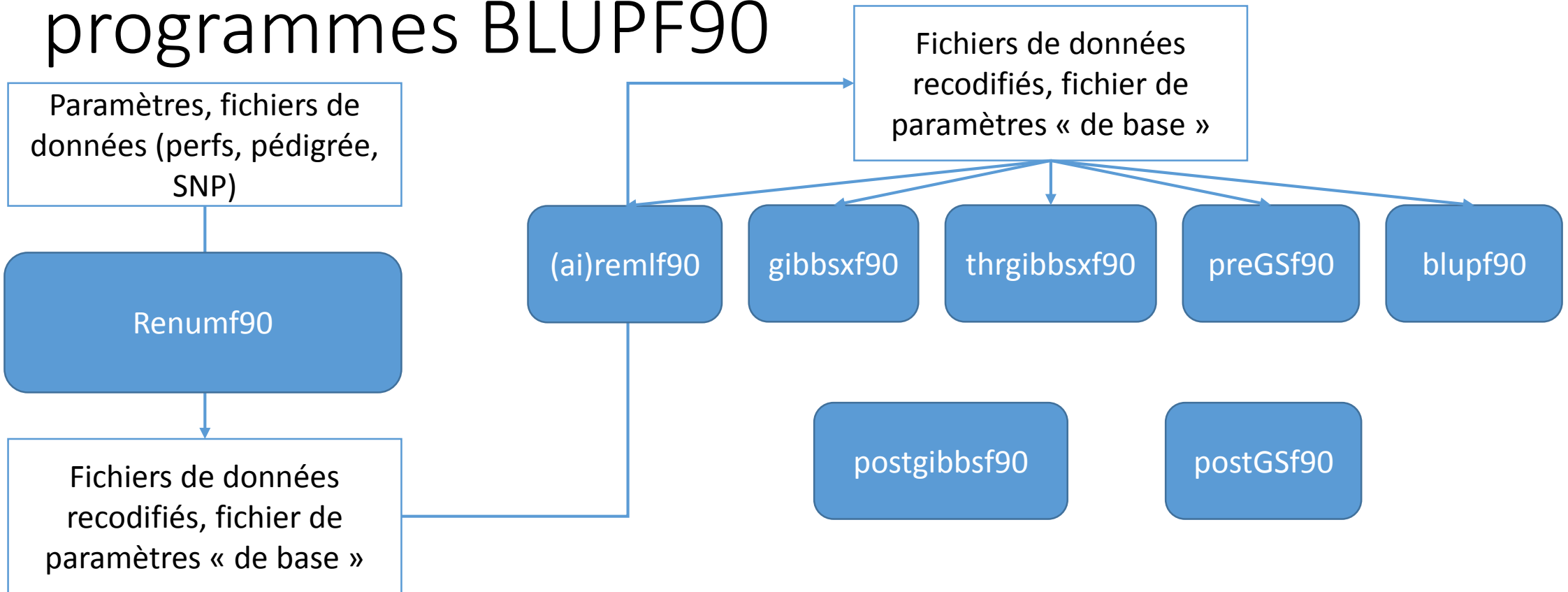


# Présentation de BLUPF90

# Architecture générale de la suite de programmes BLUPF90



# RENUMF90

- **L'intérêt** : recode les fichiers de données comme il faut, crée un fichier de paramètres général => juste des options à rajouter
- **Fichiers d'entrée (séparateurs espace)** :
  - **Performances** : animal, performances, effets (fixes, aléatoires, coraviables). Par défaut, valeur manquante = 0.
  - **Pédigrée** : animal, père, mère, éventuellement année de naissance, mère porteuse. Manquants = 0 (pas 00, pas .)
  - **SNP** : animal, génotype sans espaces, codé 0, 1 ou 2 (5 pour manquants), tous les génotypes doivent démarrer à la même position

# RENUMF90 – Fichier de paramètres

DATAFILE

./perf.txt

Nom et adresse du fichier de performances, 20 caractères max par défaut

TRAITS

Colonnes du fichier de performances dans lesquelles les caractères sont présents

234

FIELDS\_PASSED TO OUTPUT

Colonnes du fichier de performances qu'on veut transférer sans les modifier dans le fichier de données recodifié

1

WEIGHT(S)

Colonnes des poids accordés aux animaux, il n'est pas possible d'avoir des colonnes différentes pour différents caractères

RESIDUAL\_VARIANCE

949.12 236.34 14.236

236.34 797.70 1.9502

14.236 1.9502 0.70225

Matrice de (co)variances résiduelles

EFFECT

Définition des effets :

555 cross alpha

Colonnes contenant l'effet pour chacun des caractères (0 si un effet ne s'applique pas à l'un des caractères)

EFFECT

Cross ou cov, selon si l'effet est un effet à niveau ou une variable continue

666 cross alpha

Alpha ou numer (si cross) selon si la valeur de l'effet contient des caractères (alpha) ou uniquement des chiffres.

