A workflow with dedicated tools for preparing reference transcriptomes from non-model organisms has evidenced important biological information

Hicham Benzekri1, Pedro Seoane1, Rosario Carmona1, Rocio Bautista2, Darío Guerrero2, Noé Fernández-Pozo1 and M. Gonzalo Claros1,2

1) Departamento de Biología Molecular y Bioquímica (Universidad de Málaga), Campus de Teatinos, 29071 Málaga, Spain
2) Plataforma Andaluza de Bioinformática (Universidad de Málaga), Severo Ochoa, 34, 29590 Málaga, Spain

The problem

- Bioinformatic tools for NGS are mainly focused in model organism
- Non-model organisms lack reference sequences and usually are heterozygous
- De novo transcriptome assembling of non-model species requires non-reference based evaluation

Software

Discarded
- Trinity
- ABySS
- CABOG
- Newbler
- NGS Backbone

Used
- Oases
- SOAPdenovo
- MIRA
- Euler-SR
- Sma3
- MREPS
- Bowtie2
- CAP3
- CD-HIT

Newly developed
- SeqTrimNext
- Full-LengthNext
- MsaAssemblyReject

Evaluation of assemblies

<table>
<thead>
<tr>
<th>Tissue A</th>
<th>Tissue B</th>
<th>Tissue A + B</th>
</tr>
</thead>
<tbody>
<tr>
<td>k-mer 25</td>
<td>k-mer 29</td>
<td>k-mer 25</td>
</tr>
<tr>
<td>12,523</td>
<td>5,732</td>
<td>2,114</td>
</tr>
<tr>
<td>28,498</td>
<td>11,214</td>
<td>3,792</td>
</tr>
<tr>
<td>44,837</td>
<td>15,307</td>
<td>6,852</td>
</tr>
<tr>
<td>0.5511</td>
<td>0.6175</td>
<td>0.6781</td>
</tr>
<tr>
<td>0.5384</td>
<td>0.5777</td>
<td>0.5777</td>
</tr>
<tr>
<td>0.5977</td>
<td>0.6476</td>
<td>0.6476</td>
</tr>
<tr>
<td>0.5511</td>
<td>0.6476</td>
<td>0.6476</td>
</tr>
<tr>
<td>0.5384</td>
<td>0.6476</td>
<td>0.6476</td>
</tr>
</tbody>
</table>

Better assemblies

- MIRA
- MIRA3
- Euler-SR
- Sma3
- MREPS
- Bowtie2
- CAP3
- CD-HIT

References

- Benzekri H et al. (2014) De novo assembly characterization and functional annotation of Senegalia solea (Solea senegalensis) and common sole (Solea solea) transcriptomes. Integration in a database and design of a microarray. BMC Genomics in press.

Some databases constructed using these workflows

- SustainPineDB
- Plant Biotechnol J
- CABOG