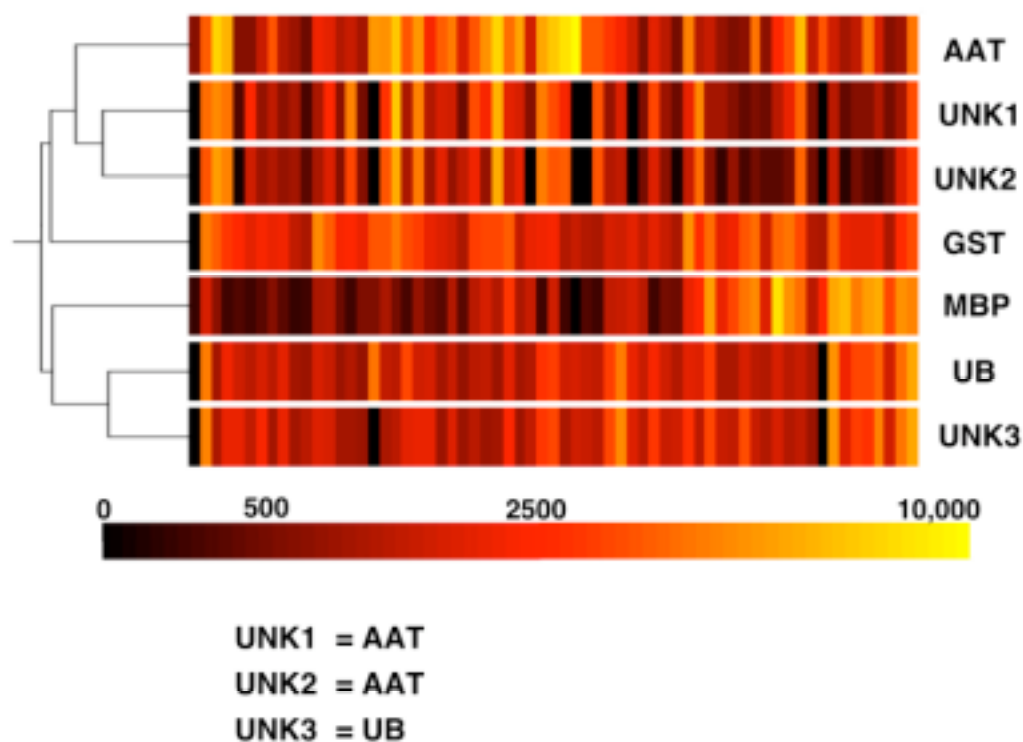


Figure 58 Blinded samples are identified based on hybridization patterns to the peptoid array.

Three unknown samples were hybridized to the peptoid array and the binding patterns of the 65 promiscuous peptoids were used to identify the protein using cluster analysis.

Discussion

We had shown previously that unique and reproducible “protein fingerprints” are obtained when a labeled protein is hybridized to a microarray comprised of 7680 peptoids (146). Similar patterns could be obtained by hybridizing mixtures of native proteins and then visualizing the pattern of a particular protein of interest with a specific

antibody and a labeled secondary antibody (146). However, these arrays were made by spotting peptoids derived from one bead/one compound libraries after separation of the beads into the wells of microtiter plates and cleavage of the peptoids. While we can create more than 1000 slides from each library, it is impossible to reproduce an array once the library is exhausted. This is highly problematic if one imagines eventually employing these arrays as diagnostic tools based on pattern recognition. The goal of this study was to ask if it was possible to fingerprint proteins using a much smaller number of peptoids. If so, then these could be synthesized individually and spotted in a spatially addressable fashion, providing an inexhaustible source of arrays.

As discussed above, analysis of three hybridization data sets revealed that almost 0.8% of the peptoids exhibited highly promiscuous binding behavior, retaining enough labeled protein in each of the three experiments to exhibit fluorescence intensities well above background. But in each case, the quantitative levels of fluorescence distinguished the proteins from one another. In other words, these 65 promiscuous peptoids provided enough discrimination to distinguish the GST, MBP and Ubiquitin fingerprints. Although, these features were able to bind to three different proteins, the critical experiment was to ask if these same features were also able to capture another protein with unknown binding properties with enough affinity to provide a distinguishing fingerprint. As is shown in Figure 57, this was indeed the case. For the serum protein AAT, 45 of the 65 features exhibit strong binding intensities (57 of the 65 features had reproducible intensities two standard deviations above the average negative control) and quantitation of these values provided a unique signature

for AAT. Furthermore, the binding pattern of these 65 features is able to accurately identify proteins in a blinded study (Figure 58). Thus, we believe that these promiscuous peptoids have considerable promise in the construction of much simpler, spatially addressable arrays for protein fingerprinting.

It is ironic that these compounds appear to be of such utility in this application. “Sticky”, non-specific protein ligands tend to be the bane of most screening experiments that aim to identify selective protein-binding agents for pharmacology/molecular genetics. This highlights the different requirements of using an array as a screening platform or a bioanalytical tool.

Finally, the results presented here also bear on how one should think about library screening experiments. At least for the case of peptoids, we have shown that 2.4-7.8% of the features on the array score as “hits” in any individual protein hybridization experiment, but as discussed above, many are quite promiscuous, binding two, three or even all four proteins employed in this study (see Venn diagram in Figure 53). These observations suggest that when one is using small molecule microarrays as a screening platform, it will be important to conduct several screens against different proteins. Comparison of these data sets would then allow the investigator to eliminate non-specific ligands prior to time-consuming biochemical characterization of the small molecule•protein complexes.

We show in this work that the 65 promiscuously binding peptoids, while lacking individual specificity, are informative and specific collectively. Interestingly, 45 of these peptoids were able to bind an unrelated fourth protein. We feel this is a significant

step towards development of new bioanalytical tools where specificity is defined not through a single specific ligand, but rather through the unique binding patterns of multiple non-specific ligands.

Materials and Methods

Direct Labeling of Proteins.

Proteins were purchased from the following commercial sources, Sigma-Aldrich (Ubiquitin), New England BioLabs Inc. (MBP), AAT (Calbiochem). GST was expressed and purified as described previously(4). Proteins were labeled using protocols described previously(4, 12, 145).

Peptoid array construction.

The details of array construction will be provided elsewhere (146).

Microarray hybridization and data acquisition.

Microarray slides were equilibrated with 1 X TBST for 15 minutes and blocked with 100-fold *E. coli* lysate in 1 ml for 1 hour at 4°C. Following blocking, microarray slides were rinsed in 1 X TBST before applying the target protein. Each protein was diluted (500nM) with 1 X TBST containing 100-fold excess of *E. coli* lysate and applied to the slides. Microarray slides were then incubated for 2 hours at 4 °C with gentle shaking. The slides were briefly rinsed once with 1X TBST followed by 3 washes with 1 X TBST (4 minutes each with gentle shaking), then dried by

centrifugation. All slides were scanned after incubation with labeled protein using a ScanArray ExpressHT scanner (Perkin Elmer, MA) using a Blue laser with an excitation wavelength of 488nm. The laser power and PMT gain was set at 100% and 70% for all slides, respectively. Emission values were collected at 519nm at a pixel resolution of 10 microns. This image was used to calculate fluorescence intensities at all peptoid features on the array.

Image/Data analysis.

Individual spot values were acquired using GenePix pro v5.0 software (Axon Instruments, CA). The mean fluorescence intensity of each spot was calculated and the mean local background fluorescence subtracted to obtain a local background subtracted fluorescence intensity value for each feature. In Excel (Microsoft Corp., WA) the threshold was calculated for each slide. The threshold was determined by averaging solvent (DMSO)-only features on the array and adding ten standard deviations to this value.

