BIOGRAPHICAL SKETCH

Cynthia J. Gibas

Assistant Professor, Department of Biology
Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0406
540.231.2393 (phone); 540.231.7126 (fax); cgibas@vt.edu
http://gaia.biotech.vt.edu

Research Interests

Comparative genomics; hybridization-based diagnostics; modeling and optimization of hybridization behavior on DNA arrays; macromolecular structure, electrostatics and diffusion-controlled behavior.

Skills

- Bioinformatics database and pipeline development for large-scale data analysis
- Molecular modeling of proteins and nucleic acids
- Application of bioinformatics methods (sequence analysis, structure analysis and simulation, expression array design and analysis)

Education

1996	University of Illinois at Urbana-Champaign	Ph.D.	Biophysics & Computational Biology Thesis: Computer Simulation of Titration Behavior in Proteins Advisor: Dr. Shankar Subramaniam
1990	Lawrence University	B.A	Chemistry

Awards and Honors

NIH NRSA Fellowship in Cell and Structural Biology, 1990-1993
DOE GAANN Fellowship in Computational Biology, 1994-1996

Professional Appointments

2000-	Assistant Professor, Department of Biology, Virginia Tech
1999-2000	Research Assistant Professor (Bioinformatics Program Development), Virginia Tech
1996-1999	Research Associate/Research Programmer, Computational Biology Group, National Center for Supercomputing Applications

Publications (including in review)

1. Sturgill D*, Ratushna V*, Ramamoorthy S, Reichow S, He Y, Lathigra R, Sriranganathan N, Halling S, Boyle S, **Gibas CJ**. Molecular targets for rapid identification of *Brucella* spp. 2004, in review.
2. **Gibas CJ**, Hilu KW, Berry MW, Stuart GW. A Phylogeny of Land Plants Based on Whole-Genome Analysis of Chloroplast Using Correlated Peptide Motifs. 2004, in review.
3. Ratushna, V*, Weller, J, **Gibas C**. Secondary structure in the target as a confounding factor in synthetic oligomer microarray design. 2004, in review.
4. Kaluszka, A*, **Gibas C**. Genome Organization Analysis Tool. 2004, Bioinformatics, in press.

5. **Gibas C**, Sturgill D*, Weller J. GenoMosaic: On-Demand Multiple Genome Comparison and Comparative Annotation. In: Proceedings of the 3rd IEEE Symposium on Bioinformatics and Bioengineering. 2003, IEEE Press.
6. **Gibas CJ**, Jambeck P, Subramaniam S. Continuum electrostatic methods applied to pH-dependent properties of antibody-antigen association. Methods. 20(3): 292-309, Mar 2000
7. Herrgard S, **Gibas CJ**, Subramaniam S. Role of an electrostatic network of residues in the enzymatic action of the Rhizomucor miehei lipase family. Biochemistry. 39(11): 2921-30, Mar 2000
8. **Gibas CJ**, Subramaniam S, McCammon JA, Braden BC, Poljak RJ. PH dependence of antibody/lysozyme complexation. Biochemistry. 36(50): 15599-614, Dec 1997
9. **Gibas C**, Subramaniam S. Knowledge-based design of a soluble bacteriorhodopsin. Protein Engineering. 10(10): 1175-90, Oct 1997
10. **Gibas CJ**, Subramaniam S. Explicit solvent models in protein pKa calculations. Biophysical Journal. 71(1): 138-47, Jul 1996
11. Blackwell M, **Gibas C**, Gygax S, Roman D, Wagner B. The plastoquinone diffusion coefficient in chloroplasts and its mechanistic implications. Biochimica et Biophysica Acta-Bioenergetics. 1183: 533-543, 1994
12. Nedbal L, **Gibas C**, Whitmarsh J. Light saturation curves show competence of the water splitting complex in inactive photosystem-II reaction centers. Photosynthesis Research. 30: 85-94, 1991

Books & Book chapters

13. Halling S, **Gibas C**, Boyle S. Comparative genomics of Brucella melitensis, B. suis, and B. abortus. In: Frontier in the Molecular and Cellular Biology of Brucella. Ignacio Lopez-Goni et al., eds. Horizon Scientific Press, 2004.
14. **Gibas C**, Jambeck P. Developing Bioinformatics Computer Skills. Cambridge, MA: O'Reilly & Assoc., 2001, 427 pp.

In preparation

15. **Gibas C**, Kaluszka A*, Hilu KW. Comparative analysis of chloroplast genomes: an online resource. 2004, in preparation.
16. Gharaibeh, R*, **Gibas C**. Performance comparison of common tools for oligonucleotide design. 2004, in preparation.
17. Sturgill D*, Gharabieh, R, **Gibas C**. GenoMosaic: integrating comparative and annotative genomic analysis. 2004, in preparation.
18. **Gibas C**, Weller J. Biophysical Optimization of Oligonucleotide Microarrays. 2004, in preparation.

Recent Invited Seminars & Colloquia

"New Tools for Comparative Functional Genomics", George Mason University, Apr. 2004.
"New Tools for Comparative Genomics", University of Arkansas, Sep. 2003.

Recent Conference Abstracts

1. Target secondary structure as a confounding factor in microarray experiments. Vladyslava G. Ratushna*, Jennifer W. Weller, **Cynthia J. Gibas**. (Poster at GSAC '04)
2. An integrated workflow for design of DNA microarrays using multiple criteria. Raad Gharaibeh, **Cynthia J. Gibas**. (Poster at GSAC '04)

3. A Phylogeny of Land Plants Based on Whole-Genome Analysis of Chloroplast Using Correlated Peptide Motifs. **Cynthia J. Gibas**, Khidir W. Hilu, Michael Berry, Gary Stuart. (Presentation at Genomes and Evolution '04)
4. GenoMosaic: On-Demand Multiple Genome Comparison and Comparative Annotation, David Sturgill*, **Cynthia Gibas**. (Poster at Genomes and Evolution '04)
5. An Interactive Web Resource for Comparative Analysis of Chloroplast Genomes. Aaron Kaluszka, Khidir Hilu, **Cynthia Gibas**. (Poster at Genomes and Evolution '04)
6. Incorporation of Target RNA Secondary Structure Parameter into Synthetic Oligomer Probe Design. Vladyslava G. Ratushna*, Jennifer W. Weller, **Cynthia J. Gibas**. (Poster at Research in Computational Molecular Biology (RECOMB) '04)
7. Identification of Transcribed Differentiating Genes in Brucella abortus, B.melitensis and B.suis. Vladyslava Ratushna*, David Sturgill*, Sheela Ramamoorthy, Sherry Poff, Nammalwar Sriranganathan, Stephen Boyle, **Cynthia Gibas**. (Poster at RECOMB '04)
8. Genome Organization Analysis Tool. Aaron Kaluszka, **Cynthia Gibas**. (Poster at RECOMB '04)

Completed Projects

Isolation of Genes for Transgenic Production of a Diabetes Treatment. Co-PI (With Glenda Gillaspy). $82,668, Commonwealth Health Research Board (6/01-12/02)

Pending Proposals

NSF, submitted 7/03, Optimization of Oligonucleotide Microarray Performance using Biophysical Properties (PI); 3 years, $1,042,944 (Recommended for funding)
NIH 1R01GM072619-01, 2/04, Biophysical Optimization of Oligonucleotide Microarrays (PI); 5 years; $2.3M (Scored; priority 153, percentile 9.7)
NSF, submitted 7/04, Biophysical Annotation of Nucleic Acids (PI); 3 years, $1,052,859 (no overlap with other proposals)

Teaching Experience

Bioinformatics Methods I (Molecular Sequence and Structure Analysis) (Fa2000-2004)
Bioinformatics Methods II (Genomics) (Sp2002, 2003)
The Physical Basis of Life (Laboratory section) (1993, 1994, University of Illinois)

Graduate and Undergraduate Training

1 M.S. graduate (David Sturgill – currently employed as a programmer at NIH/NIDDK)
Graduate research supervisor, 2 Ph.D. candidates (Raad Gharabieh, Vladyslava Ratushna)
Undergraduate research supervisor (Amanda Rudisin, Dominick DeDonno, Leah DiMascio, Shaun Rabah, Aaron Kaluszka (winner of 2004 Goldwater Scholarship), Stephanie Degen)

External and Departmental Service

Ad hoc reviewer, NIH (Biological Data Management, 2004; Major Research Equipment, 2001); NSF (Biological Databases and Informatics, 2003).
Conference Program Committee, O'Reilly Bioinformatics Technology Conference (2001,2002).
External Advisory Committee, Kentucky BRIN/INBRE program (2001-present).
Ad hoc reviewer, *Proteins*; IEEE Journals; *Biotechniques*; O'Reilly & Associates textbooks.
Virginia Tech Genetics, Bioinformatics, and Computational Biology graduate program development committee (2000-2003); Steering Committee (2003-present).
Instructor, *Advanced Uses of GenBank*, Fralin Biotechnology Center summer Biotechnology conference for educators, 2000-2002. Also workshops at Univ. of Alabama, Marshall University.

