The RDP-II (Ribosomal Database Project)
Center for Microbial Ecology, Michigan State University, East Lansing, MI 48824

James R. Cole,1 Timothy G. Lilburn,2 Ryan J. Farris,4 Benli Chai, Siddique Kulam,1 James M. Tiedje1
1Michigan State University, East Lansing, MI
2American Type Culture Collection, Manassas, VA

Abstract:
The Ribosomal Database Project - II (RDP-II) provides data, tools and services related to ribosomal RNA sequences to the research community. Through its website (http://rdp.cme.msu.edu), RDP-II offers aligned and annotated rRNA sequences, tools for alignment and phylogenetic inference, and services for data submission. RDP-II release 8.1 (May 21, 2001) contains 16,277 prokaryotic, 5201 eukaryotic, and 1503 mitochondrial small subunit rRNA sequences in aligned and annotated format. The sequence alignments provided by the RDP-II site are based on secondary structure conservation. The RDP-II website has been modified in a number of ways, including an object oriented database design that is now being extended to a production scale test. This has allowed us to release alignments on a larger scale, allowing the RDP-II website to support a greater number of users. We are currently working on the quality of alignments produced by this tool, and hand-tuned alignment or of the automated alignment, indicating the two alignments are equally valid at this level of resolution. The RDP intends to use this new automated method to help provide data to users in a more timely manner. In addition, the automated alignment system that directly takes into account primary similarity reduces the potential for human error or unintended bias in the alignment process. The Ribosomal Database Project - II (RDP-II) provides data, tools and services related to ribosomal RNA sequences to the research community. Through its website (http://rdp.cme.msu.edu), RDP-II offers aligned and annotated rRNA sequences, tools for alignment and phylogenetic inference, and services for data submission. RDP-II release 8.1 (May 21, 2001) contains 16,277 prokaryotic, 5201 eukaryotic, and 1503 mitochondrial small subunit rRNA sequences in aligned and annotated format. The sequence alignments provided by the RDP-II site are based on secondary structure conservation. The RDP-II website has been modified in a number of ways, including an object oriented database design that is now being extended to a production scale test. This has allowed us to release alignments on a larger scale, allowing the RDP-II website to support a greater number of users. We are currently working on the quality of alignments produced by this tool, and hand-tuned alignment or of the automated alignment, indicating the two alignments are equally valid at this level of resolution. The RDP intends to use this new automated method to help provide data to users in a more timely manner. In addition, the automated alignment system that directly takes into account primary similarity reduces the potential for human error or unintended bias in the alignment process.

Hand Alignment vs. SCFG

<table>
<thead>
<tr>
<th>&lt;100 nucleotides</th>
<th>100-499 nucleotides</th>
<th>500-1199 nucleotides</th>
<th>1200-1599 nucleotides</th>
<th>&gt;1599 nucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>%</td>
<td>Success</td>
<td>%</td>
<td>Success</td>
</tr>
<tr>
<td>Hand Alignment</td>
<td>95.2%</td>
<td>96.7%</td>
<td>95.1%</td>
<td>94.2%</td>
</tr>
<tr>
<td>SCFG Alignment</td>
<td>96.4%</td>
<td>96.8%</td>
<td>95.5%</td>
<td>94.3%</td>
</tr>
</tbody>
</table>

Some First Stage Heuristic Failures:
- Improved alignment results
- Enhanced confidence in secondary structure
- Detect common artifacts
- Implement half-delta helix state
- Modify heuristics strategy for missing data
- Provide new alignments to users

Which Tree is Better?

<table>
<thead>
<tr>
<th>Hand</th>
<th>SCFG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Best</td>
</tr>
<tr>
<td>Tree</td>
<td>-35543</td>
</tr>
<tr>
<td>Diff in L</td>
<td>27</td>
</tr>
<tr>
<td>S.D.</td>
<td>43</td>
</tr>
<tr>
<td>Worse?</td>
<td>No</td>
</tr>
</tbody>
</table>

Kishino-Hasegawa-Templeton Test

Hand Alignment: -35543
SCFG Alignment: -37648

The RDP-II email address for questions or comments is rdpstaff@msu.edu

Production Scale Test Results

<table>
<thead>
<tr>
<th>SCFG Assigned 88% of Bases to Model Positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
</tr>
<tr>
<td>250-1100</td>
</tr>
<tr>
<td>1200-1599</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Production Scale Test Details

- SCFG Alignment: -37648
- Hand Alignment: -35543
- Use expected probability value a
- Implement half-delta helix state
- Modify heuristics strategy for missing data
- Provide new alignments to users

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)

Production Scale Test

- 2D Structure Conserved from Domain to Family
- Model primary and secondary structure
- Can't model pseudoknots or higher-order interactions
- Time complexity O(|S|T)
- Space complexity O(|S|T)