Microarray-Based Resequencing of Multiple *B. anthracis* Isolates

LCDR Michael E. Zwick, USNR
Biological Defense Research Directorate,
Naval Medical Research Center
&
Assistant Professor
Department of Human Genetics
Emory University School of Medicine

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE **01 OCT 2005**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Microarray-Based Resequencing of Multiple B. anthracisIsolates**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Biological Defense Research Directorate,Naval Medical Research Center**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

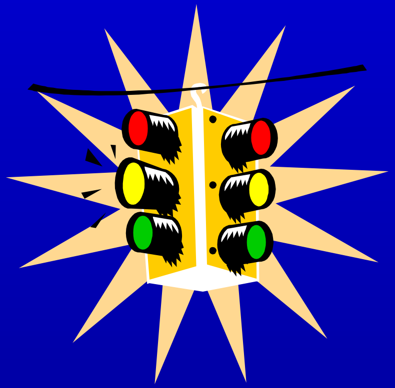
See also ADM001851, Proceedings of the 2003 Joint Service Scientific Conference on Chemical & Biological Defense Research, 17-20 November 2003. , The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**20**	

A Layered Approach: Levels of BW Testing



PRESUMPTIVE

(Hand-held assays)

CONFIRMATORY

(ELISA’s, PCR, Culture)

