postGSf90 - ssGWAS

Ignacio Aguilar
Instituto Nacional de Investigación Agropecuaria
INIA Las Brujas, Uruguay
iaguilar@inia.org.uy

Equivalent Model

VanRanden et al 2009; Goddard, 2009; Habier el al 2007

Model that estimate SNPs effects

$$y = \mu + Za + e, \qquad \mathrm{var}(a) = D\sigma_a^2 \qquad u = Za$$

Model that estimate Breeding Values

$$y = \mu + u + e, \qquad \mathrm{var}(u) = G\sigma_u^2, \quad G = ZDZ'/k$$

Genomic Information ≈ genomic relationship

Simple conversion between :

Breeding values and SNP effects

$$u = Za$$

$$a = DZ'(ZDZ')^{-1}u$$

Stranden & Garrick, 2009

Equivalent Model

- SNP effects from GEBV's (Henderson, 1973; Strandén and Garrick, 2009):

$$\hat{u} = \frac{\sigma_u^2}{\sigma_a^2} DZ' G^{-1} \hat{a}_g = DZ'[ZDZ']^{-1} \hat{a}_g$$

Differential weight to each SNP

- Also, for each SNP effect (i-th):

$$\hat{\sigma}_{u,i}^2 = \hat{u}_i^2 2 p_i (1 - p_i)$$

Genet. Res., Camb. (2012), **94**, *pp.* 73–83.
doi:10.1017/S0016672312000274

Genome-wide association mapping including phenotypes from relatives without genotypes

H. WANG[1]*, I. MISZTAL[1], I. AGUILAR[2], A. LEGARRA[3] AND W. M. MUIR[4]
[1] *Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602-2771, USA*
[2] *Instituto Nacional de Investigación Agropecuaria, INIA Las Brujas, 90200 Canelones, Uruguay*
[3] *INRA, UR631 Station d'Amélioration Génétique des Animaux (SAGA), BP 52627, 32326 Castanet-Tolosan, France*
[4] *Department of Animal Science, Purdue University, West Lafayette, IN 47907-1151, USA*

(*Received 19 September 2011; revised 8 December 2011, and 9 March 2012; accepted 13 March 2012*)

