How to present results of genomic studies in an intelligible form

Henner Simianer
Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August-University Göttingen, Germany
Purpose of a scientific talk at a conference

- Convey a message to the audience
- The audience should be able to understand your reasoning
- Provide evidence that your empirical results are sound and reliable
- The audience should be convinced by your arguments
Reports on an empirical study

- Description of your experiment should ideally enable others to repeat the experiment
- Due to restrictions (mainly in time) often not fully achievable
- But still try to be as precise and comprehensive as possible
Reports on an empirical study

Step 1: State your research question
Ideally as formal hypothesis

If you want to test whether a certain factor F (a marker, a gene, a breed) has an effect on a variable y (growth, health):

$H_0$: F does not affect y

$H_1$: F does affect y (one-sided or two sided)
Reports on an empirical study

Step 2: Describe your experiment

„Material“ and methods

Provide all relevant information on the substrate you worked with (animals, tissues, technologies) in your experiment

Describe the actual experiment (what was done, how often, when were samples taken etc.)

Describe editing and quality control of the data (outlier detection, filtering criteria, imputation etc.)
Reports on an empirical study

**Step 2:** Describe your experiment

„Material“ and methods

Describe the statistical approaches you use

Just naming a software (we used SAS/R ...) is **not** sufficient.

Provide the complete models you have used (including the relevant assumptions)

Describe in such detail that a knowledgeable person can follow
Direct model ⇔ Direct-Social model

\[ A_D \]
\[ \sigma^2_{A_D} \]

\[ A_T = A_D + 3A_S \]
\[ \sigma^2_{A_T} = \sigma^2_{A_D} + 6\sigma^2_{A_{DS}} + 9\sigma^2_{A_S} \]

Peeters et al., EAAP 2013
Reports on an empirical study

Step 3: Present your results

Graphs are much more intuitive than tables – the human brain is analog
**Experiment I:** which breed x treatment combination has the lowest performance?  
3 seconds

<table>
<thead>
<tr>
<th></th>
<th>Treatment A</th>
<th>Treatment B</th>
<th>Treatment C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breed 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breed 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breed 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breed 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Experiment I:** which breed x treatment combination has the lowest performance? 3 seconds

<table>
<thead>
<tr>
<th>Breed</th>
<th>Treatment A</th>
<th>Treatment B</th>
<th>Treatment C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.3 ± 0.2</td>
<td>2.4 ± 0.1</td>
<td>2.5 ± 0.1</td>
</tr>
<tr>
<td>2</td>
<td>2.5 ± 0.1</td>
<td>4.4 ± 0.2</td>
<td>3.3 ± 0.2</td>
</tr>
<tr>
<td>3</td>
<td>3.5 ± 0.3</td>
<td>2.1 ± 0.4</td>
<td>2.3 ± 0.2</td>
</tr>
<tr>
<td>4</td>
<td>4.5 ± 0.2</td>
<td>2.8 ± 0.2</td>
<td>4.7 ± 0.3</td>
</tr>
</tbody>
</table>
**Experiment I:** which breed x treatment combination has the lowest performance?

![Bar chart showing performance comparisons for different breeds and treatments.](chart-url)
**Experiment II:** which breeds react similar to the three treatments?

6 seconds

<table>
<thead>
<tr>
<th>Breed</th>
<th>Treatment A</th>
<th>Treatment B</th>
<th>Treatment C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breed 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breed 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breed 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breed 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Experiment II: which breeds react similar to the three treatments?

6 seconds

<table>
<thead>
<tr>
<th>Breed</th>
<th>Treatment A</th>
<th>Treatment B</th>
<th>Treatment C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breed 1</td>
<td>4,3 ± 0,2</td>
<td>2,4 ± 0,1</td>
<td>2,5 ± 0,1</td>
</tr>
<tr>
<td>Breed 2</td>
<td>2,5 ± 0,1</td>
<td>4,4 ± 0,2</td>
<td>3,3 ± 0,2</td>
</tr>
<tr>
<td>Breed 3</td>
<td>3,5 ± 0,3</td>
<td>2,1 ± 0,4</td>
<td>2,3 ± 0,2</td>
</tr>
<tr>
<td>Breed 4</td>
<td>4,5 ± 0,2</td>
<td>2,8 ± 0,2</td>
<td>4,7 ± 0,3</td>
</tr>
</tbody>
</table>
**Experiment II:** which breeds react similar to the three treatments?
Experiment II: which breeds react similar to the three treatments?
Experiment II: which breeds react similar to the three treatments?
Reports on an empirical study