Fluorescence values were imported into GeneSpring software (Silicon Genetics, CA) for comparison and additional analysis. Only features that gave positive fluorescence values above the threshold on both arrays were used to determine intra-slide reproducibility. Reproducibility was evaluated by scatter plot analysis and a calculation of the Spearman correlation values between slides. The correlation between proteins was determined in a similar manner using features that gave above-threshold values in each data set.

Identification of pan-informational peptoids was conducted by selecting features that had above threshold values in replicate experiments. Peptoid features that met this criterion for GST, MBP, and UB protein (65 peptoids) were used for further analysis. The intensity of these 65 peptoids was evaluated for AAT, and the collection of these data in addition to the original three proteins was used to generate a visual barcode within GeneSpring. This was done by clustering the 65 peptoid data set by condition (protein).

Identification of unknown samples was performed in a blinded fashion. Three independent hybridizations were performed and data extracted as described earlier. Data were imported into GeneSpring and the 65 features for the three unknowns were clustered along with the data from GST, MBP, UB, and AAT. The resultant tree structure was interpreted based on the distance between known and unknown samples to make predictions.

Cell binding experiments to peptoid microarrays

The identification of specific cell-binding reagents has been approached by multiple screening methods and ligand-libraries. I performed pilot experiments to see if spotted peptoids would be useful for identifying cell-specific binding molecules. In these experiments I took total spleenocytes from mice and incubated them with peptoid arrays for one hour at 37 degrees. The slides were subsequently washed to remove unbound cells. The cells were then fixed with 4% PFA and visualized under a microscope. Distinct foci of cells with diameters of 150-200 um were found on the

slide, however, the inability to visualize un-hybridized spots made identification of the peptoid spots with confidence difficult. It was also unclear what cell type(s) constituted these foci. Directly-labeled anti-CD3 antibody was used to determine if these were T-cells without success.

I contributed input to the experimental design and outline of this project. I further executed the data acquisition, normalization, analysis and interpretation for this work. I created all the figure (sans Figure 51) presented in this work, and contributed to the writing of the manuscript.

BIBLIOGRAPHY

1. **Adams, C., J. Pitzer, and F. C. Minion.** 2005. In vivo expression analysis of the P97 and P102 paralog families of *Mycoplasma hyopneumoniae*. *Infect Immun* **73**:7784-7.
2. **Alibek, K., and S. Handelman.** 2000. *Biohazard*. Gardners Books.
3. **Alland, D., I. Kramnik, T. R. Weisbrod, L. Otsubo, R. Cerny, L. P. Miller, W. R. Jacobs, Jr., and B. R. Bloom.** 1998. Identification of differentially expressed mRNA in prokaryotic organisms by customized amplification libraries (DECAL): the effect of isoniazid on gene expression in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **95**:13227-32.
4. **Alluri, P. G., M. M. Reddy, K. Bachhawat-Sikder, H. J. Olivos, and T. Kodadek.** 2003. Isolation of protein ligands from large peptoid libraries. *J Am Chem Soc* **125**:13995-4004.
5. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. *J Mol Biol* **215**:403-10.
6. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**:3389-402.
7. **Anderson, G. W., Jr., D. G. Heath, C. R. Bolt, S. L. Welkos, and A. M. Friedlander.** 1998. Short- and long-term efficacy of single-dose subunit vaccines against *Yersinia pestis* in mice. *American Journal of Tropical Medicine & Hygiene*. **58**:793-9.
8. **Anderson, N. L., and N. G. Anderson.** 2002. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* **1**:845-67.
9. **Andrews, G. P., D. G. Heath, G. W. Anderson, Jr., S. L. Welkos, and A. M. Friedlander.** 1996. Fraction 1 capsular antigen (F1) purification from *Yersinia pestis* CO92 and from an *Escherichia coli* recombinant strain and efficacy against lethal plague challenge. *Infection & Immunity*. **64**:2180-7.
10. **Anisimov, A. P., S. V. Dentovskaya, G. M. Titareva, I. V. Bakhteeva, R. Z. Shaikhutdinova, S. V. Balakhonov, B. Lindner, N. A. Kocharova, S. N.**

- Senchenkova, O. Holst, G. B. Pier, and Y. A. Knirel.** 2005. Intraspecies and temperature-dependent variations in susceptibility of *Yersinia pestis* to the bactericidal action of serum and to polymyxin B. *Infect Immun* **73**:7324-31.
11. **Anisimov, A. P., L. E. Lindler, and G. B. Pier.** 2004. Intraspecific diversity of *Yersinia pestis*. *Clin Microbiol Rev* **17**:434-64.
12. **Bachhawat-Sikder, K., and T. Kodadek.** 2003. Mixed-element capture agents: a simple strategy for the construction of synthetic, high-affinity protein capture ligands. *J Am Chem Soc* **125**:9550-1.
13. **Bahmanyar, M., and D. Cavanagh.** 1976. Plague Manual, p. 1-76. *In* G. World Health Organization (ed.).
14. **Bailey, T. L., and M. Gribskov.** 1997. Score distributions for simultaneous matching to multiple motifs. *J Comput Biol* **4**:45-59.
15. **Baldini, L., A. J. Wilson, J. Hong, and A. D. Hamilton.** 2004. Pattern-based detection of different proteins using an array of fluorescent protein surface receptors. *J Am Chem Soc* **126**:5656-7.
16. **Balog, R. P., Y. E. de Souza, H. M. Tang, G. M. DeMasellis, B. Gao, A. Avila, D. J. Gaban, D. Mittelman, J. D. Minna, K. J. Luebke, and H. R. Garner.** 2002. Parallel assessment of CpG methylation by two-color hybridization with oligonucleotide arrays. *Anal Biochem* **309**:301-10.
17. **Barglow, K. T., and B. F. Cravatt.** 2004. Discovering disease-associated enzymes by proteome reactivity profiling. *Chem Biol* **11**:1523-31.
18. **Barry, M. A., D. P. G. Howell, H. A. Andersson, J. L. Chen, and R. A. K. Singh.** 2004. Expression library immunization to discover and improve vaccine antigens. *Immunological Reviews* **199**:68-83.
19. **Baugh, L. R., A. A. Hill, E. L. Brown, and C. P. Hunter.** 2001. Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res* **29**:E29.
20. **Benner, G. E., G. P. Andrews, W. R. Byrne, S. D. Strachan, A. K. Sample, D. G. Heath, and A. M. Friedlander.** 1999. Immune response to *Yersinia* outer

proteins and other *Yersinia pestis* antigens after experimental plague infection in mice. *Infection & Immunity*. **67**:1922-8.

