

# Accounting for X chromosome and allele frequencies in genomic inbreeding estimation

**Juan P Nani & Paul M VanRaden**

**USDA, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD**

**Email: [juanonani@gmail.com](mailto:juanonani@gmail.com)**

**<https://aipl.arsusda.gov/>**

# Inbreeding control

Inbreeding can affect the **genetic progress** by increasing the prevalence of **harmful** or even **lethal recessive alleles** or by decreasing health, performance, and profitability through **inbreeding depression**

Breeders used pedigrees to **control** inbreeding, but genomics can provide more precise control

To measure and control inbreeding, **relationships** of each animal with the existing population are calculated when the genotype arrives

Genotypes are not available for all animals: The **correlation** between **G** and **A** is dependent on the use of different **allele frequencies** for the population

# Inbreeding control

- Genomic inbreeding estimates computed using **0.5** or **base population** allele frequency may be correlated very highly for some breeds but poorly for others
- Allele frequency of 0.5 has been used in U.S. evaluations since 2010
- New testing of options for assigning allele frequencies were needed
  
- Methods for calculating **F** generally ignored **X-Chromosome**
- **Low marker density** for the X-chromosome
- Pedigree and genomic relationships can both accurately account for the X-chromosome
- The SNPs in the **X-specific region are coded as 100% homozygous** in males ( $F_{gen}$  3% lower in females)

# Objectives of the study

- 1) To examine the impact of **allele frequency** on genomic inbreeding computation
- 2) To **adjust for the X-chromosome** to better account for differences between males and females and to better **match the scale** of genomic- to pedigree-based inbreeding and
- 3) To **improve the computational efficiency** of the procedure as the population of genotyped animals continues to grow

# Methods

- X-Chromosome corrections and allele frequencies
- scaling of **G** to **A** used only diagonal and off-diagonal genomic relationships among males
- 2,397 markers on the X-chromosome are 3.0% of the 79,060 total markers currently used

(1) statistical adjustment for females  $F_{gen} \sim = 0.97(F_{gen}^* + 0.03)$

(2) male-female or female-female off-diagonals  $G_{ij} \sim = 1.03 G_{ij}^*$

**Genomic haplotype-based inbreeding ( $F_{hap}$ ) was calculated using fixped.f90**

(<https://www.ars.usda.gov/northeast-area/beltsville-md-barc/beltsville-agricultural-research-center/agil/aip/software/fixped/>)

