Molecular Barcoding for Multiplexing

Now that you can sequence the human genome on a sequencer, what do you do for microbes?
Where we are

- 13:30-14:00 – Primer Design to Amplify Microbial Genomes for Sequencing
- 14:00-14:15 – Primer Design Exercise
- **14:15-14:45 – Molecular Barcoding to Allow Multiplexed NGS**
- 14:45-15:15 – Processing NGS Data – de novo and mapping assembly
- 15:15-15:30 – Break
- 15:30-15:45 – Assembly Exercise
- 15:45-16:15 – Annotation
- 16:15-16:30 – Annotation Exercise
- 16:30-17:00 – Submitting Data to GenBank
Why add molecular barcodes?

• Goal is to sequence as many genomes as possible, to the required depth of coverage, as economically as possible.
• All NGS platforms have their own methods for molecular barcoding, but are often expensive.
• JCVI developed a method for adding our own barcodes, less expensive, but has some drawbacks.
Early SISPA research at JCVI was developed to amplify viral genome sequences for downstream TA cloning and Sanger sequencing.

The method has been adapted to allow molecular barcoding of samples by designing 100’s of unique sequences that have the following characteristics:

- Tm and sequence composition is compatible with SISPA protocol
- Compatible with NextGen library construction and sequencing protocols
- Distinguishable from one another, allowing up to 10% sequencing error

Bar-Coding and Genome Amplification
Sequence Independent Single Primer Amplification (SISPA)
Sample Pooling/NextGen Sequencing

[Diagram showing processes related to sample pooling and next-generation sequencing]
Our typical SISPA Protocol

• For SISPA Barcoding
  o For every ~15kb of viral genome – give sample two SISPA barcodes - examples
    – Norovirus – 7.5kb – two unique SISPA barcode
    – Influenza Virus – 15kb – two unique SISPA barcodes
    – Coronavirus – 30kb – four unique SISPA barcodes

• Normalize and Pool SISPA products
  o 192 normalized SISPA products go into our typical pool

• Illumina Libraries and Pooling into Lanes
  o Each SISPA product pool is used to build an indexed Illumina PE library or Roche/454 Rapid Library
  o On Illumina HiSeq 2000, we have been pooling 3 of these libraries per lane
### Performance on HiSeq 2000

<table>
<thead>
<tr>
<th>LANE</th>
<th>Library Name</th>
<th>Yield (Mbases)</th>
<th>% PF</th>
<th># Reads</th>
<th>% of &gt;= Q30 Bases (PF)</th>
<th>Mean Quality Score (PF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>349</td>
<td>SISPA20120406-PE-IL15-01</td>
<td>4,622</td>
<td>63.85</td>
<td>71,665,850</td>
<td>62.67</td>
<td>26.87</td>
</tr>
<tr>
<td>349</td>
<td>SISPA20120407-PE-IL16-01</td>
<td>4,174</td>
<td>62.64</td>
<td>65,965,754</td>
<td>62.38</td>
<td>26.82</td>
</tr>
<tr>
<td>349</td>
<td>SISPA20120408-PE-IL23-01</td>
<td>3,623</td>
<td>63.03</td>
<td>56,901,646</td>
<td>66.01</td>
<td>27.81</td>
</tr>
<tr>
<td>350</td>
<td>SISPA20120416-PE-IL24-01</td>
<td>3,457</td>
<td>72.38</td>
<td>47,295,656</td>
<td>69.11</td>
<td>28.89</td>
</tr>
<tr>
<td>350</td>
<td>SISPA20120417-PE-IL26-01</td>
<td>5,697</td>
<td>70.94</td>
<td>79,503,452</td>
<td>68.83</td>
<td>28.81</td>
</tr>
<tr>
<td>350</td>
<td>SISPA20120418-PE-IL27-01</td>
<td>6,500</td>
<td>71.62</td>
<td>89,859,024</td>
<td>69</td>
<td>28.87</td>
</tr>
<tr>
<td>339</td>
<td>SISPA20120326RESUB-PE-IL10-01</td>
<td>4,628</td>
<td>65.31</td>
<td>70,163,402</td>
<td>68.44</td>
<td>28.67</td>
</tr>
<tr>
<td>339</td>
<td>SISPA20120328-PE-IL12-02</td>
<td>9,831</td>
<td>73.75</td>
<td>131,980,380</td>
<td>72.82</td>
<td>29.91</td>
</tr>
<tr>
<td></td>
<td>avg</td>
<td>5,317</td>
<td>67.9</td>
<td>76,666,896</td>
<td>67.4</td>
<td>28.3</td>
</tr>
<tr>
<td></td>
<td>stdev</td>
<td>2,088</td>
<td>4.7</td>
<td>25,847,948</td>
<td>3.5</td>
<td>1.1</td>
</tr>
</tbody>
</table>

**PhiX spike in lane**

<table>
<thead>
<tr>
<th>LANE</th>
<th>Library Name</th>
<th>Yield (Mbases)</th>
<th>% PF</th>
<th># Reads</th>
<th>% of &gt;= Q30 Bases (PF)</th>
<th>Mean Quality Score (PF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>352</td>
<td>20120515SISPA21-PE-IL29-01</td>
<td>4,532</td>
<td>73.56</td>
<td>60,999,882</td>
<td>69.62</td>
<td>29.03</td>
</tr>
<tr>
<td>352</td>
<td>SISPA20120425-PE-IL28-01</td>
<td>3,164</td>
<td>72.44</td>
<td>43,244,082</td>
<td>70.02</td>
<td>29.1</td>
</tr>
<tr>
<td>352</td>
<td>SISPA20120601-PE-IL30-01</td>
<td>3,117</td>
<td>72.23</td>
<td>42,725,366</td>
<td>67.33</td>
<td>28.3</td>
</tr>
<tr>
<td>352</td>
<td>PhiX spike in + unknown</td>
<td>4,413</td>
<td>93.44</td>
<td>46,765,556</td>
<td>88.37</td>
<td>34.59</td>
</tr>
<tr>
<td></td>
<td>avg</td>
<td>3,604</td>
<td>72.7</td>
<td>48,989,777</td>
<td>69.0</td>
<td>28.8</td>
</tr>
<tr>
<td></td>
<td>stdev</td>
<td>804</td>
<td>0.7</td>
<td>10,404,289</td>
<td>1.5</td>
<td>0.4</td>
</tr>
</tbody>
</table>
NOTE: We limit depth of coverage to 200x max so that assembly files remain at reasonable size, and other analyses can run in under 2 hours.
SISPA Software

- **JCVI Sequence BARcode Designer** – to design SISPA barcodes, http://sourceforge.net/projects/jcvibard/
- **JCVI DNA Barcode Deconvolution** – to demultiplex reads based on barcodes used in a NGS run, http://sourceforge.net/projects/deconolver/
NGS Vendor Barcoding

- Roche/454 – Multiplex Identifier (MID) in adaptors
- Illumina – Nextera tagmentation and Illumina Index reads
- LifeTech Ion PGM – IonXpress barcoding
- All of these are added as part of individual library construction, and demultiplexed by vendors’ software