21. **Bennett, A. M., R. J. Phillipotts, S. D. Perkins, S. C. Jacobs, and E. D. Williamson.** 1999. Gene gun mediated vaccination is superior to manual delivery for immunisation with DNA vaccines expressing protective antigens from *Yersinia pestis* or Venezuelan Equine Encephalitis virus. *Vaccine*. **18**:588-96.
22. **Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler.** 2003. GenBank. *Nucleic Acids Res* **31**:23-7.
23. **Besemer, J., and M. Borodovsky.** 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res* **33**:W451-4.
24. **Besemer, J., A. Lomsadze, and M. Borodovsky.** 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* **29**:2607-18.
25. **Bliska, J. B., K. L. Guan, J. E. Dixon, and S. Falkow.** 1991. Tyrosine phosphate hydrolysis of host proteins by an essential *Yersinia* virulence determinant. *Proc Natl Acad Sci U S A* **88**:1187-91.
26. **Boland, A., and G. R. Cornelis.** 1998. Role of YopP in suppression of tumor necrosis factor alpha release by macrophages during *Yersinia* infection. *Infect Immun* **66**:1878-84.
27. **Bonacorsi, S. P., M. R. Scavizzi, A. Guiyoule, J. H. Amouroux, and E. Carniel.** 1994. Assessment of a fluoroquinolone, three beta-lactams, two aminoglycosides, and a cycline in treatment of murine *Yersinia pestis* infection. *Antimicrob Agents Chemother* **38**:481-6.
28. **Boyce, J. D., I. Wilkie, M. Harper, M. L. Paustian, V. Kapur, and B. Adler.** 2002. Genomic scale analysis of *Pasteurella multocida* gene expression during growth within the natural chicken host. *Infect Immun* **70**:6871-9.
29. **Boyce, J. D., I. Wilkie, M. Harper, M. L. Paustian, V. Kapur, and B. Adler.** 2004. Genomic-scale analysis of *Pasteurella multocida* gene expression during growth within liver tissue of chickens with fowl cholera. *Microbes Infect* **6**:290-8.

30. **Brubaker, R. R.** 2003. Interleukin-10 and inhibition of innate immunity to *Yersinia*: roles of Yops and LcrV (V antigen). *Infect Immun* **71**:3673-81.
31. **Bullifent, H. L., K. F. Griffin, S. M. Jones, A. Yates, L. Harrington, and R. W. Titball.** 2000. Antibody responses to *Yersinia pestis* F1-antigen expressed in *Salmonella typhimurium* aroA from in vivo-inducible promoters. *Vaccine* **18**:2668-76.
32. **Butler, T.** 1983. Plague and other *Yersinia* infections. Plenum Press, New York, N.Y.
33. **Cavanaugh, D. C., and J. D. Marshall, Jr.** 1972. The influence of climate on the seasonal prevalence of plague in the Republic of Vietnam. *J Wildl Dis* **8**:85-94.
34. **Cavanaugh, D. C., and R. Randall.** 1959. The role of multiplication of *Pasteurella pestis* in mononuclear phagocytes in the pathogenesis of flea-borne plague. *J Immunol* **83**:348-63.
35. **Chambers, R. S., and S. A. Johnston.** 2003. High-level generation of polyclonal antibodies by genetic immunization. *Nature Biotechnology* **21**:1088-1092.
36. **Chenchik, A., Y. Y. Zhu, L. Diatchenko, R. Li, J. Hill, and P. D. Siebert.** 1998. Generation and Use of High-Quality cDNA from Small Amounts of Total RNA by SMART PCR, p. 305-319. *In* P. D. Siebert and J. Larrick (ed.), *Gene Cloning and Analysis by RT-PCR*. Biotechnology Books, Natick, MA.
37. **Chiang, S. L., J. J. Mekalanos, and D. W. Holden.** 1999. In vivo genetic analysis of bacterial virulence. *Annu Rev Microbiol* **53**:129-54.
38. **Christopher, G. W., T. J. Cieslak, J. A. Pavlin, and E. M. Eitzen, Jr.** 1997. Biological warfare. A historical perspective. *Jama* **278**:412-7.
39. **Cohen, R. J., and J. L. Stockard.** 1967. Pneumonic plague in an untreated plague-vaccinated individual. *Jama* **202**:365-6.
40. **Cole, S. T., and C. Buchrieser.** 2001. Bacterial genomics. A plague o' both your hosts. *Nature* **413**:467, 469-70.
41. **Cornelis, G.** 1975. Distribution of beta-lactamases A and B in some groups of *Yersinia enterocolitica* and their role in resistance. *J Gen Microbiol* **91**:391-402.

42. **Cornelis, G. R.** 2002. Yersinia type III secretion: send in the effectors. *Journal of Cell Biology*. **158**:401-8.
43. **Cornelis, G. R.** 2002. The Yersinia Ysc-Yop virulence apparatus. *Int J Med Microbiol* **291**:455-62.
44. **Cornelis, G. R., A. Boland, A. P. Boyd, C. Geuijen, M. Iriarte, C. Neyt, M. P. Sory, and I. Stainier.** 1998. The virulence plasmid of Yersinia, an antihost genome. *Microbiol Mol Biol Rev* **62**:1315-52.
45. **Craven, R. B., G. O. Maupin, M. L. Beard, T. J. Quan, and A. M. Barnes.** 1993. Reported cases of human plague infections in the United States, 1970-1991. *J Med Entomol* **30**:758-61.
46. **Crook, L. D., and B. Tempest.** 1992. Plague. A clinical review of 27 cases. *Arch Intern Med* **152**:1253-6.
47. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**:4636-41.
48. **Delcher, A. L., S. Kasif, R. D. Fleischmann, J. Peterson, O. White, and S. L. Salzberg.** 1999. Alignment of whole genomes. *Nucleic Acids Res* **27**:2369-76.
49. **Derewenda, U., A. Mateja, Y. Devedjiev, K. M. Routzahn, A. G. Evdokimov, Z. S. Derewenda, and D. S. Waugh.** 2004. The structure of Yersinia pestis V-antigen, an essential virulence factor and mediator of immunity against plague. *Structure (Camb)* **12**:301-6.
50. **Drancourt, M., L. Houhamdi, and D. Raoult.** 2006. Yersinia pestis as a telluric, human ectoparasite-borne organism. *Lancet Infect Dis* **6**:234-41.
51. **Duplaix, N.** 1988. Fleas - the leathal leapers, p. 672-694, *National Geographic*, vol. 173.
52. **Eisen, M. B., and P. O. Brown.** 1999. DNA arrays for analysis of gene expression. *Methods Enzymol* **303**:179-205.
53. **Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain.** 2002. Stochastic gene expression in a single cell. *Science* **297**:1183-6.

54. **Enscore, R. E., B. J. Biggerstaff, T. L. Brown, R. E. Fulgham, P. J. Reynolds, D. M. Engelthaler, C. E. Levy, R. R. Parmenter, J. A. Montenieri, J. E. Cheek, R. K. Grinnell, P. J. Ettestad, and K. L. Gage.** 2002. Modeling relationships between climate and the frequency of human plague cases in the southwestern United States, 1960-1997. *Am J Trop Med Hyg* **66**:186-96.
55. **Eyles, J. E., E. D. Williamson, I. D. Spiers, and H. O. Alpar.** 2000. Protection studies following bronchopulmonary and intramuscular immunisation with yersinia pestis F1 and V subunit vaccines coencapsulated in biodegradable microspheres: a comparison of efficacy. *Vaccine* **18**:3266-71.
56. **Flashner, Y., E. Mamroud, A. Tidhar, R. Ber, M. Aftalion, D. Gur, S. Lazar, A. Zvi, T. Bino, N. Ariel, B. Velan, A. Shafferman, and S. Cohen.** 2004. Generation of Yersinia pestis attenuated strains by signature-tagged mutagenesis in search of novel vaccine candidates. *Infect Immun* **72**:908-15.
57. **Friedlander, A. M., P. R. Pittman, and G. W. Parker.** 1999. Anthrax vaccine: evidence for safety and efficacy against inhalational anthrax. *Jama* **282**:2104-6.
58. **Fung, E. T., V. Thulasiraman, S. R. Weinberger, and E. A. Dalmasso.** 2001. Protein biochips for differential profiling. *Curr Opin Biotechnol* **12**:65-9.
59. **Gage, K. L., D. T. Dennis, K. A. Orloski, P. Ettestad, T. L. Brown, P. J. Reynolds, W. J. Pape, C. L. Fritz, L. G. Carter, and J. D. Stein.** 2000. Cases of cat-associated human plague in the Western US, 1977-1998. *Clin Infect Dis* **30**:893-900.
60. **Gage, K. L., and M. Y. Kosoy.** 2005. Natural history of plague: perspectives from more than a century of research. *Annu Rev Entomol* **50**:505-28.
61. **Gage, K. L., S. E. Lance, D. T. Dennis, and J. A. Montenieri.** 1992. Human plague in the United States: a review of cases from 1988-1992 with comments on the likelihood of increased plague activity. *Border Epidemiological Bulliten* **19**:1-10.
62. **Galimand, M., A. Guiyoule, G. Gerbaud, B. Rasoamanana, S. Chanteau, E. Carniel, and P. Courvalin.** 1997. Multidrug resistance in Yersinia pestis mediated by a transferable plasmid. *N Engl J Med* **337**:677-80.