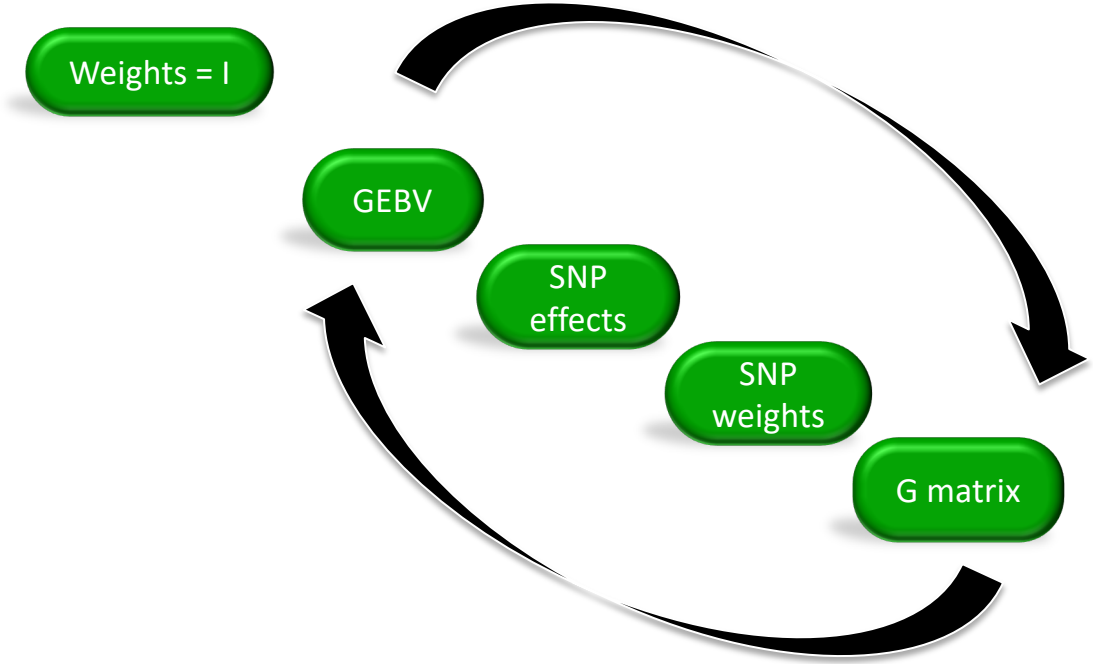
Computing algorithm

- Denote *t* as an iteration number and *i* as the *i*-th SNP

1. t=0, $D_{(t)}$=I, $G_{(t)}$=$ZD_{(t)}Z'\lambda$
2. Compute $\hat{a}_g$ by ssGBLUP
3. Calculate $\hat{u}_{(t)} = \lambda D_{(t)} Z' G_{(t)}^{-1} \hat{a}_g$
4. Calculate $d^*_{i_{(t+1)}} = \hat{u}^2_{i_{(t)}} 2p_i(1-p_i)$ for all SNPs (Zhang et al., 2010)
5. Normalize $D_{(t+1)} = \frac{tr(D_{(0)})}{tr(D^*_{(t+1)})} D^*_{(t+1)}$
6. Calculate $G_{(t+1)} = ZD_{(t+1)}Z'\lambda$
7. t=t+1
8. Exit , or loop to step 2 or 3

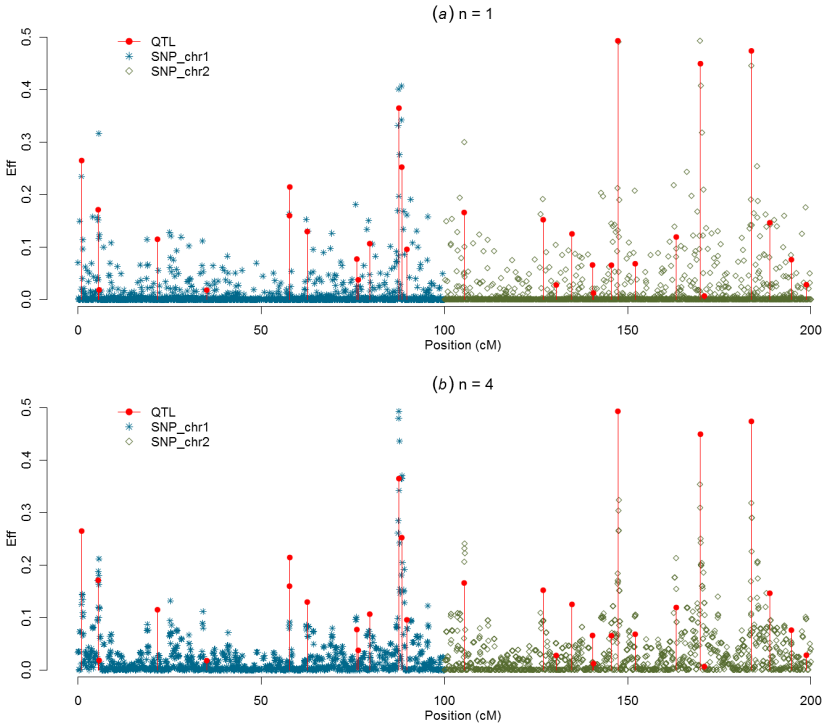
Weighted ssGBLUP

WssGBLUP(Wang et al., 2012)