From: *Brucella*: Molecular and Cellular Biology. Edited by: Ignacio López-Goñi and Ignacio Moriyón

Page Proofs

Chapter 5

Comparative Genomics of *Brucella melitensis*, *B. suis*, and *B. abortus*

Shirley M. Halling*, Cynthia Gibas and Stephen Boyle

Abstract

The genomes of the classical *Brucella* species and their biovars have two chromosomes with the exception of *B. suis* biovar 3 strain 686 which has a single chromosome. The larger chromosome has approximately 2.1 Mbp and has a bacterial origin of replication. The smaller chromosome has approximately 1.2 Mbp and has plasmid replication functions. There is a large inversion within the small chromosome of *B. abortus* biovars 1, 2, and 4. There is a single large unique genetic island among the genomic sequences of *B. melitensis, B. suis*, and *B. abortus*. This island resides in the small genome of *B. suis* and encodes homologs of transfer functions and phage related genes. Given the high similarity among the genomic sequences of brucellae, differences among them with regards to host preference, virulence and infectious cycle could be due to subtle variations in the conserved DNA and differential expression of conserved genes, rather than due to unique genomic DNA fragments. Detailed comparative sequence analysis identified common and unique regions and diverged regions within conserved genes, and suggests sequence targets to be used in a comparative approach to functional genomics experiments.

1. Introduction

Brucellosis is a widespread disease of agriculturally important animals and is the most prevalent bacterial zoonosis. In the host animal, it often manifests itself by abortion. There are six classical species of *Brucella*, named for their host preference. Most of the infections caused by *Brucella* are due to four of the six classical *Brucella* species; *B. melitensis*, *B. abortus*, *B. suis* and *B. ovis*. These species of *Brucella* preferentially infect caprine, bovine, swine and ovine, respectively. *B. canis*, which is found primarily in dogs, is recognised as an emerging disease agent. *B. neotomae*, the sixth recognised classical species, was isolated from the desert rat. *Brucella* have recently been isolated from a wide variety of sea mammals (Foster *et al.,* 2002). Though the *Brucella* are closely related and proposed to constitute biovars of a single species in the bacterial genus *B. melitensis* (Verger *et al*., 1985), the classical names are still in use. There is an

*For correspondence email shalling@nadc.ars.usda.gov

ongoing discourse among taxonomists on what constitutes a species (Moreno *et al,,* 2002; see also Chapter 1). There is an added dimension to the discussion when the bacteria are intracellular pathogens like *Brucella* that have undergone selective pressure during host adaptation and in isolation from other bacteria.

Taxonomically, the genus *Brucella* is in the class Proteobacteria, subdivision α-2. This subdivision also includes rickettsiae, agrobacteria, and rhizobiae (Moreno *et al*. 1990). These bacteria, like brucellae, have a close association with either plant or animal cells. Some of the bacteria in this subdivision are obligate intracellular pathogens while others are facultative intracellular bacteria, either pathogens or symbionts. The structures of the genomes among the α-2 subdivision vary. There may be more than one chromosome and the presence of large plasmids.

Page Proofs

2. Sequence and Characteristics of the *Brucella* Genomes

The whole genomic sequences of *B. melitensis* (Acc. NC_00317 and NC_00318, DelVecchio *et al.,* 2002) and *B. suis* (Acc. NC_004310 and NC_004311, Paulsen *et al.,* 2002) have been determined. The draft sequence of *B. abortus* 9-941, a field strain isolated from naturally infected bovine, has been completed through a collaboration of National Animal Disease Center, Agricultural Research Service, United States Department of Agriculture and the University of Minnesota and is in draft form. TIGR in a collaboration with Virginia Tech, National Animal Disease Center, and Walter Reed Army Institute of Research, has nearly completed the whole genomic sequence of *B. ovis*. Of the four *Brucella* genomes that have or are being sequenced, only *B. ovis* is not pathogenic for man. As there are differences in host preference and pathogenicity of these four *Brucella* species, their genomic sequences provide a basis for designing experiments to determine the genetic basis of virulence, pathogenicity, evolution, and host-pathogen relationships.

Whole genomic sequence comparisons were made between *Brucella* and other α-2 Proteobacteria (Paulsen et al, 2002). The whole genomic sequence of brucellae is comprised of a large chromosome, designated Chr I, and a smaller chromosome, designated Chr II. Chr I from *B. suis* and *B. melitensis* shares broad gene synteny or gene order with the large chromosome of *Mesorhizobium loti* but only limited synteny with the genomic sequences of *Agrobacterium tumefaciens* and *Sinorhizobium meliloti*. Chr II has limited synteny to genomic sequences of *M. loti, A. tumefaciens* or *S. meliloti*. Approximately 70% of the ORFs of *Brucella* are shared with at least one of the following species, *M. loti, A. tumefaciens* or *S. meliloti* and approximately 55% are shared by all three. The closest bacterial species to *Brucella* are the *Ochrobactrum* spp. (Jumas-Bilak *et al*., 1998). These are free-living soil bacteria but are also opportunistic human pathogens.

The high degree of relatedness of the genomes of the classical *Brucella* species was first demonstrated over thirty-five years ago by DNA-DNA hybridisation (Hoyer and Mc Cullough, 1968 a, b). This observation was extended by more recent pulsed field gel electrophoresis (PFGE) studies. Genomes of the Office International Des Epizooties (OIE) type species for the *Brucella* classical species and their biovars digested with *Xba*I produced similar but distinct profiles (Allardet-Servant, 1991). Further PFGE studies, revealed that the genomes of *Brucella* except for one biovar of *B. suis* have two chromosomes (Michaux, 1993).