63. **Galimand, M., A. Guiyoule, G. Gerbaud, B. Rasoamanana, S. Chanteau, E. Carniel, and P. Courvalin.** 1997. Multidrug resistance in *Yersinia pestis* mediated by a transferable plasmid.[comment]. *New England Journal of Medicine*. **337**:677-80.
64. **Garner, H. R., R. P. Balog, and K. J. Luebke.** 2002. The evolution of custom microarray manufacture. *IEEE Eng Med Biol Mag* **21**:123-5.
65. **Gish, W., and D. J. States.** 1993. Identification of protein coding regions by database similarity search. *Nat Genet* **3**:266-72.
66. **Glynn, A., L. C. Freytag, and J. D. Clements.** 2005. Effect of homologous and heterologous prime-boost on the immune response to recombinant plague antigens. *Vaccine* **23**:1957-65.
67. **Glynn, A., C. J. Roy, B. S. Powell, J. J. Adamovicz, L. C. Freytag, and J. D. Clements.** 2005. Protection against aerosolized *Yersinia pestis* challenge following homologous and heterologous prime-boost with recombinant plague antigens. *Infect Immun* **73**:5256-61.
68. **Gomes-Solecki, M. J., A. G. Savitt, R. Rowehl, J. D. Glass, J. B. Bliska, and R. J. Dattwyler.** 2005. LcrV capture enzyme-linked immunosorbent assay for detection of *Yersinia pestis* from human samples. *Clin Diagn Lab Immunol* **12**:339-46.
69. **Gottfried, R. S.** 1983. The black death. Natural and human disaster in medieval europe. The Free Press, New York, N.Y.
70. **Grandi, G.** 2001. Antibacterial vaccine design using genomics and proteomics. *Trends in Biotechnology*. **19**:181-8.
71. **Grandi, G.** 2003. Rational antibacterial vaccine design through genomic technologies. *International Journal for Parasitology*. **33**:615-20.
72. **Grosfeld, H., S. Cohen, T. Bino, Y. Flashner, R. Ber, E. Mamroud, C. Kronman, A. Shafferman, and B. Velan.** 2003. Effective protective immunity to *Yersinia pestis* infection conferred by DNA vaccine coding for derivatives of the F1 capsular antigen. *Infection & Immunity*. **71**:374-83.

73. **Guiyoule, A., G. Gerbaud, C. Buchrieser, M. Galimand, L. Rahalison, S. Chanteau, P. Courvalin, and E. Carniel.** 2001. Transferable plasmid-mediated resistance to streptomycin in a clinical isolate of *Yersinia pestis*. *Emerging Infectious Diseases* **7**:43-8.
74. **Han, Y., D. Zhou, X. Pang, Y. Song, L. Zhang, J. Bao, Z. Tong, J. Wang, Z. Guo, J. Zhai, Z. Du, X. Wang, X. Zhang, P. Huang, and R. Yang.** 2004. Microarray analysis of temperature-induced transcriptome of *Yersinia pestis*. *Microbiol Immunol* **48**:791-805.
75. **Han, Y., D. Zhou, X. Pang, L. Zhang, Y. Song, Z. Tong, J. Bao, E. Dai, J. Wang, Z. Guo, J. Zhai, Z. Du, X. Wang, P. Huang, and R. Yang.** 2005. Comparative transcriptome analysis of *Yersinia pestis* in response to hyperosmotic and high-salinity stress. *Res Microbiol* **156**:403-15.
76. **Han, Y., D. Zhou, X. Pang, L. Zhang, Y. Song, Z. Tong, J. Bao, E. Dai, J. Wang, Z. Guo, J. Zhai, Z. Du, X. Wang, P. Huang, and R. Yang.** 2005. DNA microarray analysis of the heat- and cold-shock stimulons in *Yersinia pestis*. *Microbes Infect* **7**:335-48.
77. **Heath, D. G., G. W. Anderson, Jr., J. M. Mauro, S. L. Welkos, G. P. Andrews, J. Adamovicz, and A. M. Friedlander.** 1998. Protection against experimental bubonic and pneumonic plague by a recombinant capsular F1-V antigen fusion protein vaccine. *Vaccine* **16**:1131-7.
78. **Hegde, P., R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. E. Hughes, E. Snesrud, N. Lee, and J. Quackenbush.** 2000. A concise guide to cDNA microarray analysis. *Biotechniques* **29**:548-50, 552-4, 556 passim.
79. **Hensel, M., J. E. Shea, C. Gleeson, M. D. Jones, E. Dalton, and D. W. Holden.** 1995. Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**:400-3.
80. **Hill, J., J. E. Eyles, S. J. Elvin, G. D. Healey, R. A. Lukaszewski, and R. W. Titball.** 2006. Administration of Antibody to the Lung Protects Mice against Pneumonic Plague. *Infect Immun* **74**:3068-70.

81. **Hinchliffe, S. J., K. E. Isherwood, R. A. Stabler, M. B. Prentice, A. Rakin, R. A. Nichols, P. C. Oyston, J. Hinds, R. W. Titball, and B. W. Wren.** 2003. Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis*. *Genome Res* **13**:2018-29.
82. **Hinnebusch, B. J.** 2005. The evolution of flea-borne transmission in *Yersinia pestis*. *Curr Issues Mol Biol* **7**:197-212.
83. **Hinnebusch, B. J., A. E. Rudolph, P. Cherepanov, J. E. Dixon, T. G. Schwan, and A. Forsberg.** 2002. Role of *Yersinia murine* toxin in survival of *Yersinia pestis* in the midgut of the flea vector. *Science* **296**:733-5.
84. **Hinnebusch, J., P. Cherepanov, Y. Du, A. Rudolph, J. D. Dixon, T. Schwan, and A. Forsberg.** 2000. Murine toxin of *Yersinia pestis* shows phospholipase D activity but is not required for virulence in mice. *Int J Med Microbiol* **290**:483-7.
85. **Hoheisel, J. D.** 2006. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* **7**:200-10.
86. **Hull, H. F., J. M. Montes, and J. M. Mann.** 1987. Septicemic plague in New Mexico. *J Infect Dis* **155**:113-8.
87. **Inglesby, T. V., D. T. Dennis, D. A. Henderson, J. G. Bartlett, M. S. Ascher, E. Eitzen, A. D. Fine, A. M. Friedlander, J. Hauer, J. F. Koerner, M. Layton, J. McDade, M. T. Osterholm, T. O'Toole, G. Parker, T. M. Perl, P. K. Russell, M. Schoch-Spana, and K. Tonat.** 2000. Plague as a biological weapon: medical and public health management. Working Group on Civilian Biodefense. *Jama* **283**:2281-90.
88. **Jarrett, C. O., E. Deak, K. E. Isherwood, P. C. Oyston, E. R. Fischer, A. R. Whitney, S. D. Kobayashi, F. R. DeLeo, and B. J. Hinnebusch.** 2004. Transmission of *Yersinia pestis* from an infectious biofilm in the flea vector. *J Infect Dis* **190**:783-92.
89. **Jones, H. A., J. W. Lillard, Jr., and R. D. Perry.** 1999. HmsT, a protein essential for expression of the haemin storage (Hms+) phenotype of *Yersinia pestis*. *Microbiology* **145 (Pt 8)**:2117-28.