- Gives more weight to important markers

Manhattan SNP effects for simulated data

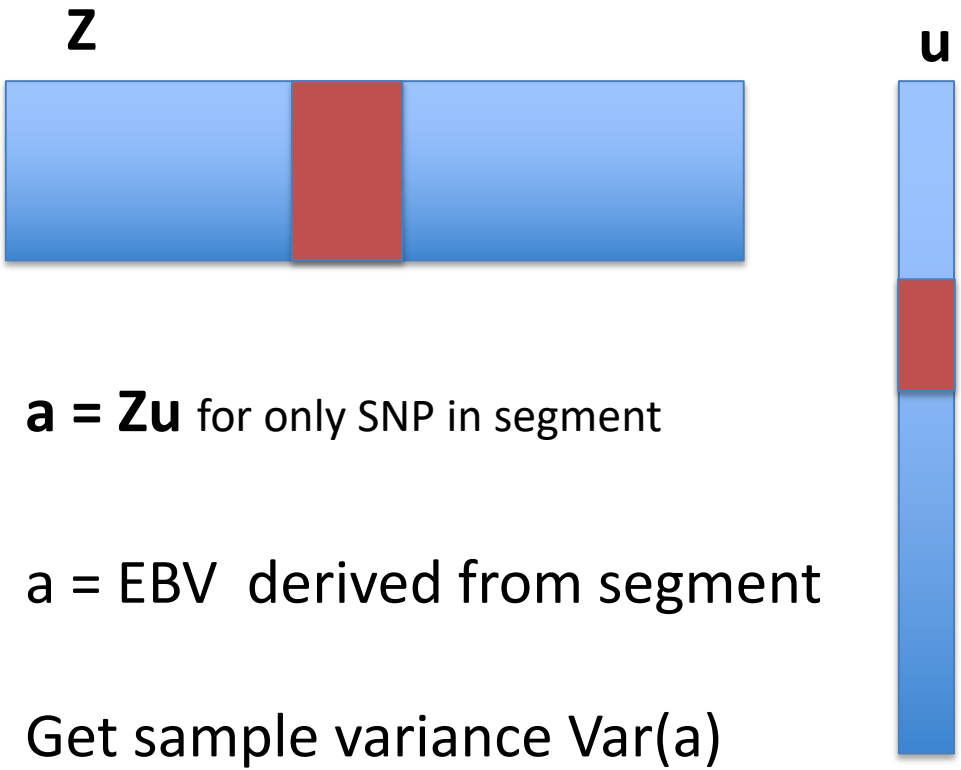


Wang et al 2012

Variances explained by segments

- ISU propose to present results from GWAS using variance explained by windows of adjacent SNP
- Fan et al 2011, Onteru et al 2011, Peters el al 2012, etc.
- Potentially use of bootstrap to get significance of detected QTL

Windows Variances



a = Zu for only SNP in segment

a = EBV derived from segment

Get sample variance Var(a) from genotyped individuals

postGSf90

- Genomic POST processing program
- Extract SNP effects from *solutions* after genomic evaluations (GLBUP and ssGLBUP)
- Calculate weights for iterative ssGWAs
- Calculate variance explained by segments

postGSf90 par files

1) Parameter files:

(1) BLUPF90 (and preGSf90)

(2) postGSf90

2) OPTIONs:

BLUPf90 / PreGSf90:

```
OPTION SNP_file marker.geno.clean
OPTION saveGInverse
OPTION weightedG w # A vector with length = M
```

postGSf90:

```
OPTION SNP_file marker.geno.clean
OPTION ReadGInverse
OPTION chrinfo mapfile #format: snpID chr pos
OPTION weightedG w
# OPTION Manhattan_plot
```

3) Document:

http://nce.ads.uga.edu/wiki/doku.php?id=readme.pregsf90#gwas_options_postgsf90

SNP map file

- OPTION chrinfo *<file>*
- For some genomic analyses (GWAS) or QC
- Format:
  - SNP number
    - Index number of SNP in the sorted map by chromosome and position
  - chromosome number
  - Position
  - SNP name (Optional)

```
1 1 135098 Hapmap43437-BTA-101873
2 1 267940 ARS-BFGL-NGS-16466
3 1 393248 Hapmap34944-BES1_Contig627_
4 1 471078 ARS-BFGL-NGS-98142
5 1 516404 Hapmap53946-rs29015852
6 1 571340 ARS-BFGL-NGS-66449
7 1 845494 ARS-BFGL-BAC-32770
8 1 883895 ARS-BFGL-NGS-65067
9 1 950841 ARS-BFGL-BAC-34682
10 1 974586 ARS-BFGL-NGS-3964
11 1 1009504 ARS-BFGL-NGS-98203
12 1 1189382 ARS-BFGL-BAC-31722
13 1 1234172 ARS-BFGL-BAC-6557
14 1 1264369 ARS-BFGL-BAC-7196
15 1 1359951 Hapmap53766-rs46526150
```

- First row corresponds to first column SNP in genotype file !!!