# Dataset

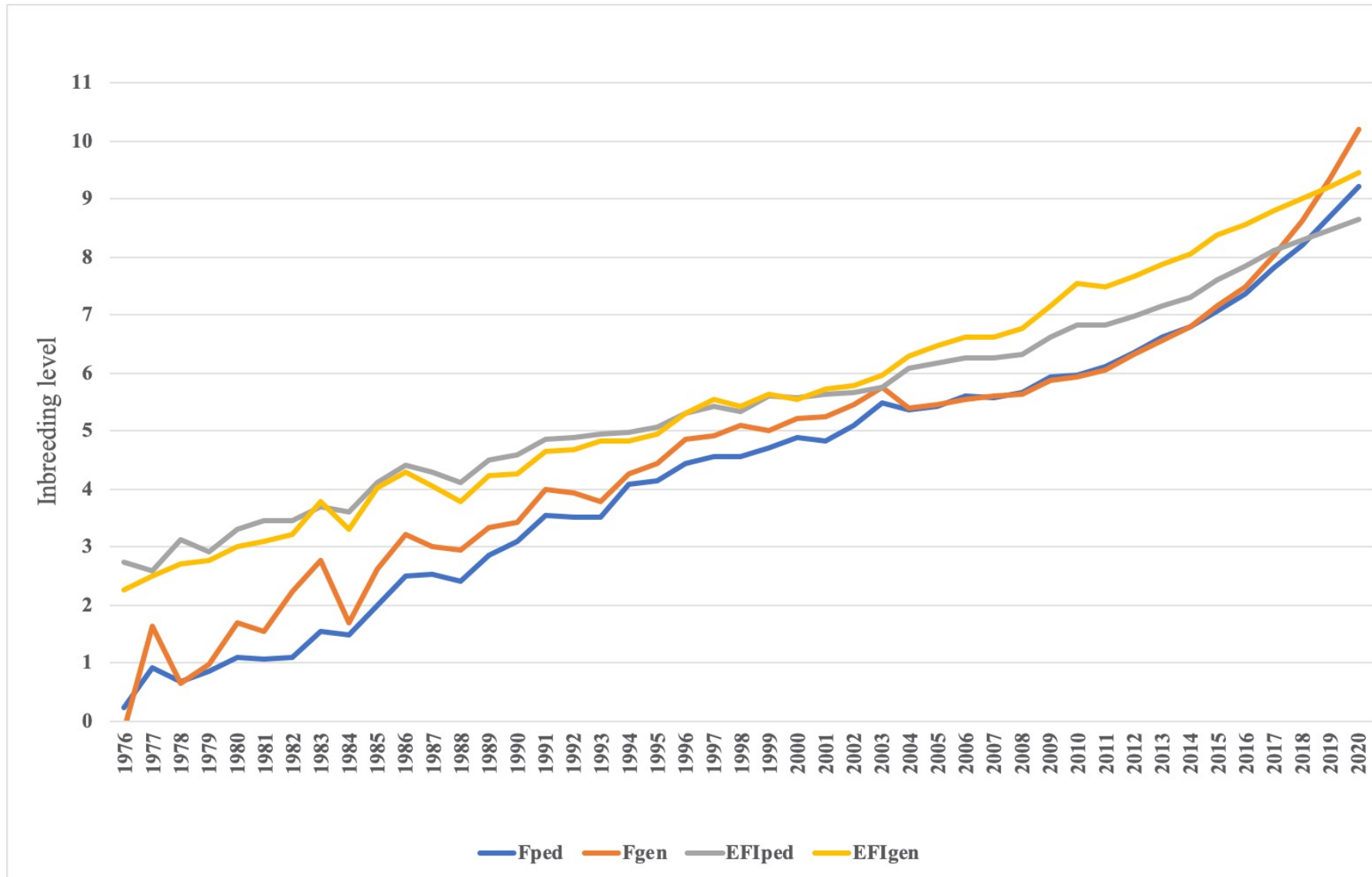
National cooperator database of the **Council on Dairy Cattle Breeding**: 3.5 million genotyped dairy cattle and 88 million animals in the pedigree file (Dec.2020)

Sex	Ayrshire	Brown Swiss	Guernsey	Holstein	Jersey
Females	7,228	11,958	4,162	3,313,344	422,700
Males	2,214	37,402	1,197	336,386	37,084

Genotypes from 44 different chips were imputed to ~80k markers using Findhap v.3 (<https://aipl.arsusda.gov/software/findhap/>)