90. **Jones, S. M., F. Day, A. J. Stagg, and E. D. Williamson.** 2000. Protection conferred by a fully recombinant sub-unit vaccine against *Yersinia pestis* in male and female mice of four inbred strains. *Vaccine*. **19**:358-66.
91. **Kane, M. D., T. A. Jatkoe, C. R. Stumpf, J. Lu, J. D. Thomas, and S. J. Madore.** 2000. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* **28**:4552-7.
92. **Kapranov, P., S. E. Cawley, J. Drenkow, S. Bekiranov, R. L. Strausberg, S. P. Fodor, and T. R. Gingeras.** 2002. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**:916-9.
93. **Karlyshev, A. V., P. C. Oyston, K. Williams, G. C. Clark, R. W. Titball, E. A. Winzeler, and B. W. Wren.** 2001. Application of high-density array-based signature-tagged mutagenesis to discover novel *Yersinia* virulence-associated genes. *Infect Immun* **69**:7810-9.
94. **Kato-Maeda, M., Q. Gao, and P. M. Small.** 2001. Microarray analysis of pathogens and their interaction with hosts. *Cell Microbiol* **3**:713-9.
95. **Kerschen, E. J., D. A. Cohen, A. M. Kaplan, and S. C. Straley.** 2004. The plague virulence protein YopM targets the innate immune response by causing a global depletion of NK cells. *Infect Immun* **72**:4589-602.
96. **Kirillina, O., J. D. Fetherston, A. G. Bobrov, J. Abney, and R. D. Perry.** 2004. HmsP, a putative phosphodiesterase, and HmsT, a putative diguanylate cyclase, control Hms-dependent biofilm formation in *Yersinia pestis*. *Mol Microbiol* **54**:75-88.
97. **Knirel, Y. A., S. V. Dentovskaya, S. N. Senchenkova, R. Z. Shaikhutdinova, N. A. Kocharova, and A. P. Anisimov.** 2006. Structural features and structural variability of the lipopolysaccharide of *Yersinia pestis*, the cause of plague. *J Endotoxin Res* **12**:3-9.
98. **Kodadek, T.** 2001. Protein microarrays: prospects and problems. *Chem Biol* **8**:105-15.

99. **Koornhof, H. J., R. A. Smego, Jr., and M. Nicol.** 1999. Yersiniosis. II: The pathogenesis of Yersinia infections. *European Journal of Clinical Microbiology & Infectious Diseases*. **18**:87-112.
100. **Lathem, W. W., S. D. Crosby, V. L. Miller, and W. E. Goldman.** 2005. Progression of primary pneumonic plague: a mouse model of infection, pathology, and bacterial transcriptional activity. *Proc Natl Acad Sci U S A* **102**:17786-91.
101. **Leary, S. E., K. F. Griffin, E. E. Galyov, J. Hewer, E. D. Williamson, A. Holmstrom, Forsberg, and R. W. Titball.** 1999. Yersinia outer proteins (YOPS) E, K and N are antigenic but non-protective compared to V antigen, in a murine model of bubonic plague. *Microbial Pathogenesis*. **26**:159-69.
102. **Leary, S. E. C., E. D. Williamson, K. F. Griffin, P. Russell, S. M. Eley, and R. W. Titball.** 1995. Active immunization with recombinant v antigen from yersinia pestis protects mice against plague. *Infection & Immunity* **63**:2854-2858.
103. **Lerm, M., G. Schmidt, and K. Aktories.** 2000. Bacterial protein toxins targeting rho GTPases. *FEMS Microbiol Lett* **188**:1-6.
104. **Li, F., and G. D. Stormo.** 2001. Selection of optimal DNA oligos for gene expression arrays. *Bioinformatics* **17**:1067-76.
105. **Lindler, L. E., G. V. Plano, V. Burland, G. F. Mayhew, and F. R. Blattner.** 1998. Complete DNA sequence and detailed analysis of the Yersinia pestis KIM5 plasmid encoding murine toxin and capsular antigen. *Infect Immun* **66**:5731-42.
106. **Little, S. F., and B. E. Ivins.** 1999. Molecular pathogenesis of Bacillus anthracis infection. *Microbes Infect* **1**:131-9.
107. **Lorange, E. A., B. L. Race, F. Sebbane, and B. Joseph Hinnebusch.** 2005. Poor Vector Competence of Fleas and the Evolution of Hypervirulence in Yersinia pestis. *J Infect Dis* **191**:1907-1912.
108. **Lowell, J. L., D. M. Wagner, B. Atshabar, M. F. Antolin, A. J. Vogler, P. Keim, M. C. Chu, and K. L. Gage.** 2005. Identifying sources of human exposure to plague. *J Clin Microbiol* **43**:650-6.

109. **Lucchini, S., A. Thompson, and J. C. Hinton.** 2001. Microarrays for microbiologists. *Microbiology* **147**:1403-14.
110. **Luebke, K. J., R. P. Balog, and H. R. Garner.** 2003. Prioritized selection of oligodeoxyribonucleotide probes for efficient hybridization to RNA transcripts. *Nucleic Acids Res* **31**:750-8.
111. **Lukashin, A. V., and M. Borodovsky.** 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* **26**:1107-15.
112. **Lukaszewski, R. A., D. J. Kenny, R. Taylor, D. G. Rees, M. G. Hartley, and P. C. Oyston.** 2005. Pathogenesis of *Yersinia pestis* infection in BALB/c mice: effects on host macrophages and neutrophils. *Infect Immun* **73**:7142-50.
113. **Madden, T. L., R. L. Tatusov, and J. Zhang.** 1996. Applications of network BLAST server. *Methods Enzymol* **266**:131-41.
114. **Mahan, M. J., J. M. Slauch, and J. J. Mekalanos.** 1993. Selection of bacterial virulence genes that are specifically induced in host tissues. *Science* **259**:686-8.
115. **Managan, J. A., I. M. Monahan, and P. D. Butcher.** 2002. *Functional Microbial Genetics*, vol. 33. Academic, Amsterdam.
116. **Marketon, M. M., R. W. DePaolo, K. L. DeBord, B. Jabri, and O. Schneewind.** 2005. Plague bacteria target immune cells during infection. *Science* **309**:1739-41.
117. **Matz, M. V., N. O. Alieva, A. Chenchik, and S. Lukyanov.** 2003. Amplification of cDNA ends using PCR suppression effect and step-out PCR. *Methods Mol Biol* **221**:41-9.
118. **Mecsas, J., I. Bilis, and S. Falkow.** 2001. Identification of attenuated *Yersinia pseudotuberculosis* strains and characterization of an orogastric infection in BALB/c mice on day 5 postinfection by signature-tagged mutagenesis. *Infect Immun* **69**:2779-87.
119. **Meyer, K. F.** 1970. Effectiveness of live or killed plague vaccines in man. *Bull World Health Organ* **42**:653-66.
120. **Meyer, K. F.** 1961. Pneumonic plague. *Bacteriol Rev* **25**:249-61.