Répéter les deux lignes autant qu'il y a d'effets (hors effets aléatoires)

# RENUMF90 – Fichier de paramètres

EFFECT	Pour les effets aléatoires : même définition que les autres effets, immédiatement suivie de RANDOM puis du type d'effet aléatoire (animal, sire, diagonal)
1 1 1 cross alpha	
RANDOM	
animal	
FILE	Si effet animal ou sire, suivi immédiatement du fichier de pédigrée et de son adresse, 20 caractères max par défaut
./pedigree.txt	
FILE_POS	Définition des colonnes du fichier de pédigrée : colonne animal, colonne père, colonne mère, colonne mère porteuse, colonne année de naissance. Si aucune position indiquée, le défaut est animal, père, mère
1 2 3 0 0	
SNP_FILE	Si on veut utiliser une information génomique, utiliser les lignes SNP_FILE et adresse du fichier contenant les SNP, 20 caractères max par défaut. Si on veut faire un BLUP, supprimer ces lignes.
./snpfilessgblup	
PED_DEPTH	
3	Profondeur de pédigrée remontée (défaut : 3, si 0 remonte tout)
(CO)VARIANCES	
215.52 -76.513 3.4841	Matrice de (co)variance génétique
-76.513 81.825 -1.0122	
3.4841 -1.0122 0.20737	

# RENUMF90 – Fichier de paramètres

- Les effets d'environnement permanent, maternel, ...
  - A ajouter entre « RANDOM animal » et « FILE fichier de pedigree »
  - Lignes : OPTIONAL
    - pe, mat et/ou mpe : ajoute un effet corrélé à l'effet animal
- Les groupes de parents inconnus : indiquer UPG\_TYPE après PED\_DEPTH nombre de générations remontées, puis
  - job suivi d'une liste d'années => définit des groupes de parents inconnus selon les dates de naissance présentes dans le fichier de pedigree
  - in\_pedigrees : les parents inconnus sont déjà indiqués dans le fichier de pedigree sous la forme -X
- L'option alpha\_size suivi d'un nombre donne la longueur maximale des adresses des fichiers (maximum 50) : OPTION alpha\_size 50

# RENUMF90 – Les sorties

- **renf90.dat** : fichier de performances recodifié. Contient, dans l'ordre :
  - Les caractères
  - Les effets recodifiés (dans l'ordre de leur définition)
  - Les colonnes non modifiées (fields\_passed to output)
- **renaddXX.ped** : fichier de pédigrée recodifié. XX = nombre d'effets. Contient, dans l'ordre :
  - Animal recodifié
  - Père recodifié
  - Mère recodifiée
  - 3-nombre de parents connus
  - Année de naissance
  - Nombre de parents connus (si animal non génotypés), 10+nombre de parents connus (si animal génotypé)
  - Nombre de lignes de performances pour l'animal
  - Nombre de descendants en tant que père
  - Nombre de descendants en tant que mère
  - Identifiant animal non recodifié

# RENUMF90 – Les sorties

- **renf90.tables** : recodification des effets. Pour chaque effet,

- Valeur d'origine
- Nombre d'animaux dans l'effet
- Valeur recodifiée

```
Effect group 3 of column 1 with 4 levels
Value # consecutive number
1 2 1
3 2 2
5 1 3
8 1 4
Effect group 4 of column 1 with 3 levels
Value # consecutive number
xx 2 1
yy 2 2
zz 2 3
```

- **nomfichierSNP\_XrefID** : fichier associant identifiant recodifié – animal pour les animaux génotypés.
  - Animal recodifié
  - Numéro animal
- **renf90.par** : le fichier de paramètres général pour d'autres programmes



# renf90.par : fichier de paramètres général

DATAFILE	renf90.dat	Adresse du fichier de données recodifié
NUMBER_OF_TRAITS	3	Nombre de caractères
NUMBER_OF_EFFECTS	4	Nombre d'effets
OBSERVATION(S)	1 2 3	Colonnes des caractères
WEIGHT(S)		Colonnes des poids
EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT[EFFECT NESTED]		Liste des effets, avec, pour chaque effet : les colonnes où ils se trouvent pour chacun des caractères, le nombre de niveaux, le type d'effet (à niveau ou covariable)
	4 4 4 12887 cross	
	5 5 5 3 cross	
	6 6 6 10 cross	
	7 7 7 284350 cross	

# renf90.par : fichier de paramètres général

RANDOM\_RESIDUAL VALUES

949.12	236.34	14.236
236.34	797.70	1.9502
14.236	1.9502	0.70225

RANDOM\_GROUP

4

RANDOM\_TYPE

add\_animal

FILE

renadd04.ped

(CO)VARIANCES

215.52	-76.513	3.4841
-76.513	81.825	-1.0122
3.4841	-1.0122	0.20737

OPTION SNP\_file ./snpfilessgblup

OPTION missing -9999

OPTION sol se

Matrice de (co)variances résiduelles

Numéro de l'effet aléatoire dans la liste des effets

(il ne s'agit donc plus de la colonne de l'effet)

Type d'effet aléatoire (add\_animal, add\_sire, add\_an\_upg, diagonal...)