Detect to Treat

2 tests

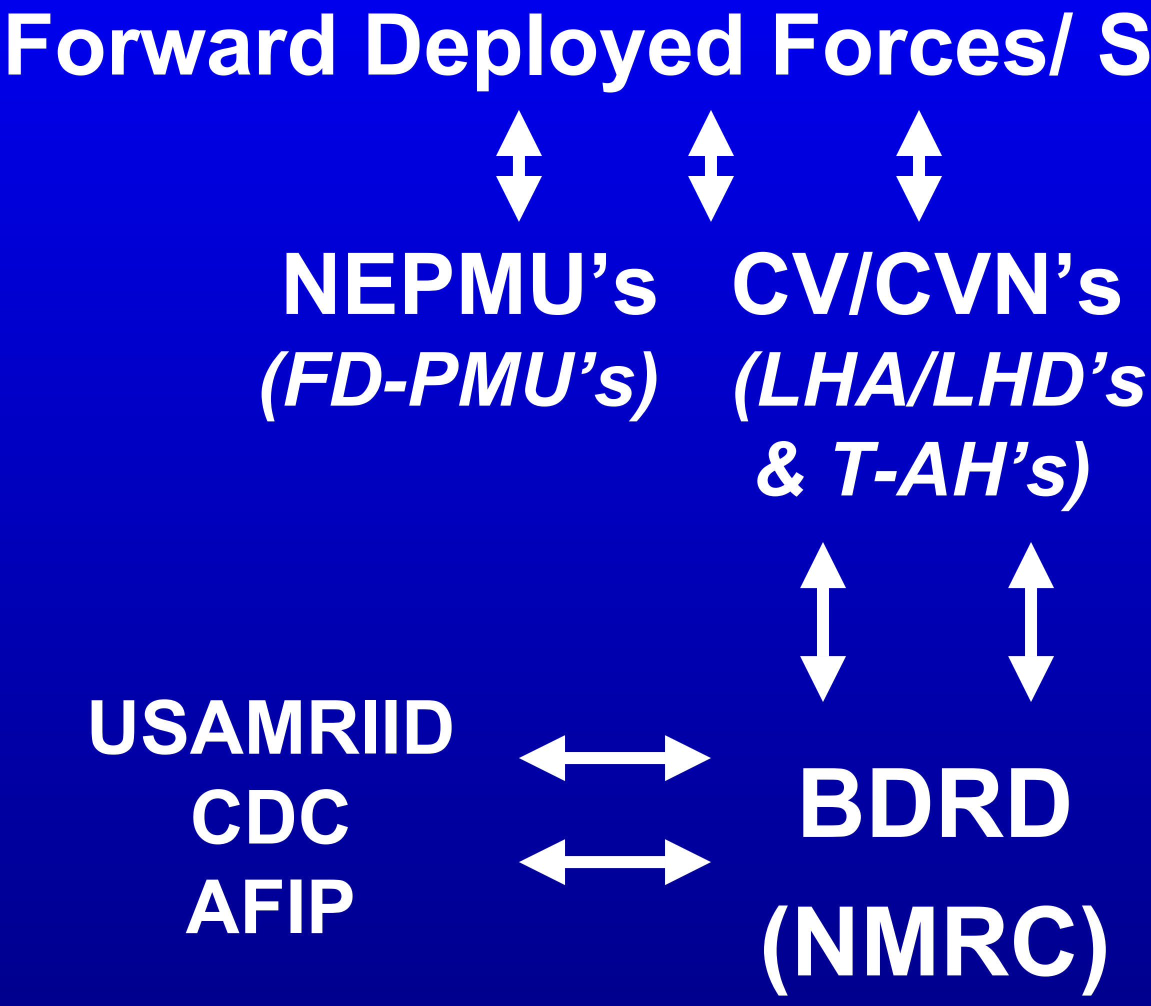
<24 hours

DEFINITIVE

(Technical Reachback, Monthly QC)

(Full-scale analytical work up by the experts)

Navy BW Testing Assets



Forward Deployed Forces/ Small Ships

NEPMU's
(FD-PMU's)

CV/CVN's
(LHA/LHD's
& T-AH's)

USAMRIID
CDC
AFIP

BDRD
(NMRC)

Coronado
Enterprise
George Washington
Harry S. Truman
John F. Kennedy
Theodore Roosevelt
Abraham Lincoln
Carl Vinson
John C. Stennis
Kitty Hawk
Nimitz
Bataan
Iwo Jima
Kearsarage
Nassau
Saipan
Wasp
Bonhomme Richard
Boxer
Essex
Peleliu
Tarawa
Comfort
Mercy

How Can We Detect and Identify BW Agents?

Genotype markers known to show variation

- Fixed species specific variants, previously identified
- Rapid detection of a small number of sites
- Example: Real-Time PCR (Confirmatory Lab)