SNP solution output

```
snp_sol
```

contains solutions of SNP and weights

- 1: trait
- 2: effect
- 3: SNP
- 4: Chromosome
- 5: Position
- 6: SNP solution
- 7: weight

if `OPTION windows_variance` is used

- 8: variance explained by *n* adjacents SNP.

ssGWAS — less snp_sol — 96×24

1	2	1	1	0	0.6176172E-03	1.938355	0
1	2	2	1	44260	0.1606191E-03	0.6675455E-01	0
1	2	3	1	88520	0.7086239E-03	2.765576	0
1	2	4	1	132780	-0.1968899E-03	0.1616454	0
1	2	5	1	221300	-0.1728076E-03	0.1641911	0
1	2	6	1	265560	0.9542164E-04	0.4274880E-01	0
1	2	7	1	309820	0.6385818E-03	2.168658	0
1	2	8	1	354080	0.7154674E-03	2.888887	0
1	2	9	1	398340	-0.1997904E-03	0.2258326	0
1	2	10	1	442600	0.2820564E-03	0.4205221	0
1	2	11	1	531120	0.2023037E-03	0.8558627E-01	0
1	2	12	1	575380	-0.2259381E-03	0.2721033	0
1	2	13	1	619640	0.8177870E-03	3.782733	0
1	2	14	1	663900	0.5891026E-03	1.854013	0
1	2	15	1	708160	-0.1360582E-02	9.393588	0
1	2	16	1	752420	0.6877066E-03	2.161938	0
1	2	17	1	885200	0.8869610E-04	0.3488963E-01	0
1	2	18	1	929460	0.6726647E-03	2.514853	0
1	2	19	1	973720	-0.2147337E-03	0.2608317	0
1	2	20	1	1017980	-0.1783458E-03	0.1724326	0
1	2	21	1	1062240	-0.4098686E-03	0.9026605	0
1	2	22	1	1106500	-0.3223574E-03	0.5474696	0
1	2	23	1	1150760	-0.1354630E-03	0.1019349	0

snp_sol

Weights

- postGSf90 calculate weights that can be used to create a weighted genomic relationship matrix

$$G = ZDZ'/k$$

- Different type of weights
  - Default $w_i = 2p_i(1-p_i)a_i^2$
  - Non-Linear A $w_i = \mathrm{CT}^{\frac{|\widehat{a_i}|}{sd(\widehat{a})}-2}$
    - OPTION which_weight nonlinearA

Variance explained by windows

Windows segments starting at each SNP are defined

Windows size fixed or defined by Mb based on bp position

```
OPTION windows_variance n
```

Calculate the variance explained by *n* adjacents SNPs.

```
OPTION windows_variance_mbp n
```

Calculate the variance explained by *n* Mb window of adjacents SNPs.

Variance explained by segment segment information

```
windows_segment
```

contains information of windows segments used to get variance explainded

- 1: label
- 2: window size (number of SNP)
- 3: Start SNP number for the window
- 4: End SNP number for the window
- 5: identification of window: (ChrNumber)'_'(startPositionMBP)
- 6: Start (ChrNumber)'_'(Position) for the window
- 7: End (ChrNumber)'_'(Position) for the window

ssGWAS — less windows_segments — 96×24

```
window      19      1      19 1_0 1_0 1_973720
window      19      2      20 1_0 1_44260 1_1017980
window      19      3      21 1_0 1_88520 1_1062240
window      19      4      22 1_0 1_132780 1_1106500
window      20      5      24 1_0 1_221300 1_1195020
window      20      6      25 1_0 1_265560 1_1239280
window      19      7      25 1_0 1_309820 1_1239280
window      19      8      26 1_0 1_354080 1_1327800
window      19      9      27 1_0 1_398340 1_1372060
window      19     10      28 1_0 1_442600 1_1416320
window      19     11      29 1_0 1_531120 1_1504840
window      19     12      30 1_0 1_575380 1_1549100
window      19     13      31 1_0 1_619640 1_1593360
window      19     14      32 1_0 1_663900 1_1637620
window      19     15      33 1_0 1_708160 1_1681880
window      19     16      34 1_0 1_752420 1_1726140
window      21     17      37 1_0 1_885200 1_1858920
window      21     18      38 1_0 1_929460 1_1903180
window      21     19      39 1_0 1_973720 1_1947440
window      20     20      39 1_1 1_1017980 1_1947440
window      20     21      40 1_1 1_1062240 1_2035960
window      20     22      41 1_1 1_1106500 1_2080220
window      20     23      42 1_1 1_1150760 1_2124480
windows_segments
```