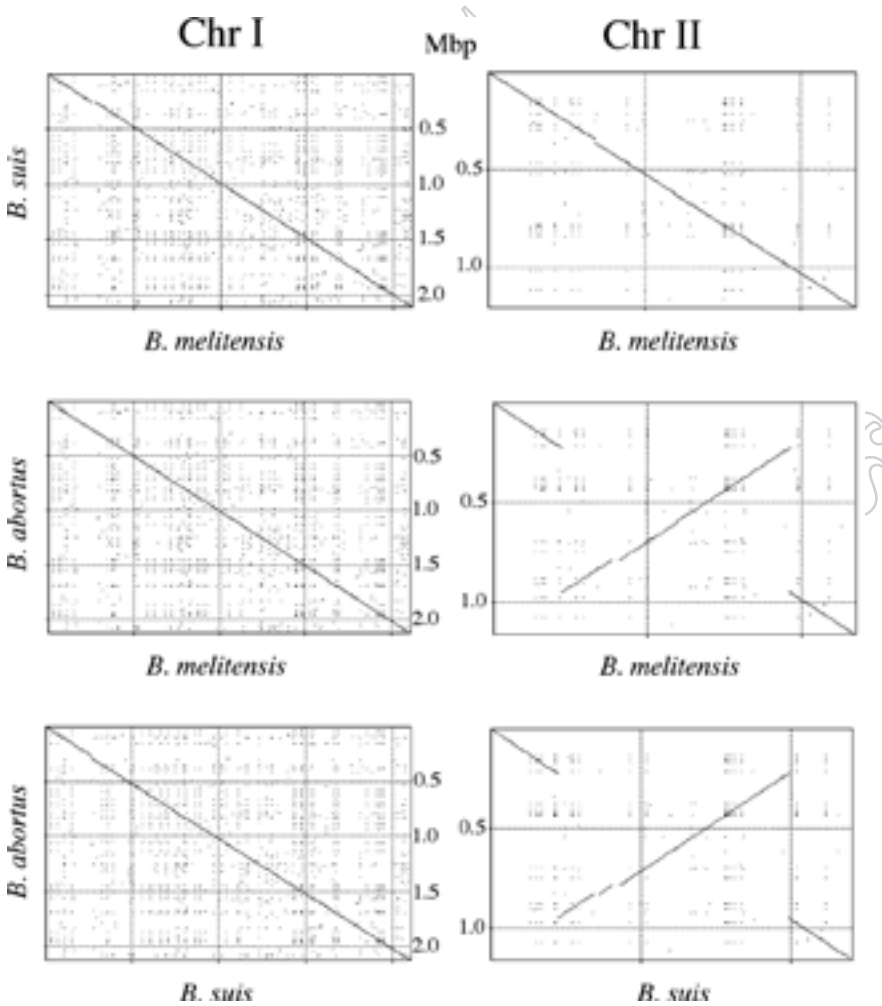

Figure 1. Pustell analyses (McVector 7.2, Accerlys Inc., San Diego, California, US) were carried out using the genomics sequences of *B. melitensis*, *B. suis*, and *B. abortus*. The chromosomes were oriented in the same direction and start at a common locus. The window size was 50. The minimum score was 90% and the hash value was 6. The number of base pairs is given in Megabasepairs (Mbp).

The degree of similarity among genomes of the *Brucella* is illustrated also by the number of single nucleotide polymorphisms (SNPs) found when genomic sequences of *Brucella* were compared. The number of SNPs between the genomic sequences of *B. suis* and *B. melitensis* was 7,301 over 3.2 Mbp (Paulsen *et al*., 2002). This is less than that for *Streptococcus pneumoniae* strains TIGR4 and R6 and *Escherichia coli* strains K-12 and O157:H7. The SNPs between strains of these two species were 8,303 over 1.9 Mbp and 36,676 over 3.9 Mbp, respectively.

The similarity of the genomic sequences and synteny among *B. melitensis*. *B. suis*, and *B. abortus* is evident by direct comparisons of their genomic sequences (Figure 1). There are more differences between the small chromosomes than the

Table 1. Unique, shared, and conserved genetic islands among *B. melitensis*, *B. suis*, and *B. abortus*

Size (bp)	Genetic Island (Loci)	[1]Chr	[2]S	[3]M	[4]A
18,290	BRA0362-BRA0379; tra genes and phage related genes	II	+	-	-
3,538	BR0588-BR0593. phage related	I	+	-	+
20,883	BMEI1674-BME1702; many small hypothetical ORFs, Phage integrase, Flj	I	-	+	+
25,245	BRA0419-BRA0439	II	+	+	-
3,952	BR1852-BR1854; Tn*2020*	I	+	-, +*	+

[1]Chromosome; [2]*B. suis*; [3]*B. melitensis;* [4]*B. abortus*

* Tn*2020* is not in the genomic sequence of *B. melitensis* 16 M but there is one copy of IS*2020* (Halling and Zuerner, 2002).

large. There are two large genetic islands (Table 1) among the genomic sequences of *B. suis* and *B. melitensis* that are not in Chr II of *B. abortus*. Given that Chr II has a plasmid-type replication origin; encodes homologs of two secretion systems (type IV secretion first described for DNA translocation and flagella); has a large putative composite transposon encoding amino acid, dipeptide, and sugar transport genes; contains at least two genetic islands encoding phage related proteins, it is reasonable to hypothesise that Chr II and many of its genes were horizontally acquired by *Brucella* (Paulsen *et al*., 2002; Tsolis, 2002). Likely, some of the genes on Chr II were acquired at the same time as the plasmid replication locus. This chromosome became essential after acquiring chromosomal genes.

Many features of the genome of the classical *Brucella* species were established before any of the whole genomic sequences was determined. The whole genomic sequences confirmed previous observations and provided details that furthered our understanding of other observations (DelVecchio *et al*. 2002; Paulsen *et al*., 2002). The base composition was calculated to be 56 to 58 % G + C from DNA hybridisation studies (Hoyer and McCullough, 1968a). This is in agreement with that determined from the whole genomic sequences of *B. suis* and *B. melitensis*, 57%. The PFGE maps established that the genomes of *Brucella* contained about 3.3 Mbp distributed on two large circular DNA chromosomes (Michaux *et al*., 1993) of approximately 2.1 Mbp and approximately 1.15 Mbp. The whole genomic sequences of both *B. suis* and *B. melitensis* are consistent with these sizes (DelVecchio *et al*., 2002; Paulsen *et al*., 2002).

The larger chromosome has a bacterial-like origin of replication while the smaller chromosome has a plasmid-like origin of replications. Small plasmids have not been isolated from *Brucella*. Attempts to isolate small plasmids from 600 strains of *Brucella* including all the classical species and their biovars using three different protocols at the University of California were not successful (Meyer, 1990). The

Page Proofs

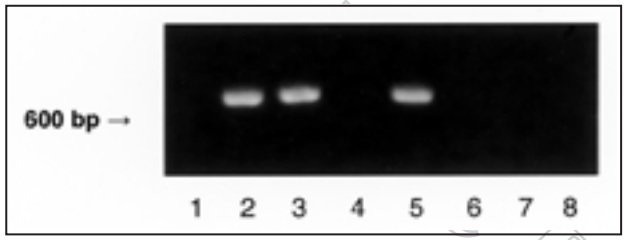

Figure 2. PCR products amplified using inversion specific primers. Lane 1: *B. suis* 1330; lane 2: *B. abortus* 544; lane 3: *B. abortus* biovar 2; lane 4: *B.* abortus biovar 3; lane 5: *B. abortus* biovar 4; lane 6: *B. abortus* biovar 5; lane 7: *B. abortus* biovar 6; lane 8: *B. abortus* biovar 9. The size marker is in base pairs (bp) on the left size of the figure. Forward primer was 5′-CCT-TTT-CCG-GAG-GCC-AAA-ATA-TGA-GCC-AT-3′ and reverse primer was 5′-CGC-CCG-ATA- TTT-CTC-TTC-ACT-TGA-CGC-CA-3′. Cells were heated to 95° C for 5 min. Melting, annealing and elongation temperatures and times were 95° C for 15 seconds, 60° C for 30 seconds, and 72° C for one and one-half minutes.

smaller, plasmid-based DNA molecule meets the criterion for a chromosome as it carries two tRNA synthetases and tRNA-Cys; these are essential genes. Most of the genes necessary for protein synthesis reside on the large chromosome while those encoding enzymes for sugar metabolism, protein regulators, and membrane transport proteins for sugars, dipeptides, and amino acids, reside on the small chromosome. Many of the transport protein coding sequences are within a 50 Kbp putative composite transposon, Tn*1953* (Bricker, Acc. No. AF454951).

There is a large inversion within the Chr II of *B. abortus* (Michaux-Charachon, 1997). The inversion was identified by PFGE maps prepared from genomic DNA digested with the restriction endonuclease *Pac*I. The inversion was identified in *B. abortus* 544, a biovar 1 strain and the type species for *B. abortus*. They also reported an inversion in Chr II of *B. abortus* biovars 2, 3 and 4. We designed primers based on the sequence flanking the inversion site of *B. abortus* 9-941 and confirmed the inversion in *B. abortus* biovars 1, 2, and 4 (Figure 2). We were unable to amplify genomic DNA from *B. abortus* biovar 3 Tuyla using these primers. This result was confirmed by use of primers that flank the genomic sequence of *B. suis* 1330 at the site where the inversion occurs relative to the genomic sequence of *B. abortus* 9-941. These primers amplified *B. suis* biovar 1 and *B. abortus* biovar 3 but not *B. abortus* biovar 1 strains 544 and 9-941, *B. abortus* biovar 2 and *B. abortus* biovar 4. The difference in results between the two laboratories may be due to different sources or strains of *B. abortus* biovar 3 used.