DNA sequence regions/genomes of interest

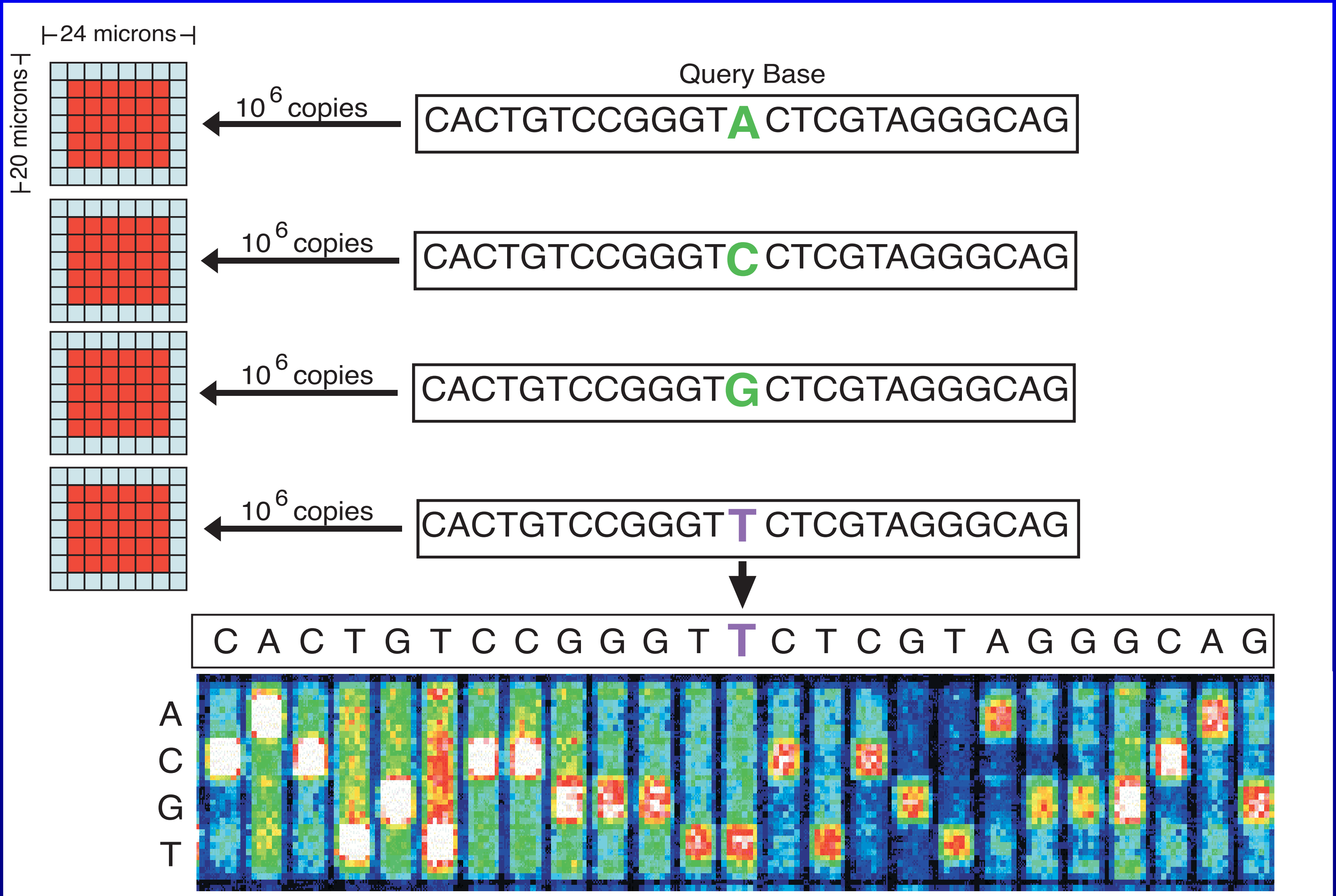
- Maximally informative:

The sequence is the genotype!

- Detects common and rare variants
- Strain identification/origin (Definitive Lab)

The future detection and identification of BW agents will increasingly depend upon DNA sequencing technologies

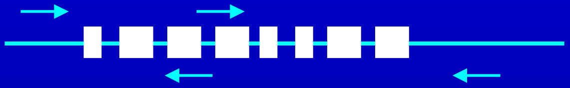
Design of Resequencing Arrays



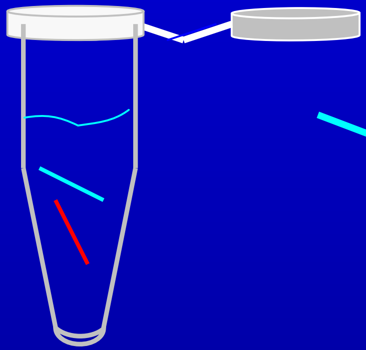
🟥 Raw Data Pixels ⬜ Masked Pixels **A,C,G** Possible SNPs **T** Reference base

Resequencing Assay

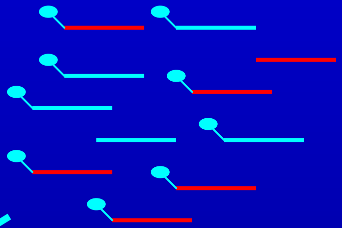
Long PCR/Whole Genome Amplification



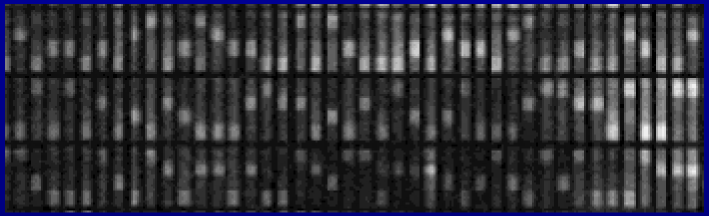
PCR products
pooled by individual;
DNAse I treated
(50 bp average)