# Results

## Mean F and EFI of dairy cattle across breeds by birth year



# Results: Inbreeding and EFI for all purebred genotyped animals

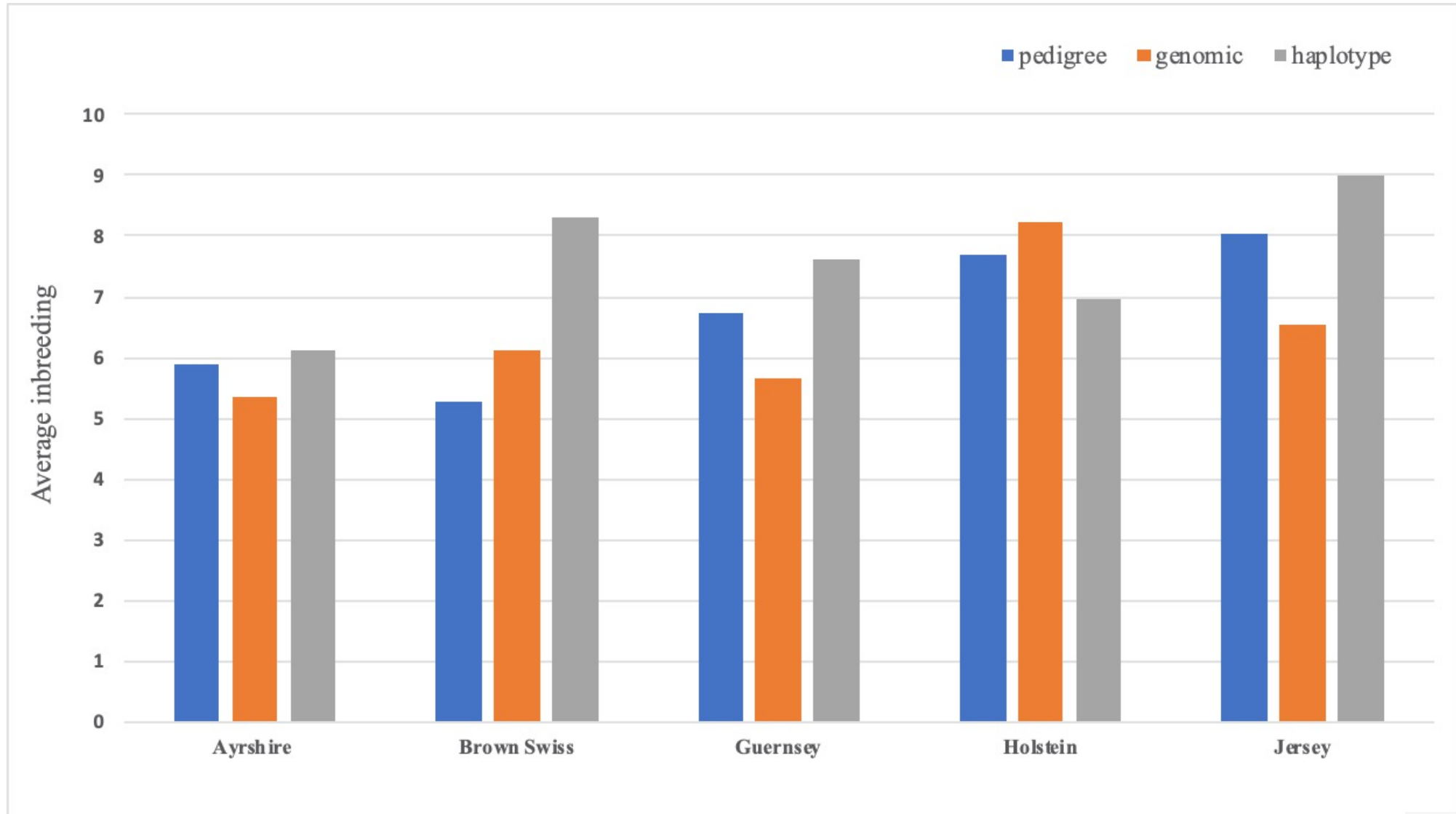
	Female			Male		
F/EFI	Minimum	Maximum	Mean	Minimum	Maximum	Mean
$F_{ped}$	0	49.1	7.65	0	34.6	8.09
$F_{gen}$	-22.40	46.7	7.92	-14.90	40.1	8.81
$F_{hap}$	0	44.3	7.14	0	37.13	7.84
$EFI_{ped}$	0.4	11.8	7.57	0.4	11.6	7.04
$EFI_{gen}$	-16.50	12.9	7.69	-8.10	12.8	7.3

average  $F_{gen}$  sex difference was 0.9 and 0.7 for  $F_{hap}$



# Results

## Inbreeding by breed



# Results

## Effects of different allele frequencies

Mean F and EFI for genotyped animals by breed and sex

	Ayrshire			Brown Swiss			Guernsey			Holstein			Jersey		
F/EFI	F (5,957)	M (1,823)	Dif.	F (9,161)	M (27,853)	Dif.	F (3,893)	M (1,172)	Dif.	F (2,687,038)	M (313,394)	Dif.	F (206,034)	M (24,428)	Dif.
$F_{ped}$	6.00	5.80	-0.2	6.91	4.81	-2.1	6.87	6.26	-0.61	7.78	8.63	0.85	7.87	7.42	-0.45
$F_{gen0.5}$	2.72	5.88	3.16	5.81	5.61	-0.2	5.85	8.83	2.98	5.62	9.66	4.04	4.28	7.15	2.87
$F_{genbase}$	6.61	7.72	1.11	4.68	6.81	2.13	5.47	7.02	1.55	5.89	9.73	3.84	5.17	7.95	2.78
$F_{genX}$	<b>5.55</b>	<b>5.88</b>	<b>0.33</b>	<b>8.55</b>	<b>5.61</b>	<b>-2.94</b>	<b>5.70</b>	<b>5.79</b>	<b>0.09</b>	<b>8.36</b>	<b>9.66</b>	<b>1.3</b>	<b>7.06</b>	<b>7.15</b>	<b>0.09</b>
$F_{hap}$	6.23	6.39	0.16	10.12	7.95	-2.17	7.64	7.61	-0.03	7.04	7.88	0.84	9.23	9.07	-0.16
$EFI_{ped}$	7.04	6.73	-0.31	7.18	5.26	-1.92	7.60	7.42	-0.18	7.95	8.15	0.2	8.05	7.63	-0.42
$EFI_{gen0.5}$	6.83	6.55	-0.28	7.29	5.67	-1.62	7.42	7.47	0.05	8.68	9.08	0.4	7.76	7.47	-0.29
$EFI_{genbase}$	9.51	8.68	-0.83	7.03	5.98	-1.05	8.17	7.84	-0.33	9.06	9.41	0.35	8.18	7.85	-0.33
$EFI_{genX}$	7.01	6.65	-0.36	7.42	5.72	-1.7	7.18	7.22	0.04	8.88	9.27	0.39	7.98	7.64	-0.34