The relative small difference in the lengths of Chr I of *B. suis* and *B. melitensis* (1,207,381 bp and 1,177,787 bp, respectively) is mostly due to an unique genomic island in *B. suis* that appears to have been acquired by horizontal transfer. Note that here DNA fragments are referred to as genetic islands if they are at least 500 bp and encode phage-related ORFs, insertion sequences (IS), or fragments that either vary significantly in their G + C content from that of the genome or vary in codon usage from that of the genome statistics even though the G + C content may not vary. Only one unique island (Table 1) that was not transposon related was identified among the genomic sequences of *B. suis*, *B. melitensis* and *B. abortus* when the *B. abortus* draft sequence was compared with that of both *B. suis* and *B. melitensis*

and that genetic is in the genome of *B. suis*. This island is carried on Chr II and encodes bacterial transfer functions and phage related ORFs and hypothetical proteins. No homologs of virulence factors were confirmed on any of the islands. The report of the genomic sequence of *B. melitensis* was accompanied by an excellent review by Moreno and Moriyón (2002). They listed the factors shown previously to affect the virulence and pathogenicity of *Brucella* and identified new putative factors from the genomic sequence of *B. melitensis*. *Brucella* have none of the obvious virulence factors identified for many bacterial pathogens such as capsules, pili or fimbriae, proteases, exotoxins, phages and cytolysins.

Page Proofs

3. Genomic Diversity: Polymorphisms and Insertion Sequences

By the early 1990's, a number of genes had been cloned and sequenced from *Brucella*. The sequence data revealed that there was a high degree of similarity among the genomes, suggesting that it may be difficult to develop assays to distinguish the classical *Brucella* species and their biovars. Sequence diversity among the brucellae at the DNA level was observed among outer membrane proteins (Omps). Much of this work began with the observations that Omp profiles varied among brucellae isolates (Verstreate and Winter, 1984). Polymorphism observed among the Omps could be due to selection based on host, as the Omp are in contact with cells in the first phase of infection, adherence and attachment. The study of DNA based polymorphism at loci encoding Omps, *omp2a*/*omp2b* (Ficht *et al*., 1989; 1990), *omp31*, and *omp25* (Cloeckaert *et al*., 1995), lead to development of polymerase chain reaction-restriction length polymorphism (PCR-RFLP) assays that could identify *Brucella* and to varying extents distinguish the classical species and their biovars. This work is summarised in a recent review by Vizcaino and colleagues (Vizcaino, *et al*., 2000; see also Chapters 1 and 2).

Insertion sequences (IS) are a source of genomic instability and diversity by moving within the genome causing gene reassortment by chromosomal deletions and inversions (Mahillon *et al*., 1999). These events create polymorphism concomitantly. The insertion sequence, IS*711*/IS*6501* (Halling *et al*., 1993, Ouahurani *et al*., 1993) appears to be an ancestral sequence as not only do all the *Brucella* have the element but they also have common copies. It is a source of genetic diversity among the *Brucella* as it has transposed to at least one unique locus in the classical species of *Brucella*. So far, it has not been shown to be involved in genomic inversions and large deletions or rearrangements.

Insertion sequences can benefit bacteria by transposing and modifying gene expression, aiding in their adaptation to a new environment such as a different host. Given that *Brucella* has little exposure to other bacteria, IS may be an important internal source of diversity especially during selection. The copy number of IS*711* ranges in *Brucella* from as few as seven to well over 30 in the marine isolates. Whether or not the other IS from *Brucella* are able to transpose is not known. Transposition of transposable elements is usually tightly regulated and, if the elements arose in *Brucella* from horizontal transfer, their promoters may not be recognised or functional in *Brucella*. Several of the elements in *B. suis* have pseudogenes and these would not be predicted to be mobile unless complemented by a second copy elsewhere in the genome.

The movement of IS alters gene linkages within the genome. This results in polymorphism that can be used as targets to identify and distinguish strains (see Chapter 1). Though the classical species have copies of IS*711* that are found at the same loci, they also have at least one IS*711* copy at a unique loci so that the classical species can be distinguished on Southern blots when they are probed with IS*711* (Ouahrani *et al*., 1993; Bricker and Halling, 1994). Other assays based on IS*711* have been developed and continue to be developed (Bricker, 2002).

Though IS*711* elements transpose in *Brucella*, they do not appear to transpose at a high frequency (Halling and Zehr, 1990). This element in *B. ovis* isolates from different time periods and geographical locations did not appear to differ in copy numbers, even though there are approximately 30 copies. The element has been observed to be mobile in *B. abortus*. The *B. abortus* biovar 1 strains 2308 and RB51 (Schurig *et al*., 1991) have more copies of IS*711* than the *B. abortus* type species, strain 544 (Bricker and Halling, 1995; Vemulapalli *et al*., 1999). In *B. abortus* S2308, there are two tandem direct copies of IS*711*. During selection of the rough vaccine strain RB51 (Schurig, *et al*., 1991), IS*711* mobilised again (Vemulapalli *et al*., 1999). This copy interrupts the *wbo*A gene encoding a glycosyltransferase that is essential for the synthesis of O-side chain. These copies of IS*711* made it possible to develop PCR assays to identify and distinguish the challenge and vaccine strains, *B. abortus* S2308 and RB51. Though there are other insertion sequences in *Brucella*, there are not any reports of their transposition yet.

Genetic diversity by IS may be directed to a certain extent as most IS do not transpose randomly but, rather, transpose to target sequences. What other factors may be involved in internal transposition of an element such as structural changes caused by palindromes or other sequences is not known. The target sequence for IS*711* is pyrimidine-A-T-purine, often CATG, but IS*711* elements do not randomly transpose to those sequences either. Several of the elements from *B. ovis* were found to flank a repeated DNA sequence, Bru-RS1 (Halling and Bricker, 1994). There did not seem to be a pattern with regards to the relative orientation of IS*711* and Bru-RS1. The Bru-RS1 elements might have an effect on DNA secondary structure that affects the efficiency of either the transposition process or the exposure and recognition of target sites. Thus, Bru-RS1 may be a source of genetic instability and diversity for *Brucella* working hand and hand with IS*711*. As *B. ovis* has more than 20 copies of IS*711,* the genomic sequence of *B. ovis* will further our understanding of the association of IS*711* and Bru-RS1. Also, it will be of interest to note if phenotypic traits of *B. ovis* can be traced to transposition of IS*711* in *B. ovis*. These data may be useful in generating hypotheses regarding both transposition targets of IS*711*, the most mobile transposable element in brucellae, and their effect on genomic structure.

Repeat sequences can generate polymorphism during replication due to mispair slippage. Slippage backwards or forwards can result in deletions or duplications of sequences, respectively. If the repeats reside in or near genes, gene expression can be affected and have been associated with genetic diseases in humans (Cummings and Zoghbi, 2000). Similar or identical tandem repeats of variable lengths are often found at several loci in a genome. These are designated as VNTRs for variable nucleotide tandem repeats. VNTRs have been used for strain typing for epidemiological trace back for bacterial pathogens, such as *Bacillus*

anthracis (Keim and Smith, 2002) and *Mycobacterium tuberculosis* (Frothingham and Meeker-O'Connell, 1998). A VNTR was identified in *Brucella*, AGGGCAGT (Bricker *et al*., 2003). Most of the loci that have VNTRs in *Brucella* have ORFs annotated upstream of them but the effect of the VNTR on expression of the ORFs is not known. Interestingly, most of these VNTRs in *Brucella* are physically linked to a conserved region that may influence the frequency of slippage.

Page Proofs

4. Genomics and Taxonomy

Moreno and Moriyón hypothesized that *B. melitensis* and *B. abortus* share a common ancestor and became isolated at the same time that their hosts did, about 20 million years ago (Moreno and Moriyón, 2001). The ability of *B. melitensis* to cause abortions in cattle, sheep and goats may be due to its having a common ancestor with *B. abortus*. *B. suis* can cause abortions in swine but not cattle. A number of studies support that *B. abortus* and *B. melitensis* are more closely related to each other than either one is to *B. suis*. This relative closeness of the genomes of *B. melitensis* and *B. abortus* is demonstrated by dendrograms constructed from data generated from the classical species of brucellae and their biovars by data obtained using: arbitrarily primed PCR (Fekete *et al*. 1992); random amplified polymorphic DNA (Tcherneva *et al*. 2000); PFGE (Michaux-Charachon *et al*., 1997, Jumas-Bilak *et al*., 1998); Western blot protein patterns including whole cell antigens and soluble antigens reacted with sera from rabbits hyperimmunised with either *Ochrobactrum anthropi* or *B. suis* 1330 (Velasco *et al*., 1998), transcript analyses (see below) and metabolic capabilities (Meyer, 1990).

