

Title : **Internship proposal for a Master's student in
Quantitative/Statistical Genetics**

Time of the training period : second semester of the 2013-2014 academic year

Keywords : Genomic Selection, Ridge Regression-Best Linear Unbiased Predictors, Identity By Descent, model accuracy

The topic :

CONTEXT :

Partial diallels or merely sets of biparental crosses are the standard calibration populations for Genomic Selection in maize at Limagrain Europe. In comparison with using a calibration model just built on a biparental population, that format allows to :

- increase our calibration population size
- increase our genetic base
- possibly combine data from several years (different sets of crosses, seen over one year each)

The same holds true with synthetic populations (such as MAGIC).

However, when moving from stand alone biparental crosses to sets thereof or to synthetics :

- allele diversity increases
- marker-gene linkage phase may change from one cross to the other because there is only small LD amongst the parents of these populations, even on short chromosome distances, with multi-K marker chips.

Hence, fitting SNP marker data as is, i.e. with only 2 alleles per locus, may be judged, *a priori*, as an unsatisfactory solution.

On the other hand, multi-allelic loci would be more likely to present one allele that is in good LD with a causal mutation in a nearby gene, even more so if the alleles are identified according to (i.e. named after) the founders that supplied the chromosome fragment.

Statistical approaches and informatics tools were developed in-house to infer Identity By Descent likelihoods of current maize lines all along their genome relative to a set of founders.

THE TOPIC :

The **overall** project's principle would consist in taking these IBD likelihoods into our GS models.

We have first concrete leads (*the student would not be left on unexplored ground with a risk of no working solution to the problem asked*) that would consist, for example, in implementing an upgrade in R of the RR-BLUP close form solution to the fitted marker effects. Students may consult Hickey *et al.* (2012) to get a flavor of the topic. Compared to the authors, we would pursue a solution, although different, in the same family of methods nevertheless.

Nature of the student's work :

- The nature of the student's work would be to develop further the equations of the models and to program them in R.
- The routines would then be applied on real data by the student to assess the method through cross validation.
- Only if time allows, simulation work will be performed also, in order to assess the method more accurately.

Available in-house expertise to support and train the student :

Statistical, genetics and programming expertise is available in Zivan Karaman's team and with Claude Lebreton. There already exists strong, concrete leads, as presented above, for the projects.

Anticipated benefit for Limagrain:

The anticipated benefit of that work to Limagrain lies in the possible increase in our GS models accuracy and thereby in our marker assisted selection efficiency through a bigger genetic gain for a given selection pressure.

location :

Rustenhart corn breeding station (in Elsass, near Colmar), with Claude Lebreton, with frequent trips to Limagrain Chappes Research Center.

Related literature :

J. M. Hickey, B. P. Kinghorn, B. Tier, S. A. Clark, J. H. J. van der Werf & G. Gorjanc **2012** Genomic evaluations using similarity between haplotypes, J. Anim. Breed. Genet. 1–11

Person to contact :

Claude Lebreton

Claude.lebreton@limagrain.com
Limagrain Europe
8 Rheinfelderhof
68 740 RUSTENHART
FRANCE

Tél : + 33 3 89 62 40 25
mobile : + 33 6 31 60 64 25
Fax : + 33 3 89 62 40 29