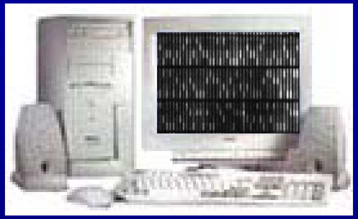
DNA fragments biotinylated



Tagged fragments hybridized
to an oligonucleotide array;
stained with streptavidin phycoerythrin

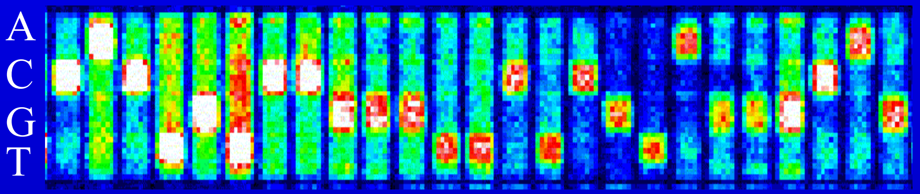


Analyzed by ABACUS
to detect variation



Resequencing *B. anthracis*

- 29.5 kb of unique sequence per chip.
- Each array has ~320,000 features.
- Forward and reverse strands tiled.
- 1 design, 6 LPCR assays
- pXO1, pXO2, Main Chromosome: All or part of 32 genes
- lef, pag, cap, vrr, rpoB, sasB



CACTGTCCGGGTTCTCGTAGGGCAG

CACTGTCCGGGTACTCGTAGGGCAG
CACTGTCCGGGTCCTCGTAGGGCAG
CACTGTCCGGGTGCTCGTAGGGCAG
CACTGTCCGGGTTCTCGTAGGGCAG

ACTGTCCGGGTTATCGTAGGGCAGT
ACTGTCCGGGTTCTCGTAGGGCAGT
ACTGTCCGGGTTGTCGTAGGGCAGT
ACTGTCCGGGTTTTCGTAGGGCAGT

…

❖ How certain are we of this G ?

ABACUS: An Automated Statistical Algorithm for Base/Genotype Calling

- Within any given feature, florescence intensities of individual pixels are assumed to be independent and identically distributed Gaussian variables.
- Forward and reverse strands are treated as independent replicates (with different parameters).
- All parameters are fit by maximum likelihood.
- 5 models for haploid data (null,A,C,G,T).
- 11 models for diploid data (null, AA,CC,GG,TT,AC, AG, AT, CG, CT, GT).
- Neighborhood quality rules are used.