Comparison of genomic sequences entails not only the DNA sequence itself but their ORFs as well. Confidence in an annotated genomic sequence is higher for those ORFs for which conserved homologs are identified in protein databases searches. Even though sequences of *B. suis* and *B. melitensis* are highly similar they have been annotated differently in some instances. In some cases, ORFs are on opposite strands or overlap. Annotation of the *Brucella* genomes contains a number of hypothetical proteins and conserved hypothetical proteins. The function of some of the hypothetical proteins will likely be discovered with further experimental work on *Brucella* and other bacteria. In time, it is likely that it will be determined that some of these putative ORFs do not encode gene products. In any case in order to have an inclusive microarray, it is better to saturate the genome with ORFs and use data from other studies including microarray studies to identify which ORFs encode gene products. Proteomics can aid in the identification of start codons and expressed ORFs.

One of the most difficult tasks in annotation of any genome is to identify which short ORFs express polypeptides. There are many short ORFs annotated for the *B. suis* genome but very few for the *B. melitensis* genome. Mostly, selection of ORFs when there are no homologs in the databases is done by use of software such as Glimmer (Salzberg *et al*., 1998; Delcher *et al*., 1999). This software identifies ORFs based on using highly conserved homologous regions within the target genomic sequence. This approach works best to identify genes in regions of the chromosome where the G + C content matches the statistical G + C content for the whole genome. The ability to detect ORFs that are expressed, especially small ORFs, may be improved by studies to identify promoters from *Brucella*.

Another difficulty in the annotation of genomic sequences of *Brucella* is that the α-Proteobacteria may use a start codon other than ATG. Annotation is an ongoing process and data from microarrays, RT-PCR, and transcript mapping will be needed to identify all the genes and their ORFs. These data will be needed to determine the genetic basis for differences in pathogenicity, virulence, and infectious cycle of the *Brucella*.

Various kinds of repeated sequences have been reported among bacteria. The accumulation, distribution and function of these elements are still largely speculative. They may affect gene expression indirectly by stabilising m-RNA (Hulton *et al*., 1991). In enterobacteriae, two such sequences, designated Enterobacterial Repetitive Intergenic Consensus (ERIC) and Repetitive Extragenic Palindromic (REP) sequences, have been described (Stern *et al*., 1984; Hulton *et al*., 1991). Two sequences designated Bru-RS1 and Bru-RS2 that are similar to ERIC sequences in size and occurrence in non-coding regions but not in DNA sequence were discovered in brucellae (Halling and Bricker, 1994). The Bru-RS sequences like ERIC sequences occur singly. A comparison of the genomic sequences supports that the elements have not been involved in genomic rearrangements. The two Bru-RS elements are 65% similar and are flanked by inverted repeats. Within the element, there is another copy of the repeat. This results in a 70 bp subsequence in Bru-RS1 that is bounded by 8 bp direct repeats and a 38 bp sequence bounded by 8 bp inverted repeats. Some of the Bru-RS1 elements are truncated, having only the 70 bp subsequence. While the Bru-RS2 elements have a similar structure, their inverted repeats are shorter, 6 bp. However, 17 bp of the left end of Bru-RS2 is directly repeated within the element to bound an 81 bp subsequence. The internal direct repeat pairs with the left end to form a 40 bp element. Both Bru-RS1 and Bru-RS2 have several repeats of the sequence 5′-GAAA-3′, a sequence shown to stabilise RNA hairpins (Heus and Pardi, 1991). As discussed above Bru-RS1 sequences are hot spots for IS*711* insertion. This may be due to the highly palindromic nature of the elements that could disrupt DNA structure. The sequences of the elements are not identical and appear to have drifted (Halling and Bricker, 1994).

5. Genomics and Diagnostics

Relatively small differences in the predicted gene content of *Brucella* are unlikely to completely explain differences in virulence and host preference. However, these minor gene content differences are sufficient to uniquely identify each of the species in an expression experiment.

In a genomic sequence comparison, we identified a total of 101 genes that are uniquely present, or uniquely absent, in one of the three *Brucella* genomes (Figure 3). Gene boundaries were defined based on existing annotation for *B. suis* and *B. melitensis*, and predicted on a draft sequence of *B. abortus* using Glimmer 2.0 (Delcher *et al*., 1999). Complete nucleotide sequences were compared directly using an early prototype of the GenoMosaic system (Gibas *et al*., 2003), and differentiating regions were identified and localised within gene boundaries. RT-PCR was used to test for expression of these differentiating genes *in vitro*, and 54 of the genes identified produced transcripts under the conditions of our experiment (Sturgill *et al*., unpublished).

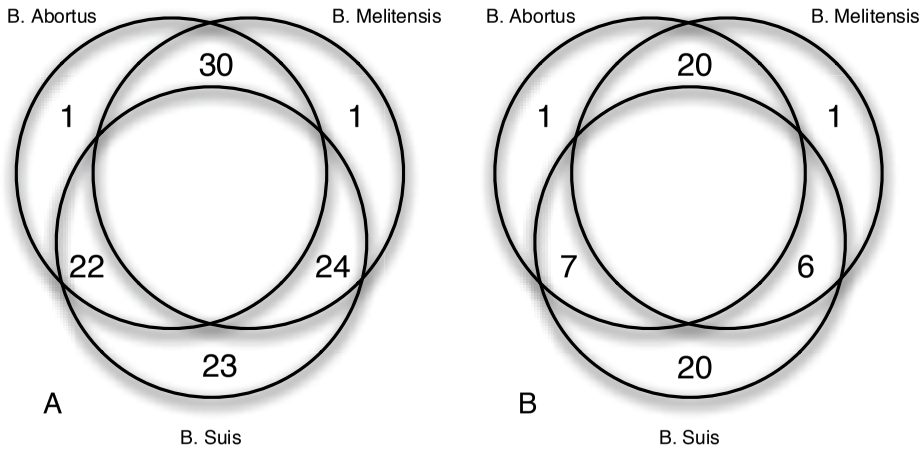

Figure 3. Distribution of unique and common genes in three *Brucella* genomes as A) predicted by sequence comparison and detected in the genomic DNA by PCR, B) detected by transcripts by RT-PCR.

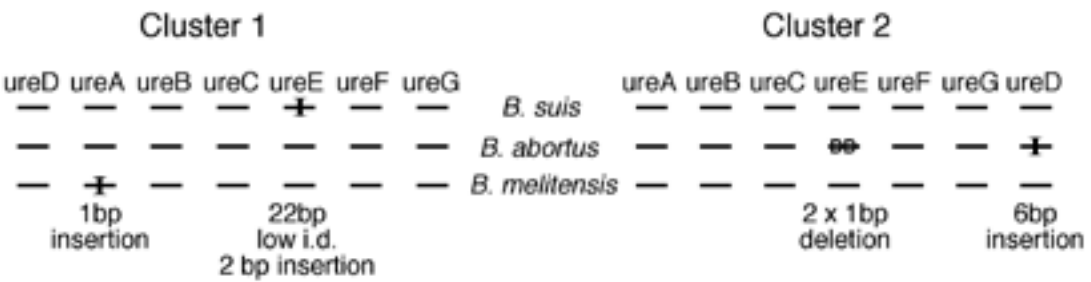

Figure 4. Schematic of urease 1 and 2 clusters within *Brucella* spp. genomes.
I represent insertion(s) and D represents deletion(s) within indicated urease subunit gene.

Image Proofs

Genes with homologs in all three species were surveyed to identify additional differentiating features comprising less than a complete gene. Of 3,163 genes with homologs in all three genomes, nearly all were found to be more than 95% identical in sequence to their homologs at the nucleotide level, with 95% or greater coverage of the sequence. In every case where homologs were present in all three genomes, high scoring sequence match regions among them had sequence identity greater than 90%. Therefore, the only additional group of genes that could be considered as potential differentiating features in an expression context was a group of 214 genes where the sequence match detected extended over less than 80% of the sequence, leaving a distinguishing region as a target for a unique primer or probe. These partial homologs may be useful as additional differentiating features but have not yet been tested in expression experiments (Sturgill, 2003).

