State of the Art for Autonomous Detection Systems using Genomic Sequencing
BioWatch Needs

1. Affordable Continuous Coverage of At-Risk Populations
2. Accuracy and Precision Supporting High Regret Responses
3. Responsiveness to Full Scope of BW/BT Threat
High-Throughput DNA/RNA Sequencing Technologies
Past ---> Present ---> Future

- Sanger - 1975
- ABI gel “automated” - 1986
- ABI Capillary 1999 – current

Capillary Based Sequencer, 70 kb / run
- Needed higher throughput...new technologies
- Must be relatively cost effective & flexible

Who will win and What’s next ???

Seq. by Synthesis
- 2005
  - 400bp / 400Mb Roche 454 – 2nd Gen
  - 250bp / 600Gb Hiseq2500 – 2nd Gen
- 2010
  - 4000bp / 80Mb PacBio – 3rd Gen
- 2011
  - 2 hrs
  - 2011 (miseq)
  - 60 min
  - 400bp / 60Mb Ion Torrent 2nd Gen

Semiconductor Seq.
- 2hrs
- Needed higher throughput...new technologies
- Must be relatively cost effective & flexible

Pyrosequencing
- 8 hrs
NGS process workflow

1. Sample (throat swab, blood, soil surface, water, etc.)
2. Preparation of a sequencing library
3. Clonal amplification of library fragments
4. Sequencing
5. Computational analysis of sequencing reads
NGS applications

Detection and characterization of pathogens:
- in the environment (soils, rivers, etc)
- facilities (hospitals, dorms, offices, etc)
- animals
- people

<table>
<thead>
<tr>
<th>Sample types</th>
<th>Clinical</th>
<th>Environmental</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Swabs</td>
<td>Swipes</td>
</tr>
<tr>
<td></td>
<td>Blood</td>
<td>Soil</td>
</tr>
<tr>
<td></td>
<td>Stool</td>
<td>Liquids</td>
</tr>
</tbody>
</table>

- Culturing and DNA/RNA extraction (pure cultures)
- DNA / RNA extraction
- Preparation of sequencing libraries (amplicons or shotgun libraries)
- Sequencing on next generation platforms
  - MiSeq (Illumina)
  - PGM / Proton (Ion Torrent/Life Tech)
# Sequencing Applications

<table>
<thead>
<tr>
<th>Pipeline</th>
<th>Description</th>
<th>Information provided</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amplicon Sequencing</strong></td>
<td>Rapid sequencing of small portions of target genomes</td>
<td>• ID &amp; characterize known pathogens and features&lt;br&gt;• Can test 100’s of samples</td>
</tr>
<tr>
<td><strong>Sequencing mixed samples</strong></td>
<td>Shotgun sequencing of environmental or clinical samples</td>
<td>• ID &amp; characterize both known and emerging pathogens</td>
</tr>
<tr>
<td><strong>Sequencing an isolate</strong></td>
<td>Full genome sequence of a single organism after laboratory isolation and growth</td>
<td>• ID sequences associated with specific outbreaks&lt;br&gt;• Allows rapid detection of the same pathogen in future outbreaks</td>
</tr>
</tbody>
</table>
Rapid Response Scenario
Sample A

- **Environmental Sample**
- Sample prep issues
- Standard and WGA seq’ed
- 4 platforms generated data for target

**Francisella tularensis**
- PacBio ID target first at 2.5d (WGA product)
- Illumina data sets obtained F. tul genome at 19x coverage with 99.6% of the genome covered (as compared to the FSC198 reference)
- Individual & combined assemblies and annotations generated and uploaded to BioAgent
Rapid Response Scenario
Sample B

- **Clinical Sample**
- **Sample prep issues**
- **Standard and WGA seq’d**
- **4 platforms generated data for target**

- **Hepatitis B virus**
- **PacBio ID target first at 22hrs (standard product)**
- **>19,000x (100%) coverage generated**
- **Assembly/Mapping & Annotation generated and uploaded to BioAgent**

- **Yersinia enterolitica**
- **LANL ID in all platforms.**
  - PB: 5 of 226,787 filtered reads (0.002%) map to Yersinia
  - 454: 8 of 398,565 reads (0.002% of reads) (by blast) map to Yersinia
  - Illumina: 279 of 815,836 (.00003%) mapped to Yersinia

- **Described symptoms suggested causative agent was likely Yersinia, not Hep B**
Potential for Sequencing to Meet BioWatch Autonomous Detection Requirement Areas:

- Sensitivity
- Specificity
- Size and Functionality
- Measuring Timelines
- Measuring Interval
- Cost
- Automation
- Operation Environment
Readiness Levels

• Tier 1- TRL 6 by 2016
  – Uses High-throughput Amplicon Sequencing
  – Benefits from comprehensive knowledgebase generated from NGS
  – Requires engineering of existing technology & integration into BioWatch system
  – Detects all known pathogens, antibiotic resistance markers, & virulence factors
  – Has limited ability to identify emerging threats, especially RNA viruses

• Tier 2- TRL 6 by 2020
  – Uses improved NGS technology to provide for metagenomic sequencing
  – Can also detect many emerging pathogens, including engineered ones
  – Even with improved sequencing speeds, the requirement to prepare libraries slows the overall process down

• Tier 3- TRL 6 >2020
  – Use a future sequencing technology (i.e. Oxford Nanopore or Gen2 Pacbio) to rapidly perform metagenomic sequencing of environmental samples
  – Similar to Tier 2, but no library prep needed and has a much reduced reagent requirement
Key Points

1. Sequencing is rapidly becoming the Gold standard for Biological Identification
   - Phylogeny is being revisited
   - Can cover the entire threat (e.g. SARS, Novel Coronavirus, etc.)

2. We don’t yet know the full potential of sequencing for Bio Detection in the field
   - Designed for high-throughput (Human Genome)
   - Hardware and Knowledgebases are already being built (dual use platforms will be most effective....share cost, risk, & processes)
   - Forensics and Biosurveillance share a similar Seq adoption path
   - Has transformed microbiology in the background (e.g. NIH’s human microbiome project, BGI’s 10k microbial genome project, JGI’s microbial earth project)
3. Bioinformatics and Metadata will be key
   - Opportunity to have a shared knowledgebase across all venues
   - Possible core analysis brain with many functional nodes
   - Sequence everything and let the bioinformatics gods sort it out!
   - Need a rapid decision process

4. Need autonomous and modular platform
   - Integration & Engineering will be challenging, but huge unknown since we have not really tried it yet
   - Ease of use is critical for wide acceptance (i.e. Apple products)

5. Must be responsive to the public health official’s needs
   - Provide acceptable & actionable information for their decision making process