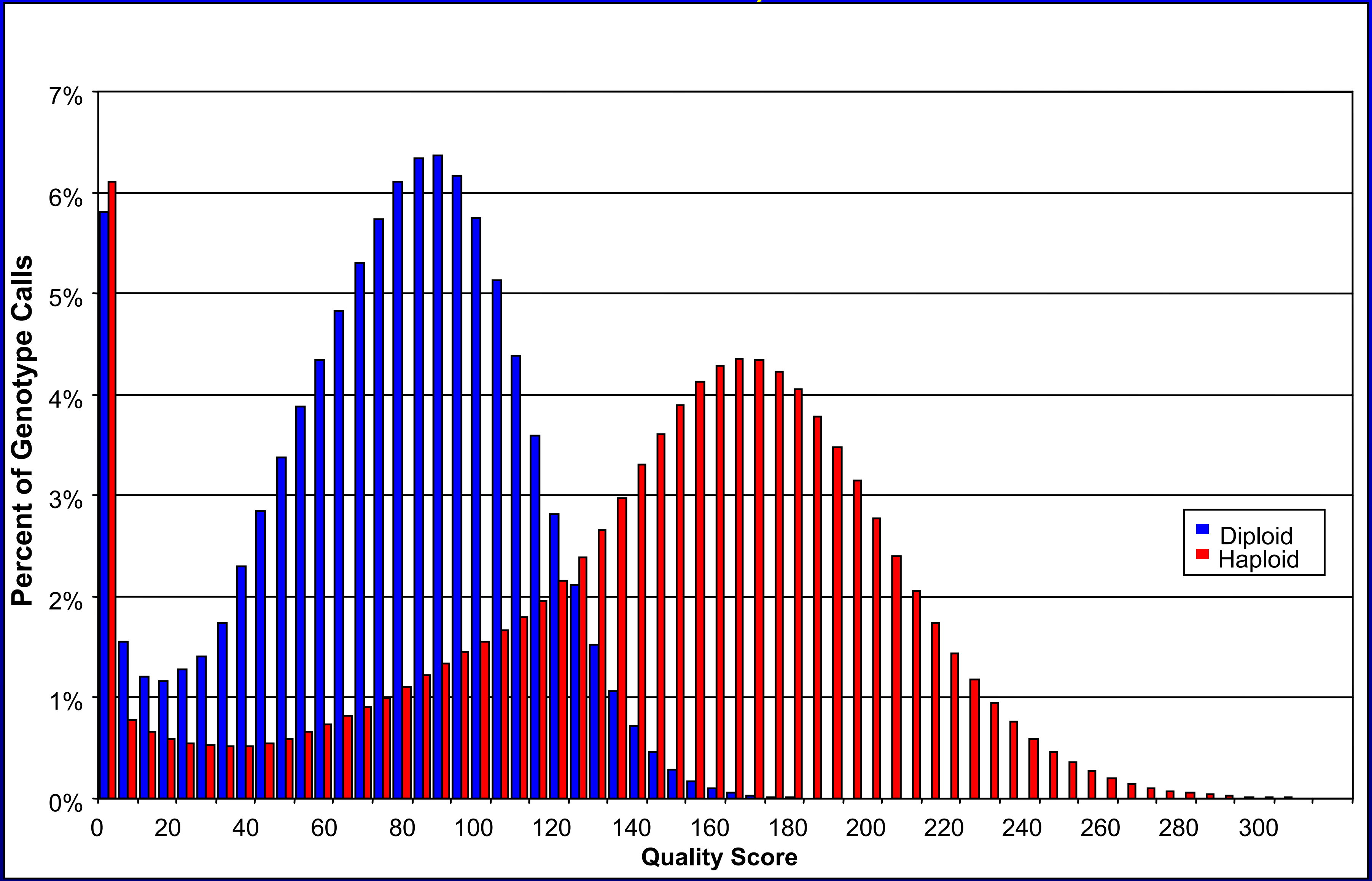
ABACUS Assigns Quality Scores to Each Base/Genotype Call

- A *Quality Score*, the difference between the $\log_{10}$ likelihood of the best fitting and second best fitting model, is assigned to each genotype.
- Information from both the forward and reverse strands is incorporated into the *Quality Score*.
- Genotypes inferred only when a *Quality Score* threshold is reached.

For more detail, see Cutler, DJ, Zwick, ME *et al.*

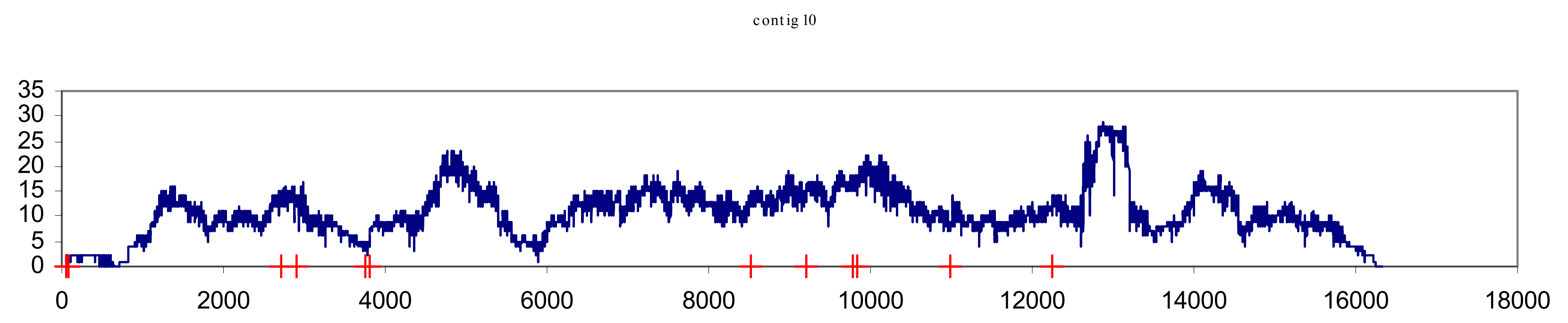
Genome Res. 2001 11: 1913-1925

Distribution of *Quality Scores* (Human Data)