Page Proofs

6. Differential Gene Expression

The small number of truly unique coding sequences in the *Brucella* genomes suggests that differences in virulence and host preference are most likely a function of differential gene expression as well as unique gene content. Very few comparative studies of differential gene expression in *Brucella* exist, but recent results of a comparative study of urease gene expression (S. Boyle, unpublished) suggest that a differential approach to expression will yield interesting results.

There are two separate urease operons on Chr I of the three *Brucella* spp. genomes located at approximately 0,.273 and 1.316 Mb. Species-specific differences are present in the two gene clusters of urease subunits present in *B. suis*, *B. abortus*, and *B. melitensis* (*ure*A-G-1 BR0267-BR0273 and *ure*A-G-2 BR1356-BR1362 in *B. suis*). In the *ure*A-1 gene (BR0268) of *B. melitensis,* there is a 1 bp insertion representing a potential frameshift. In the *ure*D-2 (BR1362) gene of *B. abortus*, a 6 bp insertion was identified. In the *ure*E-2 gene (BR1359) of *B. abortus,* two separate single base deletions are present and representing possible frame shift in translation. Finally, the last 22 bp of *ure*E-1 (BR0271) were shown to be 100% identical in *B. abortus* and *B. melitensis* but significantly diverged in *B. suis*, including a 2 bp deletion. This variation within these urease gene clusters could prove to be significant to the expression of functional urease subunits (Figure 4).

Using primers specifically designed to amplify mRNA transcripts from the three species of *Brucella* grown in trypticase soy broth, we detected transcripts of a predicted size from all the subunit genes in both clusters except for *ure*D-1 and a smaller than expected transcript from ureC-2. All 3 *Brucella* spp. exhibited urease activity when tested in urea broth (Difco). However, when a kanamycin resistant marker was introduced and most of the *ure*A-1 and *ure*B-1 genes were deleted by recombination, *B. suis* did not produce urease activity. In contrast, when a kanamycin resistant marker was introduced and most of the *ure*B-2 and *ure*C-2 genes were deleted, *B. suis* retained urease activity. These results suggest that there is likely some type of post-translational complementation occurring among the subunits of the two urease gene clusters. A similar observation was reported (Sangari and Garcia-Lobo, 2003) when the *ure*C cluster was mutated by transposon insertion, only *ure*C-1 but not *ure*C-2 mutants exhibited loss of urease

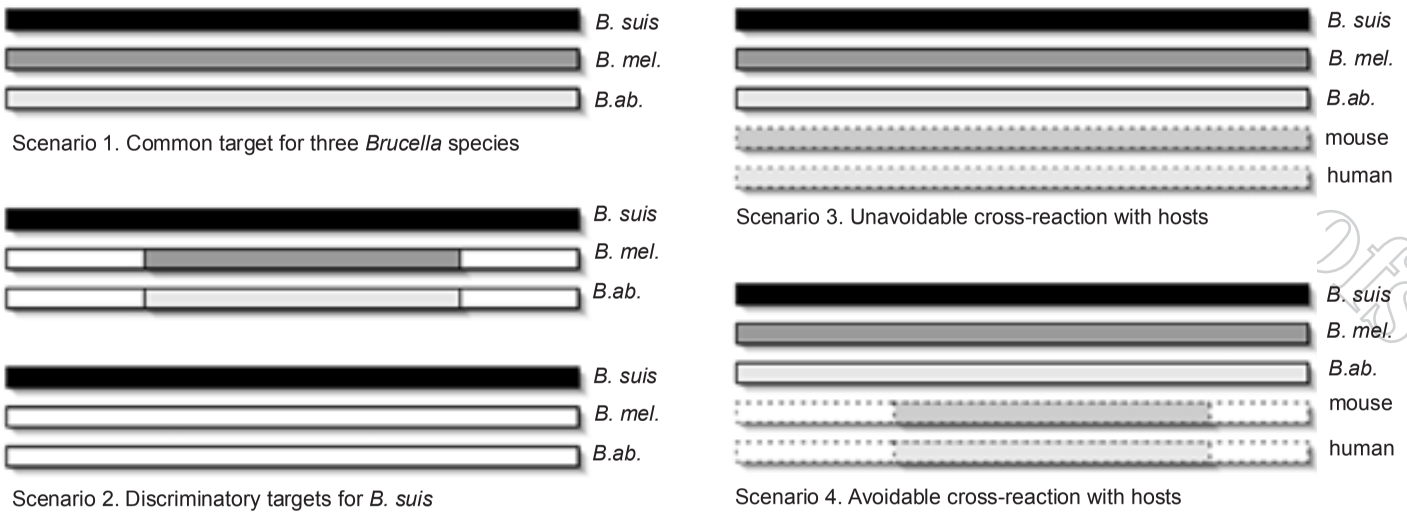

Figure 5. Some probe design scenarios for a multiple-genome array design problem. In scenario 1, a gene homolog exists in each of the three species with high identity along the full length of the gene. Scenario 2 describes a situation in which a differentiating probe can be made; the gene homolog in one or more of the genomes is missing, partially or completely. In scenario 3, this extensive identity includes genes in the host genomes and cross-reaction is unavoidable. In scenario 4, cross-reaction with the host can be avoided where identity to host genes is low.

activity in all three *Brucella* species. Certainly more experimental work needs to be performed in order to clarify the nature of the expression of the two urease operons of the *Brucella*.

7. Designing Arrays that Address Closely Related Sequences

DNA microarrays are widely used to study global transcription levels. A comparative expression study of the *Brucella* presents a unique challenge and therefore a unique opportunity in DNA array design. The extensive identity between gene homologs in the three *Brucella* species suggests a comparative genomics approach to construction of a DNA microarray. Rather than using separate arrays, or separate probes designed for each gene in each species, it would be possible to construct a generic *Brucella* array that would include probes designed for sequence regions common to all three genomes. Probes for differentiating genes or regions could also be added, to serve as a built-in diagnostic.

In fact, constructing such a design presents a significant challenge. Currently available oligonucleotide array design software packages have no capabilities for optimising probe picks to fall within highly similar regions of related targets. Complicating the issue further is the need to avoid regions of the target sequence that are highly similar to host sequences, in order to avoid cross-reaction with host transcripts in experiments where *Brucella* samples are isolated from host cells, as shown in Figure 5.

Even in sequences greater than 90% identical, it is difficult to avoid mismatches, and while the effects of single base pair mismatches on hybridisation is well-studied for oligonucleotides 25-30 nt in length (Lipshutz *et al*., 1999; Ramakrishnan *et al*., 2002; Peterson *et al*., 2002; Riccelli *et al*., 2003), there has been little corresponding work for longer oligonucleotides or for cDNA arrays. However, longer oligonucleotides have been shown to provide optimal sensitivity and specificity (Kane *et al*., 2000; Hughes *et al*., 2001). In the absence of quantitative study of the thermodynamic effects of slight mismatches in long oligonucleotides and development of a modelling approach to compensate for these effects in a comparative study, mismatches should be avoided when possible. A truly "generic" *Brucella* array for comparison among sequenced species would be limited to probes for regions that are 100% identical in all three species.

In practice, we were limited to designing an array using existing oligonucleotide selection methodologies pick70 (Bozdech *et al*., 2003) and OligoArray (Rouillard *et al*., 2002; 2003). These programs select probes based on commonly used design criteria, among them sequence uniqueness, low self-complementarity, and specific GC content or sequence complexity. OligoArray adds biophysical criteria which affect hybridisation – duplex T_m and T_m of secondary structures predicted to form within the probe – but at the present time these criteria are applied in a rather rudimentary fashion and information needed to predict hybridisation behaviour is not available.