Adresse du fichier de pédigrée recodifié

Matrice de (co)variances génétiques

Options : celles présentes dans le fichier de paramètres de renum sont automatique incluses, d'autres peuvent être rajoutées

# BLUPF90

- Fichiers d'entrée (séparateurs espace) :
  - Performances recodifiées (. est une valeur)
  - Pédigrée recodifié (format renadd)
  - Fichier de SNP + fichier associant identifiant recodifié – identifiant initial nommé nomfichierSNP\_XrefID
- Fichier de sortie (séparateurs espace) : solutions. Contient, dans l'ordre :
  - Caractère
  - Numéro de l'effet
  - Niveau de l'effet
  - Solution
  - Erreur standard, si option sol se

# BLUPF90 – Les possibilités

- Faire un BLUP : définir les effets, fournir les fichiers d'entrée, ne pas mettre l'option SNP\_file
- Faire un Single-Step : exactement pareil, avec l'option SNP\_file
- Faire un GBLUP : exactement pareil qu'un single-step, avec uniquement des animaux génotypés dans le pédigrée
  
- Quelques options pratiques
  - missing suivi d'un nombre : définit les valeurs manquantes
  - sol se : stocker les erreurs résiduelles
  - saveG, saveGInverse, saveA22... => sauvegarder les différentes matrices, pour pouvoir les utiliser par la suite
  - readG, readGinverse, ou RANDOM\_TYPE user\_file => utiliser une matrice de parenté particulière
  - tunedG => définir la façon de mettre G et A22 sur la même échelle
  - Diverses options pour modifier G
  - Possibilité, par PostGSf90, de calculer les effets des SNP, puis de refaire tourner BLUPF90 en donnant des poids variables aux SNP

# Commande de lancement

- `echo renf90.par | repertoireprogramme/blupf90 | tee log_ssgblupf90`

# Quelques conseils

- Vérifiez soigneusement vos options : il n'y a pas toujours de message d'erreur quand elles ne sont pas reconnues par le logiciel
- En cas de gros jeu de données sur lesquels vous voulez faire plusieurs analyses, vous pouvez faire tourner une fois preGS, sauvegarder les matrices, puis les lire dans vos différents essais (pas de contrôle qualité, pas de création de G...)

# Chaine en développement

- Recode : imposer une taille de cellule minimale, un certain nombre de descendants par père...
- Mise en forme des fichiers de performance et de pédigrée en sortie de recode
- Récupération des phases, extraction des phases des animaux dans le fichier de pédigrée de sortie de recode, préparation du fichier de SNP
- Renumf90 : recodification des fichiers
- BLUPF90 : BLUP/Single-Step/GBLUP...

=> les options ne sont pas encore paramétrées, il faut un paramètre du type de calcul (génomique/polygénique), ajouter une étape de mise en forme des données

=> supporte des modèles uni et multicaractères, la prise en compte de poids, la présence ou non d'effets fixes...

# Quelques adresses utiles

- Vous pouvez trouver les exécutables (téléchargés en octobre 2015) dans `/g2b/icroue/prog/blupf90`
- Il y a tout un tas de documentation sur BLUPF90 là : <http://nce.ads.uga.edu/wiki/doku.php>. Vous y trouverez notamment :
  - les infos sur le cours de Misztal de 2016 => [http://nce.ads.uga.edu/wiki/doku.php?id=course\\_information\\_-\\_uga\\_2016](http://nce.ads.uga.edu/wiki/doku.php?id=course_information_-_uga_2016)
  - Le manuel de BLUPF90 : [http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90\\_all2.pdf](http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf)
  - Des documentations plus détaillées que le manuel sur chacun des programmes : [http://nce.ads.uga.edu/wiki/doku.php?id=application\\_programs](http://nce.ads.uga.edu/wiki/doku.php?id=application_programs)



# Les sessions de formation de la Miszta team

- Cours sur la programmation en Fortran et l'évaluation génétique en 3 semaines ce printemps, à UGA.
- Semaine évaluation génétique :
  - Théorie les matinées
    - Modèles d'évaluation (unicaractère, multicaractère, groupes de parents inconnus, régression aléatoire, effets aléatoires maternels, croisement...)
    - Estimation des paramètres génétiques
    - Théorie du single-step (différence par rapport à la génomique « classique », intérêt, définition et inversion de H), présentation d'APY
    - Les différentes matrices de parenté génomique possibles
    - Le contrôle qualité des données génomiques
    - Techniques de validation
    - Quelques points sur la mise en place du single-step en pratique
    - Présentation de la suite de programmes, surtout de renum et blupf90
  - TD l'après-midi pour prendre en main le logiciel