Variance explained by segments

```
windows_variance
```

contains variance explained for the biggest non-overlapping windows segments

- 1: trait
- 2: effect
- 3: Start SNP number or SNP name for the window
- 4: End SNP number or SNP name for the window
- 5: window size (number of SNP)
- 6: Start (ChrNumber)'_'(Position) for the window
- 7: End (ChrNumber)'_'(Position) for the window
- 8: identification of window: (ChrNumber)'_'(startPositionMBP)
- 9: variance explained by *n* adjacents SNP

ssGWAS — less windows_variance — 96×24

1	2	1	19	19	1_0	1_973720	1_0	0.05208
1	2	26	46	21	1_1327800	1_2301520	1_1	0.13269
1	2	47	69	23	1_2390040	1_3363760	1_2	0.07656
1	2	72	90	19	1_3585060	1_4558780	1_3	0.09174
1	2	94	116	23	1_4780080	1_5753800	1_4	0.14594
1	2	121	142	22	1_5975100	1_6948820	1_5	0.13505
1	2	149	169	21	1_7258640	1_8232360	1_7	0.14097
1	2	170	190	21	1_8276620	1_9250340	1_8	0.09060
1	2	191	212	22	1_9383120	1_10356840	1_9	0.01600
1	2	213	233	21	1_10401100	1_11374820	1_10	0.14180
1	2	241	261	21	1_11817420	1_12791140	1_11	0.05473
1	2	264	286	23	1_12968180	1_13941900	1_12	0.05335
1	2	289	307	19	1_14074680	1_15048400	1_14	0.01402
1	2	313	335	23	1_15313960	1_16287700	1_15	0.02237
1	2	336	357	22	1_16331940	1_17305660	1_16	0.07048
1	2	358	378	21	1_17349920	1_18323640	1_17	0.01520
1	2	383	404	22	1_18589200	1_19562920	1_18	0.02668
1	2	405	426	22	1_19607180	1_20580900	1_19	0.06459
1	2	429	449	21	1_20757940	1_21731660	1_20	0.02089
1	2	450	471	22	1_21775920	1_22749640	1_21	0.04006
1	2	484	506	23	1_23413540	1_24387260	1_23	0.04621
1	2	512	533	22	1_24697080	1_25670800	1_24	0.01662
1	2	542	561	20	1_26113400	1_27087120	1_26	0.05705

windows_variance

Manhattan plots options

```
OPTION Manhattan_plot
```

Plot using GNUPLOT the Manhattan plot (SNP effects) for each trait and correlated effect.

```
OPTION Manhattan_plot_R
```

Plot using R the Manhattan plot (SNP effects) for each trait and correlated effect.
`pdf` images are created: *manplot_St1e2.pdf*, but other formats can be specified.
Note: *t1e2* corresponds to trait 1, effect 2.

```
OPTION Manhattan_plot_R_format <format>
```

Control the format type to create images in R
`format` values accepted:

- pdf (default)
- png
- tif

```
Graphic control files
```

Several files are created to generate graphics using either GNUPLOT or R
File names rules:

```
e.g.: 'Sftle2.R'

first letter indicate
  'S' for solutions of SNP
  'V' for variance explained

tle2
  indicates that the file is for the trait 1 and the effect 2

filename extension
   .gnuplot
   .R
   .pdf
   .png
   .tif
```

Prediction for based on makers

- File **snp_pred** contains
  - allele frequencies
  - SNP estimates

- Program: **`predf90`** indirect genomic predictions based on maker information and SNP estimates from snp_pred file

