breedR: statistical methods for forest genetic resources analysis

breedR workshop
Jaca - Spain June 30th - July 3rd 2015
Plan

• **Why** do we need this?
• What are the **objectives**?
• What do we **need**?
• Where we are **now**?
• What are the **difficulties**?
• demo
Why do we need this?

- Statistical modelling: preliminary key step when assessing forest trials
- Experimental data is becoming increasingly complex and massive
Experimental data: complex & massive

Maritime pine pedigree

Douglas-fir data, INRA

[Designing Trees for the Future]
Why do we need this?

- Statistical modelling: preliminary key step when assessing forest trials
- Experimental data is becoming increasingly complex and massive
- Appropriate statistical tools are not an easy choice: daunting start for beginners / source of heterogeneity among advanced users
- Need of interchange & expertise & sharing in statistical issues
What are the objectives?

- Seek a **consensual starting seed package** comprising essential and novel statistical tools that are relevant to the community of forest genetics and breeding.

- Develop a **collaborative effort** to make these tools reliable, readily available and conceptually accessible to the community.

- Make all developments from publicly accessible open-source components and within a framework that allows for **complete openness** to new additions by the community of users.

- Foster the **interchange and expertise** on statistical issues within the consortium and beyond.
What do we need?

1. Elaborate a development plan, a road-map

- essential toolkit
  - Spatial statistics
  - G×E analysis
  - Mixed-model framework
  - Pedigree-free evaluation
  - Competition evaluation

- estimation of causal (co)variance components

- prediction of individual genetic values

- computationally demanding tasks in FORTRAN

- access to databases (WP1 and WP2)

- population/individual estimates (to meta-modeling, to WP3)
What do we need?

1. Elaborate a development plan, a road-map
2. Identify and test appropriate solutions

<table>
<thead>
<tr>
<th>Module</th>
<th>Solution</th>
<th>Experts</th>
<th>Degree of accessibility and development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed-model</td>
<td>BLUPF90 publicly available suite of 11 main programs, including several REML procedures and several Bayesian Gibbs samplers for solving general mixed-models. Briefly described in Misztal (2008) and downloadable from [1]</td>
<td>External: Ignacy Misztal, Jon Hallander and Luis Varona; internal: already in use for in-house simulation purposes at INRA</td>
<td>Readily available as FORTRAN90 codes and usable as standalone programs via the appropriate compilation. Some coding is expected and concerns mainly the input and output interfaces with users, among the different codes and with other modules</td>
</tr>
<tr>
<td>SS</td>
<td>Collection of independent programs for various phases of spatial analyses and developed by Eduardo Cappa and collaborators, and made available to T4F. Firstly described in Cappa and Cantet (2007)</td>
<td>External: Eduardo Cappa</td>
<td>Readily available as FORTRAN77 codes and usable as standalone programs via the appropriate compilation. Some substantial coding is needed in order for it to interface with the mixed-model module</td>
</tr>
<tr>
<td>GxE analyses</td>
<td>Module published for the R platform of statistical programming: Agricolae (downloadable from [2]) and random model alternatives as described by Smith et al (2001) and Piepho (1997)</td>
<td>Internal: Catherine Bastien (INRA)</td>
<td>Available as R module, readily usable and modifiable within R. Some substantial coding is needed in order for it to interface with the mixed-model module</td>
</tr>
</tbody>
</table>

---

**PFEM**

Published procedures for the estimation of realized genetic gains in the Mating Program for the extension of efficiency in the Breeding Program (EPE), and improving the exploitation of genetic variation. The coding is needed for incorporating published protocols.
What do we need?

1. Elaborate a development plan, a road-map
2. Identify and test appropriate solutions
3. Organize a community for development/testing/training/dissemination

• **recruitment of a mathematician / developer**
  September 2013 (up to end of T4F): development and integration of the different programs with help from local INRA team

• **involving external experts into our development**
  Up to 6 external experts with gradual involvements from late-2013 onwards

• **setting-up a network of beta-testers**
  Mostly belonging to T4F consortium: 13 people, 12 participants & up to 32MM + several volunteers without MM + several users outside T4F
Where we are now?

- Programming and development

- Beta-testing (from M31 to final release M42)

- Training, dissemination and updating
  one international workshop and national follow-ups, repositories, wiki-documents, scientific publications using the tool
What are the difficulties?

• Development is not a well paved way... but we are succeeding
• We rely very much on our own local circle of testers... but they tend to see things like us
• We desperately need feedback from (T4F) beta-testing

Beta-test NOW!
Where to get BreedR?

- BreedR general web page: http://famuvie.github.io/breedR/
- BreedR Forge: https://github.com/famuvie/breedR
Thanks!