Neither of these design methodologies addresses the problem of design for multiple transcriptomes, so it was not possible to eliminate mismatch entirely in the first round of oligonucleotide selection. *B. suis* was chosen as the base sequence for design of the array, and probes from specific differentiating targets to address

unique regions in *B. melitensis* and *B. abortus* were subsequently added. The prototype array designed using these methods was analysed and found to contain approximately 10% of probes that contained unusually stable secondary structures, and approximately 10% of probes that did not match equally well with targets in each of the three genomes (Ratushna and Gibas, unpublished).

Because we can identify precisely where secondary structure and mismatch effects may interfere with hybridisation on this prototype array, the experiment becomes a valuable exploration not only of comparative *Brucella* functional genomics, but of our ability to design, optimise and interpret data gathered from a multi-transcriptome experiment. Multi-transcriptome approaches are likely to become more common as rapid genome sequencing allows microbiologists to take the approach of sequencing families of closely related microbial genomes to identify the basis of virulence and other phenotypic differences (Bhattacharyya *et al*., 2002a; 2002b).

The differences in host specificity and virulence among the brucellae can be more easily investigated now that genomic sequences are available. The similarity among the genomic sequences suggests that those differences may be due in part to sequence variations such as SNPs and pseudogenes. The RT-PCR studies support that differential gene expression is also likely to play a part as well. Differences in gene expression will also be applicable to rapid identification of agents of brucellosis. Clearly the genomic sequence is not the end but is another beginning.

Page Proofs

References

Allardet-Servent, A., Carles-Nurit, M.J., Bourg, G., Michaux, S., and Ramuz, M. 1991. Physical map of the *Brucella melitensis* 16 M chromosome. J. Bacteriol. 173: 2219-2224.

Alton, G.G., Jones, L.M., and Pietz, D.E. 1975. Laboratory techniques in brucellosis. In: Monograph series World Health Organization, Geneva, Switzerland.

Bhattacharyya, A., Stilwagen, S., Ivanova, N., D'Souza, M., Bernal, A., Lykidis, A., Kapatral, V., Anderson, I., Larsen, N., Los, T., Reznik, G., Selkov, E. Jr., Walunas, T.L., Feil, H., Feil, W.S., Purcell, A., Lassez, J.L., Hawkins. T.L., Haselkorn, R., Overbeek, R., Predki, P.F., and Kyrpides, N.C. 2002a. Whole-genome comparative analysis of three phytopathogenic *Xylella fastidiosa* strains. Proc. Natl. Acad. Sci. U S A. 99: 2403-2408.

Bhattacharyya, A., Stilwagen, S., Reznik, G., Feil, H., Feil, W.S., Anderson, I., Bernal, A., D'Souza, M., Ivanova, N., Kapatral, V., Larsen, N., Los, T., Lykidis, A., Selkov, E. Jr., Walunas, T.L., Purcell, A., Edwards, R.A., Hawkins, T., Haselkorn, R., Overbeek, R., Kyrpides, N.C., and Predki, P.F. 2002b. Draft sequencing and comparative genomics of *Xylella fastidiosa* strains reveal novel biological insights. Genome Res. 12:1556-63.

Bricker, B.J. 2002. PCR as a diagnostic tool for brucellosis. Vet. Microbiol. 90: 435-446.

Bricker, B.J., Ewalt, D.R., and Halling, S.M. 2003. *Brucella* 'HOOF-Prints': strain typing by multi-locus analysis of variable number tandem repeats (VNTRs). BMC Microbiol. 3:15.

Bozdech, Z., Zhu, J., Joachimiak, M.P., Cohen, F.E., Pulliam, B., and DeRisi, J.L. 2003. Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. Genome Biol. 4: R9.

Bricker, B.J., Ewalt, D.R., MacMillan, A.P., Foster, G., and Brew, S. 2000. Molecular characterization of *Brucella* strains isolated from marine mammals. J. Clin. Microbiol. 38:1258-1262.

Bricker, B.J., and Halling, S.M. 1994. Differentiation of *Brucella abortus* bv. 1, 2, and 4, *Brucella melitensis*, *Brucella ovis*, and *Brucella suis* bv. 1 by PCR. J. Clin. Microbiol. 32: 2660-2666.

Bricker, B.J., and Halling, S.M. 1995. Enhancement of the *Brucella* AMOS PCR assay for differentiation of *Brucella abortus* vaccine strains S19 and RB51. J. Clin. Microbiol. 33:1640-1642.

Cloeckaert, A., Grayon, M., and Grepinet, O. 2000. IS*711* element downstream of the bp26 gene is a specific marker of *Brucella* spp. isolated from marine mammals. Clin. Diagn. Lab. Immunol. 7: 835-839.

Cloeckaert, A., Verger, J.M., Grayon, M. , Grépinet, O. 1995. Restriction site polymorphism of the genes encoding the major 25 kDa and 36 kDa outer membrane proteins of *Brucella*. Microbiol. 141: 2111–2121.

Cloeckaert, A., Verger, J.M., Grayon, M., and Vizcaino, N. 1996. Molecular and immunological characterization of the major outer membrane proteins of *Brucella*. FEMS Microbiol. Lett. 145:1-8.

Cummings, C.J., and Zoghbi, H.Y. 2000. Trinucleotide repeats: mechanisms and pathophysiology. Annu. Rev. Genomics: Hum. Genet. 1: 281-328.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27: 4636-4641.

DelVecchio, V.G., Kapatral, V., Redkar, R.J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goltsman, E., Selkov, E., Elzer, P.H., Hagius, S., O'Callaghan, D., Letesson, J.J., Haselkorn, R., Kyrpides, N., and Overbeek, R. 2002. The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. Proc. Natl. Acad. Sci. U S A. 99: 443-448.

Fekete, A., Bantle, J.A., Halling, S.M., and Stich, R.W. 1992. Amplification fragment length polymorphism in *Brucella* strains by use of polymerase chain reaction with arbitrary primers. J Bacteriol. 174: 7778-7783.

Ficht, T.A., Bearden, S.W., Sowa, B.A. and Marquis, H. 1990. Genetic variation of the *omp2* porin locus of the *Brucella*: species-specific markers. Mol. Microbiol. 4: 1135–1142.

Ficht, T.A., Bearden, S.W., Sowa, B.A., Adams, L.G. 1989. DNA sequence and expression of the 36-kilodalton outer membrane protein gene of *Brucella abortus*. Infect. Immun. 57: 3281-3291.

Foster, G., MacMillan, A.P., Godfroid, J., Howie, F., Ross, H.M., Cloeckaert, A., Reid, R.J., Brew, S., and Patterson, I.A. 2002. A review of *Brucella sp*. infection of sea mammals with particular emphasis on isolates from Scotland. Vet. Microbiol. 90: 563-580.

Please Inquire

Frothingham, F., Meeker-O'Connell, W.A. 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiol. 144:1189-1196.

Gibas, C., Sturgill, D., and Weller, J. 2003. GenoMosaic*:* On-Demand Multiple Genome Comparison and Comparative Annotation. In: Third IEEE Symposium on Bioinformatics and Bioengineering. Bethesda, MD: IEEE Computer Society.

Halling, S.M., and Bricker, B.J. 1994. Characterization and occurrence of two repeated palindromic DNA elements of *Brucella* spp.*:* Bru-RS1 and Bru-RS2. Mol Microbiol. 14: 681-689.

Halling, S.M., Tatum, F.M., and Bricker, B.J. 1993. Sequence and characterization of an insertion sequence, IS*711*, from *Brucella ovis*. Gene. 133: 123-127.

Halling, S.M., and Zehr, E.S. 1990. Polymorphism in *Brucella* spp. due to highly repeated DNA. J. Bacteriol. 172: 6637-6640.

Halling, S.M., and Zuerner, R.L. 2002. Evidence for lateral transfer to *Brucella*: characterization of a locus with a Tn-like element (Tn*2020*). Biochim. Biophys. Acta. 1574: 109-116.

Heus, H., and Pardi, A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. Science. 253: 191-194.

Hoyer, B.H., and McCullough, N.B. 1968a. Polynucleotide homologies of *Brucella* deoxyribonucleic acids. J. Bacteriol. 95: 444–448.