# Results: Across breeds and sexes correlations for the different inbreeding definitions

Pedigree ( $F_{ped}$ ), genomic with 0.5 ( $F_{gen0.5}$ ), base population ( $F_{genbase}$ ), 0.5 + X correction ( $F_{genX}$ )

	Ayrshire		Brown Swiss		Guernsey		Holstein		Jersey	
F/EFI correlation	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males
$F_{genbase}, F_{ped}$	0.31	0.38	0.59	0.28	0.54	0.65	0.71	0.81	0.54	0.57
$F_{gen0.5}, F_{ped}$	0.67	0.70	0.61	0.68	0.64	0.62	0.72	0.83	0.56	0.69
$F_{genX}, F_{ped}$	<b>0.67</b>	<b>0.70</b>	<b>0.61</b>	<b>0.68</b>	<b>0.64</b>	<b>0.62</b>	<b>0.72</b>	<b>0.83</b>	<b>0.56</b>	<b>0.69</b>
$F_{hap}, F_{ped}$	0.66	0.70	0.51	0.62	0.64	0.60	0.66	0.77	0.59	0.69
$F_{genbase}, F_{gen0.5}$	0.29	0.26	0.82	0.22	0.65	0.56	0.95	0.96	0.59	0.58
$F_{genX}, F_{gen0.5}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$F_{hap}, F_{gen0.5}$	0.93	0.95	0.87	0.94	0.90	0.89	0.93	0.94	0.91	0.93
$EFI_{gen0.5}, EFI_{ped}$	0.88	0.92	0.68	0.85	0.84	0.85	0.84	0.91	0.65	0.85
$EFI_{genbase}, EFI_{ped}$	0.67	0.81	0.78	0.87	0.87	0.93	0.86	0.91	0.85	0.90
$EFI_{genX}, EFI_{ped}$	0.88	0.93	0.67	0.85	0.83	0.85	0.84	0.91	0.65	0.85
$EFI_{gen0.5}, EFI_{genbase}$	0.63	0.77	0.76	0.93	0.83	0.82	0.97	0.98	0.74	0.83
$EFI_{gen0.5}, EFI_{genX}$	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00

# Discussion

- Correlations of  $F_{\text{gen}}$  with  $F_{\text{ped}}$  were higher for most breeds by centering the genotypes using 0.5
- Correlations of  $\text{EFI}_{\text{gen}}$  with  $\text{EFI}_{\text{ped}}$  tended to be higher using the base population frequencies
- $F_{\text{hap}}$  had lower correlations with  $F_{\text{ped}}$  than using genotypes centered by 0.5
- X-chromosome corrections allowed the averaged differences between male and female to be smaller
- Corrections for the X-chromosome contribution also eliminated the differences within full-sib families
- Improving of computation efficiency by CDCB's Gerald Jansen

# Conclusions

- Genomic measures of relationship and inbreeding **differ** for males and females.
- Adjustments to **scale genomic measures** to match pedigree relationships were developed and applied.
- Adjusting pedigree relationships for **variation on the X-chromosome** is feasible
- Inbreeding calculated from **haplotypes** did not give higher correlations with  $F_{ped}$
- Improvements in computation efficiency resulted in **rapid calculations** allowing more frequent updates to account for changing pedigrees and genotypes.

# Acknowledgments

- Council on Dairy Cattle Breeding (**CDCB**) and its industry suppliers for data
- Agricultural Research Service (ARS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE; Oak Ridge, TN)