**Step 3:** Present your results

Report your results with some indication of the precision of your estimates

- standard errors
- confidence intervals
- posterior distributions
<table>
<thead>
<tr>
<th>Weight (kg)</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Box plots are a good way of characterising the entire distribution.
You can also show the ‘confidence band’ around an estimated (non-linear) regression.
Bayesian statistics provide posterior distributions of estimated quantities
Reports on an empirical study

Step 3: Present your results

Report the statistical significance of your results

- You have stated earlier which hypotheses were tested and which statistical tests were used

- Report p-values or 'usual' error levels
  (* = 0.05; ** = 0.01; *** = 0.001)
Account for multiple testing

Nominal error level $\alpha = 0.05 (*)$ means that under the $H_0$ (no effect) 1 out of 20 tests finds an effect.

$N = 1000$ tests $\rightarrow$ 50 'significant' effects even under the $H_0$

- **Bonferroni (1935) correction**: use in each test the test level $\alpha / N$; the global error probability is kept but testing is very (too) conservative – improved versions e.g. by Holm (1979) and Hochberg (1988)

- **False Discovery Rate (Benjamini and Hochberg, 1995)**: Test such, that a proportion $\alpha$ of the significant results are false positive

- **Permutation test (Doerge and Churchill, 1996)**: generate the distribution of the test statistic under $H_0$ through permutation of the data.
Genome-wide vs. chromosome-wise testing

With multiple testing the power decreases with the number of tests

- If you can *a priori* restrict the location of a QTL to a subregion (e.g. a chromosome), then chromosome-wise testing is legitimate

- Otherwise you always have to account for all tests you are doing → only genome-wide results should be communicated
Reports on an empirical study

Step 3: Present your results – some basics

- state clearly what is shown on a slide
- put titles on axes in sufficiently large and readable fonts
- add a legend, if necessary
- Use colours and symbols in a systematic way
- Assign colours, symbols, and line types consistently across slides to the same object
Figures – which type?

- Yield (kg) for wheat and barley over months May, June, July, and August.
- Year 1 and Year 2 data shown.

Legend:
- Blue line and dots: wheat - yr 1
- Green line and stars: barley - yr 1
- Red diamond: wheat - yr 2
- Pink star: barley - yr 2
Figures – which type?

- Wheat - yr 1
- Barley - yr 1
- Wheat - yr 2
- Barley - yr 2

Yield (kg)

Month: May, June, July, August

- Red: Wheat - yr 1
- Blue: Barley - yr 1
- Open Red: Wheat - yr 2
- Open Blue: Barley - yr 2
Reports on an empirical study

Step 3: Present your results

Be creative!

Avoid default settings of standard software

Make use of the great opportunities of modern statistics and graphics programs

Be inspired by what the leaders in the field are doing

But: Fanciness should not be at the expense of clarity
Brawand et al., 2014 Nature

![Graph showing evolutionary relationships and genetic divergence](image)

- **Genealogy 0 (1)**
  - Mz
  - Pn
  - Ab

- **Genealogy 2**
  - Mz
  - Pn
  - Ab

- **Genealogy 3**
  - Mz
  - Pn
  - Ab

- **Divergence time**
  - 45.2% (11.5%)
  - 20.4%
  - 22.6%

- **T_{mb}**
- **T_{m}**

**Figure b:**
- **Species pairs:***P. nyororoi vs. P. punxamila***
- **F_{ST}**
  - 0, 2, 000, 4, 000, 6, 000, 8, 000, 10, 000

**Figure c:**
- **Genomic divergence**
- **All sites**
- **Ancient variant sites**
- **F_{ST} threshold**

**Color-coding:**
- Intron
- Exon
- 25-kb flank
- None
Groenen et al., 2014 Nature
Final remarks

<table>
<thead>
<tr>
<th></th>
<th>Poor presentation</th>
<th>Good presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor science</td>
<td>![Sad Face]</td>
<td>![Sad Face]</td>
</tr>
<tr>
<td>Good science</td>
<td>![Sad Face]</td>
<td>![Happy Face]</td>
</tr>
</tbody>
</table>

It is worth spending some thought and time on preparing the **perfect** presentation of your research.
Learn the rules so you know how to break them properly

Thank you!
Experiment II: which breeds react similar to the three treatments?