Indirect Prediction for new animals

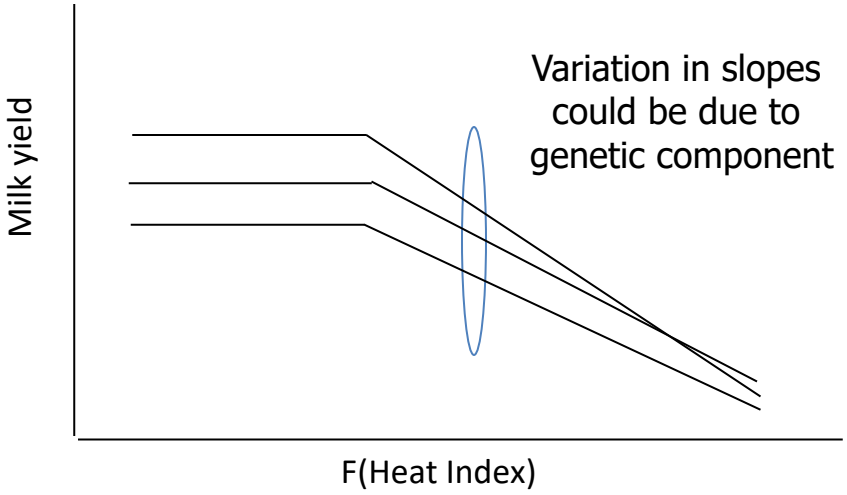
```
echo new_animals | predf90
```

ssGWAS — less pred.log — 96×24

```
 Predf90 1.04
 Predicts EBVs from genotypes based on results from single-step evaluation
 name of genotype file?
Number of SNP:    45000
Number of traits:   1
number of correlated traits:   1
       45000  SNP
The genotype file contains   45000 SNP starting from position  14
 UGA50014        1.98239330E-02
 UGA50016       -1.17402682E-02
 UGA50042       0.916237175
 UGA50058        3.55899930E-02
 UGA50060       -7.62441754E-02
 UGA50065      -0.194798529
 UGA50073      -0.370114744
 UGA50077      -0.308982432
 UGA50079      -0.913696527
Processed   200 genotypes
Average calling rate: 1.00
pred.log (END)
```

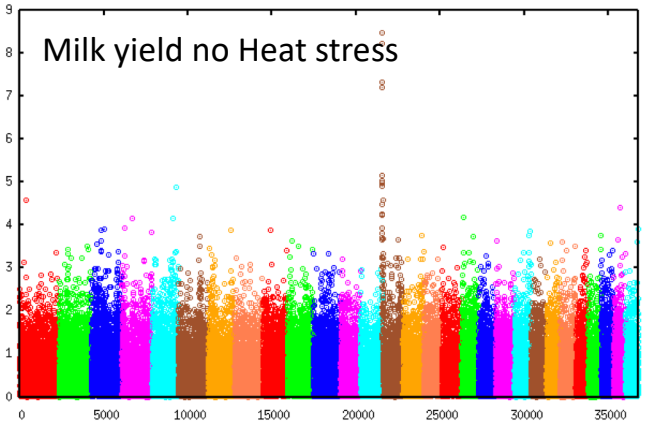
Model to study genetic of heat stress

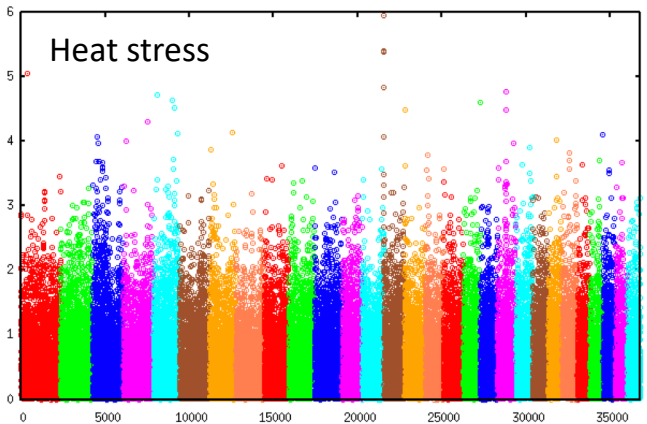
- Performance data + weather data (Ravagnolo & Misztal, 2000)



Single-Step GWAS Heat Stress

- Multiple-Trait Test-Day model heat tolerance
  - ~ 90 millions records, ~ 9 millions pedigrees
  - ~ 3,800 genotyped bulls
- Computing time
  - Complete evaluation ~ 16 h





Variance explained Heat Stress