121. **Motin, V. L., A. M. Georgescu, J. P. Fitch, P. P. Gu, D. O. Nelson, S. L. Mabery, J. B. Garnham, B. A. Sokhansanj, L. L. Ott, M. A. Coleman, J. M. Elliott, L. M. Kegelmeyer, A. J. Wyrobek, T. R. Slezak, R. R. Brubaker, and E. Garcia.** 2004. Temporal global changes in gene expression during temperature transition in *Yersinia pestis*. *J Bacteriol* **186**:6298-305.
122. **Motley, S. T., B. J. Morrow, X. Liu, I. L. Dodge, A. Vitiello, C. K. Ward, and K. J. Shaw.** 2004. Simultaneous analysis of host and pathogen interactions during an in vivo infection reveals local induction of host acute phase response proteins, a novel bacterial stress response, and evidence of a host-imposed metal ion limited environment. *Cell Microbiol* **6**:849-65.
123. **Ohyama, H., X. Zhang, Y. Kohno, I. Alevizos, M. Posner, D. T. Wong, and R. Todd.** 2000. Laser capture microdissection-generated target sample for high-density oligonucleotide array hybridization. *Biotechniques* **29**:530-6.
124. **Okinaka, R., K. Cloud, O. Hampton, A. Hoffmaster, K. Hill, P. Keim, T. Koehler, G. Lamke, S. Kumano, D. Manter, Y. Martinez, D. Ricke, R. Svensson, and P. Jackson.** 1999. Sequence, assembly and analysis of pX01 and pX02. *J Appl Microbiol* **87**:261-2.
125. **Olivos, H. J., P. G. Alluri, M. M. Reddy, D. Salony, and T. Kodadek.** 2002. Microwave-assisted solid-phase synthesis of peptoids. *Org Lett* **4**:4057-9.
126. **Organization, W. H.** 1995. Human plague in 1993. *Weekly Epidemiological Record*. **70**:45-48.
127. **Oyston, P. C., and K. E. Isherwood.** 2005. The many and varied niches occupied by *Yersinia pestis* as an arthropod-vectored zoonotic pathogen. *Antonie Van Leeuwenhoek* **87**:171-7.
128. **Parent, M. A., K. N. Berggren, L. W. Kummer, L. B. Wilhelm, F. M. Szaba, I. K. Mullarky, and S. T. Smiley.** 2005. Cell-mediated protection against pulmonary *Yersinia pestis* infection. *Infect Immun* **73**:7304-10.
129. **Parent, M. A., K. N. Berggren, I. K. Mullarky, F. M. Szaba, L. W. Kummer, J. J. Adamovicz, and S. T. Smiley.** 2005. *Yersinia pestis* V protein epitopes recognized by CD4 T cells. *Infect Immun* **73**:2197-204.

130. **Parkhill, J., B. W. Wren, N. R. Thomson, R. W. Titball, M. T. G. Holden, M. B. Prentice, M. Sebaihia, K. D. James, C. Churcher, K. L. Mungall, S. Baker, D. Basham, S. D. Bentley, K. Brooks, A. M. Cerdeno-Tarraga, T. Chillingworth, A. Cronin, R. M. Davies, P. Davis, G. Dougan, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, A. V. Karlyshev, and et al.** 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413**:523-527.
131. **Parmenter, R. R., E. P. Yadav, C. A. Parmenter, P. Ettestad, and K. L. Gage.** 1999. Incidence of plague associated with increased winter-spring precipitation in New Mexico. *Am J Trop Med Hyg* **61**:814-21.
132. **Paweletz, C. P., J. W. Gillespie, D. K. Ornstein, N. L. Simone, M. R. Brown, K. A. Cole, Q.-H. Wang, J. Huang, N. Hu, T.-T. Yip, W. E. Rich, E. C. Kohn, W. M. Linehan, T. Weber, P. Taylor, M. R. Emmert-Buck, L. A. Liotta, and E. F. P. III.** 2000. Rapid protein display profiling of cancer progression directly from human tissue using a protein biochip . *Drug Development Research* **49**:34-42.
133. **Perry, R. D., A. G. Bobrov, O. Kirillina, H. A. Jones, L. Pedersen, J. Abney, and J. D. Fetherston.** 2004. Temperature regulation of the hemin storage (Hms+) phenotype of *Yersinia pestis* is posttranscriptional. *J Bacteriol* **186**:1638-47.
134. **Perry, R. D., and J. D. Fetherston.** 1997. *Yersinia pestis*--etiologic agent of plague. *Clinical Microbiology Reviews*. **10**:35-66.
135. **Perry, R. D., S. C. Straley, J. D. Fetherston, D. J. Rose, J. Gregor, and F. R. Blattner.** 1998. DNA sequencing and analysis of the low-Ca²⁺-response plasmid pCD1 of *Yersinia pestis* KIM5. *Infect Immun* **66**:4611-23.
136. **Petricoin, E., J. Wulfkuhle, V. Espina, and L. A. Liotta.** 2004. Clinical proteomics: revolutionizing disease detection and patient tailoring therapy. *J Proteome Res* **3**:209-17.
137. **Petricoin, E. F., A. M. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, E. C. Kohn, and L. A. Liotta.** 2002. Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* **359**:572-7.

138. **Pfaffl, M. W.** 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**:e45.
139. **Qiu, J., D. Zhou, Y. Han, L. Zhang, Z. Tong, Y. Song, E. Dai, B. Li, J. Wang, Z. Guo, J. Zhai, Z. Du, X. Wang, and R. Yang.** 2005. Global gene expression profile of *Yersinia pestis* induced by streptomycin. *FEMS Microbiol Lett* **243**:489-96.
140. **Raddatz, G., M. Dehio, T. F. Meyer, and C. Dehio.** 2001. PrimeArray: genome-scale primer design for DNA-microarray construction. *Bioinformatics* **17**:98-9.
141. **Rakin, A., P. Urbitsch, and J. Heesemann.** 1995. Evidence for two evolutionary lineages of highly pathogenic *Yersinia* species. *J Bacteriol* **177**:2292-8.
142. **Ramalingaswami, V.** 2001. Psychosocial effects of the 1994 plague outbreak in Surat, India. *Mil Med* **166**:29-30.
143. **Rayner, S., S. Brignac, R. Bumeister, Y. Belosludtsev, T. Ward, O. Grant, K. O'Brien, G. A. Evans, and H. R. Garner.** 1998. MerMade: an oligodeoxyribonucleotide synthesizer for high throughput oligonucleotide production in dual 96-well plates. *Genome Res* **8**:741-7.
144. **Read, T. D., S. N. Peterson, N. Tourasse, L. W. Baillie, I. T. Paulsen, K. E. Nelson, H. Tettelin, D. E. Fouts, J. A. Eisen, S. R. Gill, E. K. Holtzapple, O. A. Okstad, E. Helgason, J. Rilstone, M. Wu, J. F. Kolonay, M. J. Beanan, R. J. Dodson, L. M. Brinkac, M. Gwinn, R. T. DeBoy, R. Madpu, S. C. Daugherty, A. S. Durkin, D. H. Haft, W. C. Nelson, J. D. Peterson, M. Pop, H. M. Khouri, D. Radune, J. L. Benton, Y. Mahamoud, L. Jiang, I. R. Hance, J. F. Weidman, K. J. Berry, R. D. Plaut, A. M. Wolf, K. L. Watkins, W. C. Nierman, A. Hazen, R. Cline, C. Redmond, J. E. Thwaite, O. White, S. L. Salzberg, B. Thomason, A. M. Friedlander, T. M. Koehler, P. C. Hanna, A. B. Kolsto, and C. M. Fraser.** 2003. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **423**:81-6.
145. **Reddy, M. M., K. Bachhawat-Sikder, and T. Kodadek.** 2004. Transformation of low-affinity lead compounds into high-affinity protein capture agents. *Chem Biol* **11**:1127-37.

