Introgression Activity
Introgression Activity
The multi-species coalescent model
The multi-species coalescent model

Milan Malinsky 2022
The multi-species coalescent model
The multi-species coalescent model
Incomplete lineage sorting
Incomplete lineage sorting

Time (generations)

Milan Malinsky 2022
Incomplete lineage sorting

Time (generations)

Mutation

Milan Malinsky 2022
Incomplete lineage sorting
Incomplete lineage sorting

Milan Malinsky 2022
Introgression

\[ D = \frac{nABBA - nBABA}{nABBA + nBABA} \]
Introgression

\[ D = \frac{n_{ABBA} - n_{BABA}}{n_{ABBA} + n_{BABA}} \]
**Introgression**

![Diagram showing introgression process]

**Formula:**

\[ D = \frac{nABBA - nBABABA}{nABBA + nBABABA} \]
Introgression

\[ D = \frac{nABBA - nBABA}{nABBA + nBABA} \]

\[ nABBA > nBABA \]
Introgression

\[ D = \frac{nABBA - nBABA}{nABBA + nBABA} \]

\[ nABBA > nBABA \]

\[ D > 0 \]
Introgression

LETTER

Butterfly genome reveals promiscuous exchange of mimicry adaptations among species

The Heliconius Genome Consortium*
Activity:
Inferring gene flow in simulated data

Simulated data (VCF):
- 20 species in 1 million generations
- Ne = 50,000
- mu, rho = 1x10e-8
- 20Mb of data
Activity: Finding gene-flow in a real dataset

Model-based inference

PhyloNet
Activity: Gene flow signal along the genome

![Graph showing gene flow signal along the genome with different annotations and coordinates.]

- Accessible genome
- Assembly gaps
- Gene annotation: RH2Aβ, RH2Aα, RH2B

**Graph details:**
- **x-axis:** Scaffold_18 coordinates
- **y-axis:** $f_{\text{tot}}$
- **Legend:**
  - Blue line: Gene flow signal
  - Black bars: Assembly gaps
  - Green markers: Gene annotations
Activity: Gene flow signal along the genome
Activity: Results
Introgression Activity

True tree, no introgression
Introgression Activity

$$D = \frac{nABBA - nBABA}{nABBA + nBABA}$$

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>Dstatistic</th>
<th>Z.score</th>
<th>p.value</th>
<th>f4.ratio</th>
<th>BBAA</th>
<th>ABBA</th>
<th>BABA</th>
</tr>
</thead>
<tbody>
<tr>
<td>171</td>
<td>S18</td>
<td>S19</td>
<td>S00</td>
<td>1.000000</td>
<td>3.01176e-314</td>
<td>NaN</td>
<td>7.33365e-06</td>
<td>179922</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>324</td>
<td>S18</td>
<td>S19</td>
<td>S01</td>
<td>1.000000</td>
<td>3.01176e-314</td>
<td>NaN</td>
<td>7.33100e-06</td>
<td>179814</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>460</td>
<td>S18</td>
<td>S19</td>
<td>S02</td>
<td>1.000000</td>
<td>3.01176e-314</td>
<td>NaN</td>
<td>7.31256e-06</td>
<td>179800</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>580</td>
<td>S18</td>
<td>S19</td>
<td>S03</td>
<td>1.000000</td>
<td>3.01176e-314</td>
<td>NaN</td>
<td>7.34992e-06</td>
<td>179778</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>685</td>
<td>S18</td>
<td>S19</td>
<td>S04</td>
<td>1.000000</td>
<td>3.01176e-314</td>
<td>NaN</td>
<td>7.34833e-06</td>
<td>179765</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>690</td>
<td>S06</td>
<td>S05</td>
<td>S11</td>
<td>0.833333</td>
<td>2.91667e+00</td>
<td>0.000353794</td>
<td>4.88206e-05</td>
<td>138120</td>
<td>11.0</td>
<td>1</td>
</tr>
<tr>
<td>776</td>
<td>S18</td>
<td>S19</td>
<td>S05</td>
<td>1.000000</td>
<td>3.01176e-314</td>
<td>NaN</td>
<td>7.34066e-06</td>
<td>179768</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>854</td>
<td>S18</td>
<td>S19</td>
<td>S06</td>
<td>1.000000</td>
<td>3.01176e-314</td>
<td>NaN</td>
<td>7.35210e-06</td>
<td>179777</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>920</td>
<td>S18</td>
<td>S19</td>
<td>S07</td>
<td>1.000000</td>
<td>3.01176e-314</td>
<td>NaN</td>
<td>7.33584e-06</td>
<td>179771</td>
<td>1.5</td>
<td>0</td>
</tr>
</tbody>
</table>

True tree, no introgression
The multi-species coalescent model
Introgression Activity

True tree, no introgression
**Introgression Activity**

Three significant values after multiple testing correction
FDR=5%

True tree, no introgression
Introgression Activity

1140 tests reduced to 400 by aggregating over P1

True tree, no introgression
Introgression Activity

True tree, five introgression events

outgroup
Introgression Activity

True tree, five introgression events
Introgression Activity

True tree, five introgression events

Dstatistic 0.805219
Z-score p-value f4-ratio 72.7811 2.3e-16 0.0950458
Introgression Activity

True tree, five introgression events

outgroup

S00
S01
S02
S03
S04
S05
S06
S07
S08
S09
S10
S11
S12
S13
S14
S15
S16
S17
S18
S19

Dmin
123
1
1016
85
39
BBAA

14%
16%
18%
13%
8%
Introgression Activity

True tree, five introgression events

D-statistic

f4-ratio
True tree, five introgression events

modi- fied from Ottenburghs (2019) Avian Research
3. Finding geneflow in a real dataset - Tanganyikan cichlids
3. Finding geneflow in a real dataset - Tanganyikan cichlids
3. Finding geneflow in a real dataset - Tanganyikan cichlids
3. Finding geneflow in a real dataset - Tanganyikan cichlids