Hoyer, B.H., and McCullough. N.B. 1968b. Homologies of deoxyribonucleic acids from *Brucella ovis*, canine abortion organisms and other *Brucella* species. J. Bacteriol. 96: 1783–1790.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., Kobayashi, S., Davis, C., Dai, H., He, Y.D., Stephaniants, S.B., Cavet, G., Walker, W.L., West, A., Coffey, E., Shoemaker, D.D., Stoughton, R., Blanchard, A.P., Friend, S.H., and Linsley, P. 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat. Biotechnol.19: 342-347.

Hulton, C.S., Higgins, C.F., and Sharp, P.M. 1991. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. Mol. Microbiol. 5: 825-834.

Jumas-Bilak, E., Michaux-Charachon, S., Bourg, G., Ramuz, M., and Allardet-Servent, A. 1998. Unconventional genomic organization in the alpha-subgroup of the *Proteobacteria*. J Bacteriol. 180: 2749–2755.

Kane, M.D., Jatkoe, T.A., Stumpf, C.R., Lu, J., Thomas, J.D., and Madore, S.J. 2000. Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. Nucleic Acids Res. 28: 4552-4557.

Keim, P., and Smith, K.L. *Bacillus anthracis* evolution and epidemiology. 2002. Curr. Top. Microbiol. Immunol. 71: 21-32.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. 1999. High density synthetic oligonucleotide arrays. Nat. Genet. 21:20-24.

Mahillon, J., Leonard, C., and Chandler, M. IS elements as constitutents of bacterial genomes. Res. Microbiol. 150: 675-687.

Meyer, M.E. 1990. Evolutionary development and taxonomy of the genus *Brucella*. In: L. Garry Adams (Ed.) Advances in Brucellosis Research, Texas A & M University Press, College Station. pp. 12-35.

Page Proof

Michaux, S., Paillisson, J., Carles-Nurit, M.J., Bourg, G., Allardet-Servent, A., and Ramuz, M. 1993. Presence of two independent chromosomes in the *Brucella melitensis* 16M genome. J. Bacteriol. 175:701-705.

Michaux-Charachon, S., Bourg, G., Jumas-Bilak, E., Guigue-Talet, P., Allardet-Servent, A., O'Callaghan, D., and Ramuz, M. 1997. Genome structure and phylogeny in the genus *Brucella*. J. Bacteriol. 179: 3244-3249.

Moreno, E., Cloeckaert, A., and Moriyón, I. 2002. *Brucella* evolution and taxonomy. Vet. Microbiol. 90: 209-227.

Moreno, E., and Moriyón, I. 2001. Genus *Brucella*. In: Dworkin (Ed.), The Procaryotes: An evolving microbiological resource for the microbiological community. Springer, New York.

Moreno, E., Stackebrandt, E., Dorsch, M., Wolters, J., Busch, M., and Mayer H. 1990. *Brucella abortus* 16S rRNA and lipid A reveal a phylogenetic relationship with members of the alpha-2 subdivision of the class Proteobacteria. J. Bacteriol. 172: 3569-3576.

Ouahrani, S., Michaux, S., Sri Widada, J., Bourg, G., Tournebize, R., Ramuz, M., and Liautard, J.P. 1993. Identification and sequence analysis of IS6501, an insertion sequence in *Brucella* spp.: relationship between genomic structure and the number of IS*6501* copies. J. Gen. Microbiol. 139: 3265-3273.

Paulsen, I.T., Seshadri, R., Nelson, K.E., Eisen, J.A., Heidelberg, J.F., Read, T.D., Dodson, R.J., Umayam, L., Brinkac, L.M., Beanan, M.J., Daugherty, S.C., Deboy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Nelson, W.C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J., Van Aken, S.E., Riedmuller, S., Tettelin, H., Gill, S.R., White, O., Salzberg, S.L., Hoover, D.L., Lindler, L.E., Halling, S.M., Boyle, S.M., and Fraser, C.M. 2002. The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proc. Natl. Acad. Sci. USA. 99: 13148-13153.

Peterson, A.W., Wolf, L.K., and Georgiadis, R.M. 2002. Hybridization of mismatched or partially matched DNA at surfaces. J. Am. Chem. Soc. 124: 14601-14607.

Ramakrishnan, R., Dorris, D., Lublinsky, A., Nguyen, A., Domanus, M., Prokhorova, A., Gieser, L., Touma, E., Lockner, R., Tata, M., Zhu, X., Patterson, M., Shippy, R., Sendera, T.J., and Mazumder, A. 2002. An assessment of Motorola Code Link microarray performance for gene expression profiling applications. Nucleic Acids Res. 30: p. e30.

Riccelli, P.V., Hall, T.S., Pancoska, P., Mandell, K.E., and Benight, A.S. 2003. DNA sequence context and multiplex hybridization reactions: melting studies of heteromorphic duplex DNA complexes. J. Am. Chem. Soc. 125:141-150.

Rouillard, J.M., Herbert, C.J., and Zuker, M. 2002. OligoArray: genome-scale oligonucleotide design for microarrays. Bioinformatics. 18:486-487.

Rouillard, J.M., Zuker, M., and Gulari, E. 2003. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. Nucleic Acids Res. 31: 3057-3062.

Salzberg, S.L., Delcher, A.L., Kasif, S., and White, O. 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. 26:544-548.

Horizon Proofs

Sangari, F.J., and Garcia-Lobo, J.M. 2003. Analysis of the urease gene clusters in *Brucella* spp. In: Brucellosis 2003 International Research Conference, Universidad de Navarra, Spain. p. 84-85.

Schurig GG, Roop RM 2nd, Bagchi T, Boyle S, Buhrman D, and Sriranganathan N. 1991. Biological properties of RB51; a stable rough strain of *Brucella abortus*. Vet. Microbiol. 28: 171-188.

Stern, M.J., Ames, G.F., Smith, N.H., Robinson, E.C., and Higgins, C.F. 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell. 37: 1015-1026.

Sturgill, D. 2003. Comparative genome analysis of three *Brucella spp*. and a data model for automated multiple genome comparison. M.Sc. thesis. Virginia Tech., Blacksburg, VA, USA.

Tcherneva, E., Rijpens, N., Jersek, B., and Herman, L.M.F. 2000. Differentiation of *Brucella* species by random amplified polymorphic DNA analysis. J. App. Microbiol. 88: 69-80.

Tsolis, R.M. 2002. Comparative genome analysis of the alpha-proteobacteria: relationships between plant and animal pathogens and host specificity. Proc. Natl. Acad. Sci. U S A. 99: 12503-12505.

Velasco, J., Romero, C., Lopez-Goni, I., Leiva, J., Diaz, R., and Moriyón, I. 1998. Evaluation of the relatedness of *Brucella* spp. and *Ochrobactrum anthropi* and description of *Ochrobactrum intermedium sp*. nov., a new species with a closer relationship to *Brucella* spp. Int. J. Syst. Bacteriol. 48: 759-768.

Vemulapalli, R., McQuiston, J.R., Schurig, G.G., Sriranganathan, N., Halling, S.M., and Boyle, S.M. 1999. Identification of an IS*711* element interrupting the *wbo*A gene of *Brucella abortus* vaccine strain RB51 and a PCR assay to distinguish strain RB51 from other *Brucella* species and strains. Clin. Diagn. Lab. Immunol. 6: 760-764.

Verger, J.M., Grimont, F., Grimont, P.A.D., and Grayon, M. 1985. *Brucella*, a monospecific genus as shown by deoxyribonucleic acid hybridization. Int. J. Syst. Bacteriol. 35: 292–295.

Verstreate, D.R., and Winter, A.J. 1984. Comparison of sodium dodecyl sulfate-polyacrylamide gel electrophoresis profiles and antigenic relatedness among outer membrane proteins of 49 *Brucella abortus* strains. Infect. Immun. 46: 182-187.

Vizcaino, N., Cloeckaert, A., Verger, J., Grayon, M., and Fernandez-Lago, L. 2000. DNA polymorphism in the genus *Brucella*. Microbes Infect. 2: 1089-1100.

Vizcaino, N., Cloeckaert, A., Zygmunt, M.S., and Fernandez-Lago, L. 2001. Characterization of a *Brucella* species 25-kilobase DNA fragment deleted from *Brucella abortus* reveals a large gene cluster related to the synthesis of a polysaccharide. Infect. Immun. 69: 6738-6748.

Page Proofs