146. **Reddy, M. M., and T. Kodadek.** 2005. Protein "fingerprinting" in complex mixtures with peptoid microarrays. *Proc Natl Acad Sci U S A* **102**:12672-7.
147. **Reed, D. S., and M. J. Martinez.** 2006. Respiratory immunity is an important component of protection elicited by subunit vaccination against pneumonic plague. *Vaccine* **24**:2283-9.
148. **Revel, A. T., A. M. Talaat, and M. V. Norgard.** 2002. DNA microarray analysis of differential gene expression in *Borrelia burgdorferi*, the Lyme disease spirochete. *Proc Natl Acad Sci U S A* **99**:1562-7.
149. **Rinn, J. L., G. Euskirchen, P. Bertone, R. Martone, N. M. Luscombe, S. Hartman, P. M. Harrison, F. K. Nelson, P. Miller, M. Gerstein, S. Weissman, and M. Snyder.** 2003. The transcriptional activity of human Chromosome 22. *Genes Dev* **17**:529-40.
150. **Robinson, V. L., P. C. Oyston, and R. W. Titball.** 2005. A dam mutant of *Yersinia pestis* is attenuated and induces protection against plague. *FEMS Microbiol Lett* **252**:251-6.
151. **Rocke, T. E., J. Mencher, S. R. Smith, A. M. Friedlander, G. P. Andrews, and L. A. Baeten.** 2004. Recombinant F1-V fusion protein protects black-footed ferrets (*Mustela nigripes*) against virulent *Yersinia pestis* infection. *J Zoo Wildl Med* **35**:142-6.
152. **Rozen, S., and H. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**:365-86.
153. **Rudolph, A. E., J. A. Stuckey, Y. Zhao, H. R. Matthews, W. A. Patton, J. Moss, and J. E. Dixon.** 1999. Expression, characterization, and mutagenesis of the *Yersinia pestis* murine toxin, a phospholipase D superfamily member. *J Biol Chem* **274**:11824-31.
154. **Sabhnani, L., and D. N. Rao.** 2000. Identification of immunodominant epitope of F1 antigen of *Yersinia pestis*. *FEMS Immunology & Medical Microbiology*. **27**:155-62.

155. **Salzberg, S. L., M. Pertea, A. L. Delcher, M. J. Gardner, and H. Tettelin.** 1999. Interpolated Markov models for eukaryotic gene finding. *Genomics* **59**:24-31.
156. **Santi, L., A. Giritch, C. J. Roy, S. Marillonnet, V. Klimyuk, Y. Gleba, R. Webb, C. J. Arntzen, and H. S. Mason.** 2006. Protection conferred by recombinant *Yersinia pestis* antigens produced by a rapid and highly scalable plant expression system. *Proc Natl Acad Sci U S A* **103**:861-6.
157. **Schageman, J. J., C. J. Horton, S. Niu, H. R. Garner, and A. Pertsemlidis.** 2004. ELXR: a resource for rapid exon-directed sequence analysis. *Genome Biol* **5**:R36.
158. **Schmittgen, T. D.** 2001. Real-time quantitative PCR. *Methods* **25**:383-5.
159. **Schramm, G., I. Bruchhaus, and T. Roeder.** 2000. A simple and reliable 5'-RACE approach. *Nucleic Acids Res* **28**:E96.
160. **Sebbane, F., D. Gardner, D. Long, B. B. Gowen, and B. J. Hinnebusch.** 2005. Kinetics of disease progression and host response in a rat model of bubonic plague. *Am J Pathol* **166**:1427-39.
161. **Selinger, D. W., K. J. Cheung, R. Mei, E. M. Johansson, C. S. Richmond, F. R. Blattner, D. J. Lockhart, and G. M. Church.** 2000. RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat Biotechnol* **18**:1262-8.
162. **Shoemaker, D. D., E. E. Schadt, C. D. Armour, Y. D. He, P. Garrett-Engele, P. D. McDonagh, P. M. Loerch, A. Leonardson, P. Y. Lum, G. Cavet, L. F. Wu, S. J. Altschuler, S. Edwards, J. King, J. S. Tsang, G. Schimmack, J. M. Schelter, J. Koch, M. Ziman, M. J. Marton, B. Li, P. Cundiff, T. Ward, J. Castle, M. Krolewski, M. R. Meyer, M. Mao, J. Burchard, M. J. Kidd, H. Dai, J. W. Phillips, P. S. Linsley, R. Stoughton, S. Scherer, and M. S. Boguski.** 2001. Experimental annotation of the human genome using microarray technology. *Nature* **409**:922-7.
163. **Simon, R. J., R. S. Kania, R. N. Zuckermann, V. D. Huebner, D. A. Jewell, S. Banville, S. Ng, L. Wang, S. Rosenberg, C. K. Marlowe, and et al.** 1992.

- Peptoids: a modular approach to drug discovery. *Proc Natl Acad Sci U S A* **89**:9367-71.
164. **Singh-Gasson, S., R. D. Green, Y. Yue, C. Nelson, F. Blattner, M. R. Sussman, and F. Cerrina.** 1999. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat Biotechnol* **17**:974-8.
 165. **Smego, R. A., J. Frean, and H. J. Koornhof.** 1999. Yersiniosis I: microbiological and clinicoepidemiological aspects of plague and non-plague *Yersinia* infections. *Eur J Clin Microbiol Infect Dis* **18**:1-15.
 166. **Snyder, J. A., B. J. Haugen, E. L. Buckles, C. V. Lockatell, D. E. Johnson, M. S. Sonnenberg, R. A. Welch, and H. L. Mobley.** 2004. Transcriptome of uropathogenic *Escherichia coli* during urinary tract infection. *Infect Immun* **72**:6373-81.
 167. **Sykes, K. F., and S. A. Johnston.** 1999. Linear expression elements: a rapid, in vivo, method to screen for gene functions. *Nat Biotechnol* **17**:355-9.
 168. **Takahashi, M., K. Nukihara, and H. Mihara.** 2003. Construction of a protein-detection system using a loop peptide library with a fluorescence label. *Chem Biol* **10**:53-60.
 169. **Talaat, A. M., S. T. Howard, W. t. Hale, R. Lyons, H. Garner, and S. A. Johnston.** 2002. Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. *Nucleic Acids Res* **30**:e104.
 170. **Talaat, A. M., P. Hunter, and S. A. Johnston.** 2000. Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. *Nature Biotechnology*. **18**:679-82.
 171. **Talaat, A. M., R. Lyons, S. T. Howard, and S. A. Johnston.** 2004. The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. *Proc Natl Acad Sci U S A* **101**:4602-7.
 172. **Talaat, A. M., and K. Stemke-Hale.** 2005. Expression library immunization: a road map for discovery of vaccines against infectious diseases. *Infect Immun* **73**:7089-98.