Haploid ABACUS Base Calls Are Highly Accurate (QS>30)

- LPCR fragments hydrosheared
- Individual 8 from FMR1
- Subcloned with end-repair into PUC Library
- Single Pass sequenced with M13 primers
- At least 6x



- 17,423 bp with at least 6x coverage, all identical to ABACUS calls
- At 2x coverage, an additional 4,081 bp, with 1 difference from ABACUS calls

ABACUS Genotype Calls Are Highly Repeatable

- Haploid
  - 0 differences / 841,236 sites (QS>30)
- Diploid
  - 0 differences / 812,944 homozygotes (QS>30)
  - 0 differences / 351 heterozygotes (QS > 30)

- Implies a phred score of at least 54

B. anthracis Resequencing Experiment

- Chips Hybridized and Scanned: 114

Successful: 112

Experimental Failure: 2

- *B. anthracis* Isolates Analyzed: 59

Replicated: 53 (106 chips)

Single Analysis: 6 (6 chips)

Microarrays Can Generates Vast Amounts of Sequence Data

- Raw Sequence Generated

Bases Called: 3,052,254

Total Possible Bases: 3,271,744

Call Rate: 93.3%

- Variant Sites Discovered

38 Single Nucleotide Polymorphisms (SNPs)

16 of 38 SNPs singletons

22 SNPs found more than once

Anthrax Resequencing is Highly Replicable

Total Comparisons	1,420,583
Total Bases Called	2,897,098
Total Discrepancies	1

- Suggests error rates of less than 1 per million
- Quality Score Threshold: 31
- Sequences on chip: 34.7% GC Content

How different are two *B. anthracis* isolates?

- Variation Estimates
  - Tajima's Estimate of Theta: 1.6×10^{-4}
  - Watterson's Estimate of Theta: 2.9×10^{-4}
- Two Isolates of *B. anthracis* are expected to differ at between:

~924 (Tajima)
and
~ 1606 (Watterson)

Resequencing can uniquely identify *B. anthracis* isolates

Acknowledgements

Biological Defense Research Directorate (BDRD),
Naval Medical Research Center (NMRC)
Farrell Mcafee
CAPT Darrell Galloway
CAPT Alfred Mateczun

Johns Hopkins University
Dr. David Cutler

NAVAL MEDICAL RESEARCH CENTER
U.S. NAVY

Points of Contact

Biological Defense Research Directorate (BDRD)

24/7 BDRD Watch Stander
877-243-1528, Secondary 877-243-1531

NMRC Officer of the Day
DSN: 285-9053, Comm: (mobile) 301-526-1649

BDRD Main Office
301-319-7510

bdrd@nmrc.navy.mil
bdrd.ops@intecwash.navy.smil.mil

STU III
301-319-7509/DSN 285, FAX 301-295-0137

Assessing ABACUS Performance

- Replicability: Comparison of haploid/diploid replicates by independent:
  - PCR amplification of genomic DNA
  - Manufacture of resequencing arrays (distinct wafers)
  - Hybridization of amplified DNA to chips
  - ABACUS genotype calls
- Accuracy: Independent Genotyping/DNA Sequencing

All genotyping technologies should be assessed using these criteria

Diploid ABACUS Genotype Calls Are Highly Accurate (QS>30)

- Homozygous genotypes
  - 0 differences / 1,515 genotypes (100% correct)
- Heterozygous genotypes
  - 3 differences / 423 genotypes (99.3% correct)

- Two of the three differences were in a single LPCR fragment
- All three differences were at high frequency sites
- Chips called heterozygote, sequencing called homozygote

Probable Cause: Sample Mixing