GENOME SEQUENCING AND ASSEMBLY

BIO 300/CMPSC 300
Spring 2016
What is a Genome?

- an organism’s complete set of DNA, including all of its genes, regulatory regions, non-coding regions, etc.
- an organism’s complete set of genetic instructions
<table>
<thead>
<tr>
<th>Organism</th>
<th>Number of genes in the genome</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mycoplasma genitalium</em></td>
<td>517</td>
</tr>
<tr>
<td><em>Saccharomyces cerevisiae</em></td>
<td>6,275</td>
</tr>
<tr>
<td><em>Arabidopsis thaliana</em></td>
<td>~20,000</td>
</tr>
<tr>
<td><em>Caenorhabditis elegans</em></td>
<td>19,099</td>
</tr>
<tr>
<td><em>Haemophilus influenzae</em></td>
<td>1,743</td>
</tr>
<tr>
<td><em>Drosophila melanogaster</em></td>
<td>13,601</td>
</tr>
<tr>
<td><em>Neisseria meningitdis</em></td>
<td>2,158</td>
</tr>
<tr>
<td><em>Homo sapiens</em></td>
<td>20,000–25,000</td>
</tr>
</tbody>
</table>
Genome Projects

• Goals:
  • Determine complete genome sequence of an organism
  • Annotate protein-coding genes and other important genome-encoded features
Genome Projects

• **Goals:**
  • Determine complete genome sequence of an organism
  • Annotate protein-coding genes and other important genome-encoded features

• **Projects:**
  • Over 15,000 [genome projects](#) in progress or completed
Genome Projects

• Goals:
  • Determine complete genome sequence of an organism
  • Annotate protein-coding genes and other important genome-encoded features

• Projects:
  • Over 15,000 genome projects in progress or completed
Genome Sequencing
It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way - in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.
Shredded Book Reconstruction

- Dickens accidently shreds first printing of Tale of Two Cities
  - first printing = 5 copies

It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of innocence...
Shredded Book Reconstruction

- Dickens accidently shreds first printing of Tale of Two Cities
  - first printing = 5 copies
  - shredding was random (can cut between different words in each copy)
  - always 5 words per fragment

It was the best of times, it was the worst of times, it was the...
Shredded Book Reconstruction

• Dickens accidently shreds first printing of Tale of Two Cities
  • first printing = 5 copies
  • shredding was random (can cut between different words in each copy)
  • always 5 words per fragment

It was the best of times, it was the worst of times, it was the
It was the best of times, it was the worst of times, it was the
It was the best of times, it was the worst of times, it was the

5 copies x 138, 656 words/5 words per fragment = 138k fragments

All short fragments are mixed together
It was the best of age of wisdom, it was the best of times, it was the age of wisdom, it was the age of foolishness, it was the worst of times, it was the worst of times, it was the age of foolishness, it was the age of wisdom, it was the best of times, it was the worst of times, it was the worst of times, it was the best of times, it was the best of times, it was the worst of times, it was the worst of times, it was the age of foolishness, it was the age of wisdom
It was the best of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of fools.
It was the best of age of wisdom, it was the best of times, it was it was the age of wisdom, it was the worse of times, it was the age of wisdom, it was the best of times, it was the worst of times, it was the age of foolishness, was the best of times, it was the age of wisdom, it was the worst of times, it was the best of times, it was
It was the best of times, it was the age of wisdom, it was the best of times, it was the age of foolishness, it was the worst of times, it was the age of wisdom, it was the best of times, it was the age of foolishness, it was the best of times, it was the age of wisdom, it was the best of times, it was the age of foolishness, it was the best of times, it was the age of wisdom.
It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the age of believe the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness.
It was the best of age of wisdom, it was
best of times, it was
it was the age of
it was the age of
it was the worst of
of times, it was the
of times, it was the
of wisdom, it was the
the age of wisdom, it
the best of times, it
the worst of times, it
worst of times, it was
wisdom, it was the age
was the worst of times,
was the best of times,
was the age of foolishness,
was the age of wisdom
times, it was the worst
It was the best of times, it was the best of times, it was the worst of times, it was the worst of times, it was the age of wisdom, it was the age of wisdom, it was the best of times, it was the best of times, it was the worst of times, it was the worst of times, it was the age of foolishness, it was the age of foolishness, it was the age of wisdom, it was the age of wisdom.
Tale of Two Cities – Charles Dickens

It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of Light, it was the season of Darkness, it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going direct the other way - in short, the period was so far like the present period, that some of its noisiest authorities insisted on its being received, for good or for evil, in the superlative degree of comparison only.
It was the best of age of wisdom, it was the best of times, it was the age of worst of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times.

Assembly Parameter:
100% identify across 4 words

Repeats pile up – actual placement of each individual fragment unknown
It was the best of age of wisdom, it was the best of times, it was the age of wisdom, it was the age of foolishness, it was the best of times, it was the worst of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times, it was the best of times, it was the worst of times.  

Repeats pile up – actual placement of each individual fragment unknown  

Repeats can cause ambiguity and prevent proper assembly  

Assembly Parameter:  
100% identify across 4 words
It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness.
Genome Sequencing

Cloned genomes

Multiple genomes are sheared into variable sized segments

Unordered sequenced segments

Computational automated assembly

Resulting overlapping sequence segments. (The higher the coverage the better the quality of the sequencing.

Overlapping sequence segments combined to construct the genome consensus.

ATGTTCCGATTAGGAAAACTATCTGTAACCTTTCATTCCAGTAAGGAGGAAATATAA
Coverage

random short sequence reads

coverage

assembly of overlapping fragments

assembled contig sequence
Finding the Largest Overlap

• Consider two fragment assembly:
  • If there is more than one overlap, choose the **longest** overlap
  • Assume the sequences are not identical
  • Assume neither sequence is a substring of the other
  • The longest **possible** overlap is length of the shorter sequence - 1
Finding the Largest Overlap

- Removing assumptions of identical subsequences and substrings:
  - set the initial $n$ to the length of the sequence rather than the length of the sequence – 1.
Dealing with Noisy Sequencing Data

- Sequencing errors
- Ambiguities leading to incorrect base-calling

- Modify the algorithm so that the overlap exceeds some threshold value (instead of being perfect match)
  - Check if the number of matching bases is \textbf{threshold value} \times n
  - With the \textbf{threshold value} being between 0 and 1
Assembling a Contig

Table 8.3 Overlaps for a hypothetical set of sequence reads.

<table>
<thead>
<tr>
<th>Fragments</th>
<th>Overlaps (Length)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. TACCTTG</td>
<td>2 (3), 3 (1), 4 (1), 7 (1)</td>
</tr>
<tr>
<td>2. TTGAT</td>
<td>1 (1), 3 (3)</td>
</tr>
<tr>
<td>3. GATATGG</td>
<td>4 (2), 7 (1)</td>
</tr>
<tr>
<td>4. GGAG</td>
<td>3 (1), 7 (1)</td>
</tr>
<tr>
<td>5. CTCTA</td>
<td>1 (2), 6 (3)</td>
</tr>
<tr>
<td>6. CTAGT</td>
<td>1 (1), 2 (1)</td>
</tr>
<tr>
<td>7. GCTCT</td>
<td>1 (1), 2 (1), 5 (4), 6 (2)</td>
</tr>
</tbody>
</table>
Assembling a Contig: graph representation