173. **Tang, D. C., M. DeVit, and S. A. Johnston.** 1992. Genetic immunization is a simple method for eliciting an immune response. *Nature* **356**:152-4.
174. **Thomas, J. G., J. M. Olson, S. J. Tapscott, and L. P. Zhao.** 2001. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Res* **11**:1227-36.
175. **Titball, R. W., and E. D. Williamson.** 2004. *Yersinia pestis* (plague) vaccines. *Expert Opin Biol Ther* **4**:965-73.
176. **Tjaden, B., R. M. Saxena, S. Stolyar, D. R. Haynor, E. Kolker, and C. Rosenow.** 2002. Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res* **30**:3732-8.
177. **Tomaso, H., E. C. Reisinger, S. Al Dahouk, D. Frangoulidis, A. Rakin, O. Landt, and H. Neubauer.** 2003. Rapid detection of *Yersinia pestis* with multiplex real-time PCR assays using fluorescent hybridisation probes. *FEMS Immunol Med Microbiol* **38**:117-26.
178. **Usui, K., T. Ojima, M. Takahashi, K. Nokihara, and H. Mihara.** 2004. Peptide arrays with designed secondary structures for protein characterization using fluorescent fingerprint patterns. *Biopolymers* **76**:129-39.
179. **Valdivia, R. H., and S. Falkow.** 1997. Fluorescence-based isolation of bacterial genes expressed within host cells. *Science* **277**:2007-11.
180. **Van Gelder, R. N., M. E. von Zastrow, A. Yool, W. C. Dement, J. D. Barchas, and J. H. Eberwine.** 1990. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci U S A* **87**:1663-7.
181. **Wang, E., L. D. Miller, G. A. Ohnmacht, E. T. Liu, and F. M. Marincola.** 2000. High-fidelity mRNA amplification for gene profiling. *Nat Biotechnol* **18**:457-9.
182. **Wheeler, D. L., D. M. Church, R. Edgar, S. Federhen, W. Helmberg, T. L. Madden, J. U. Pontius, G. D. Schuler, L. M. Schriml, E. Sequeira, T. O. Suzek, T. A. Tatusova, and L. Wagner.** 2004. Database resources of the National Center for Biotechnology Information: update. *Nucleic Acids Res* **32 Database issue**:D35-40.

183. **Williamson, E. D., S. M. Eley, A. J. Stagg, M. Green, P. Russell, and R. W. Titball.** 2000. A single dose sub-unit vaccine protects against pneumonic plague. *Vaccine* **19**:566-71.
184. **Williamson, E. D., H. C. Flick-Smith, C. Lebutt, C. A. Rowland, S. M. Jones, E. L. Waters, R. J. Gwyther, J. Miller, P. J. Packer, and M. Irving.** 2005. Human immune response to a plague vaccine comprising recombinant F1 and V antigens. *Infect Immun* **73**:3598-608.
185. **Williamson, E. D., G. J. E. Sharp, S. M. Eley, P. M. Vesey, T. C. Pepper, R. W. Titball, and H. O. Alpar.** 1996. Local and systemic immune response to a microencapsulated sub-unit vaccine for plague. *Vaccine* **14**:1613-1619.
186. **Wolf-Watz, H., D. A. Portnoy, I. Bolin, and S. Falkow.** 1985. Transfer of the virulence plasmid of *Yersinia pestis* to *Yersinia pseudotuberculosis*. *Infect Immun* **48**:241-3.
187. **Worsham, P. L., M. P. Stein, and S. L. Welkos.** 1995. Construction of defined F1 negative mutants of virulent *Yersinia pestis*. *Contributions to Microbiology & Immunology* **13**:325-8.
188. **Xu, D., G. Li, L. Wu, J. Zhou, and Y. Xu.** 2002. PRIMEGENS: robust and efficient design of gene-specific probes for microarray analysis. *Bioinformatics* **18**:1432-7.
189. **Yamada, K., J. Lim, J. M. Dale, H. Chen, P. Shinn, C. J. Palm, A. M. Southwick, H. C. Wu, C. Kim, M. Nguyen, P. Pham, R. Cheuk, G. Karlin-Newmann, S. X. Liu, B. Lam, H. Sakano, T. Wu, G. Yu, M. Miranda, H. L. Quach, M. Tripp, C. H. Chang, J. M. Lee, M. Toriumi, M. M. Chan, C. C. Tang, C. S. Onodera, J. M. Deng, K. Akiyama, Y. Ansari, T. Arakawa, J. Banh, F. Banno, L. Bowser, S. Brooks, P. Carninci, Q. Chao, N. Choy, A. Enju, A. D. Goldsmith, M. Gurjal, N. F. Hansen, Y. Hayashizaki, C. Johnson-Hopson, V. W. Hsuan, K. Iida, M. Karnes, S. Khan, E. Koesema, J. Ishida, P. X. Jiang, T. Jones, J. Kawai, A. Kamiya, C. Meyers, M. Nakajima, M. Narusaka, M. Seki, T. Sakurai, M. Satou, R. Tamse, M. Vaysberg, E. K. Wallender, C. Wong, Y. Yamamura, S. Yuan, K. Shinozaki, R. W. Davis, A.**

- Theologis, and J. R. Ecker.** 2003. Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science* **302**:842-6.
190. **Yother, J., T. W. Chamness, and J. D. Goguen.** 1986. Temperature-controlled plasmid regulon associated with low calcium response in *Yersinia pestis*. *J Bacteriol* **165**:443-7.
191. **Zhang, J., and T. L. Madden.** 1997. PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. *Genome Res* **7**:649-56.
192. **Zhumabayeva, B., L. Diatchenko, A. Chenchik, and P. D. Siebert.** 2001. Use of SMART-generated cDNA for gene expression studies in multiple human tumors. *Biotechniques* **30**:158-63.
193. **Zuckermann, R. N., J. M. Kerr, S. B. H. Kent, and W. H. Moos.** 1992. **Efficient method for the preparation of peptoids [oligo(N-substituted glycines)] by submonomer solid-phase synthesis.** *Journal of the American Chemical Society.* **114**:10646-10647.

VITAE

Jonathan Neale Lawson was born in Lawton, Oklahoma, on January 15, 1976, the son of Carolyn Elizabeth and Douglas Brownlow Lawson. After graduating from Sandia High School in Albuquerque, New Mexico, in 1994 he attended Baylor University in Waco, Texas. During his undergraduate studies he conducted independent research developing and using a novel baculovirus detection screen in order to search for members of this virus family in hymenopteran species, specifically *Solenopsis invicta*. This work was done under the guidance of Robert Baldrige Ph.D. and lab space and advice provided by Christopher M. Kearney, Ph.D. In 1998 he graduated from Baylor University with a Bachelor of Science with a major in Biology and a minor in Biochemistry. After graduation he was employed at the University of Texas Southwestern Medical Center, where he worked as a laboratory technician and assistant to Orhan Kemal Oz, M.D., Ph.D.. In August of 2000 he entered the Graduate School of Biomedical Sciences at the University of Texas Southwestern Medical Center. In May of 2001 he married Elizabeth (Lisa) Louise Mitchell. Also in May of 2001 he joined the Molecular Microbiology graduate program where he initiated his thesis work in the Center for Biomedical Inventions under the guidance of Professor and Director Stephen Albert Johnston. In October of 2002 Jonathan qualified for doctoral candidacy within the Molecular Microbiology Department. Throughout his graduate studies Jonathan participated in the Student and Teachers Access to Resources at Southwestern (STARS) program where he mentored and taught high school students from the Dallas area. On August 25th of 2005 Jonathan and Lisa Lawson became the proud parents of Hannah Louise Lawson. Jonathan and his family plan on remaining in the Dallas area for a period after completion of his Doctoral degree.

Permanent Address: 3049 Modella Ave.
Dallas, TX 75229