

Levels of heterozygosity calculations

The comparison between level of heterozygosity and inbreeding coefficient for single animals

The correlation between level of heterozygosity and inbreeding coefficient is **-0.47** when all the Danish pure and crossbred animals are pooled together. The crossbred animals in the following plots are from crossbred routine evaluation. The animals that cannot be traced back at least for 5 generations were excluded (28554 out of 342819 animals).

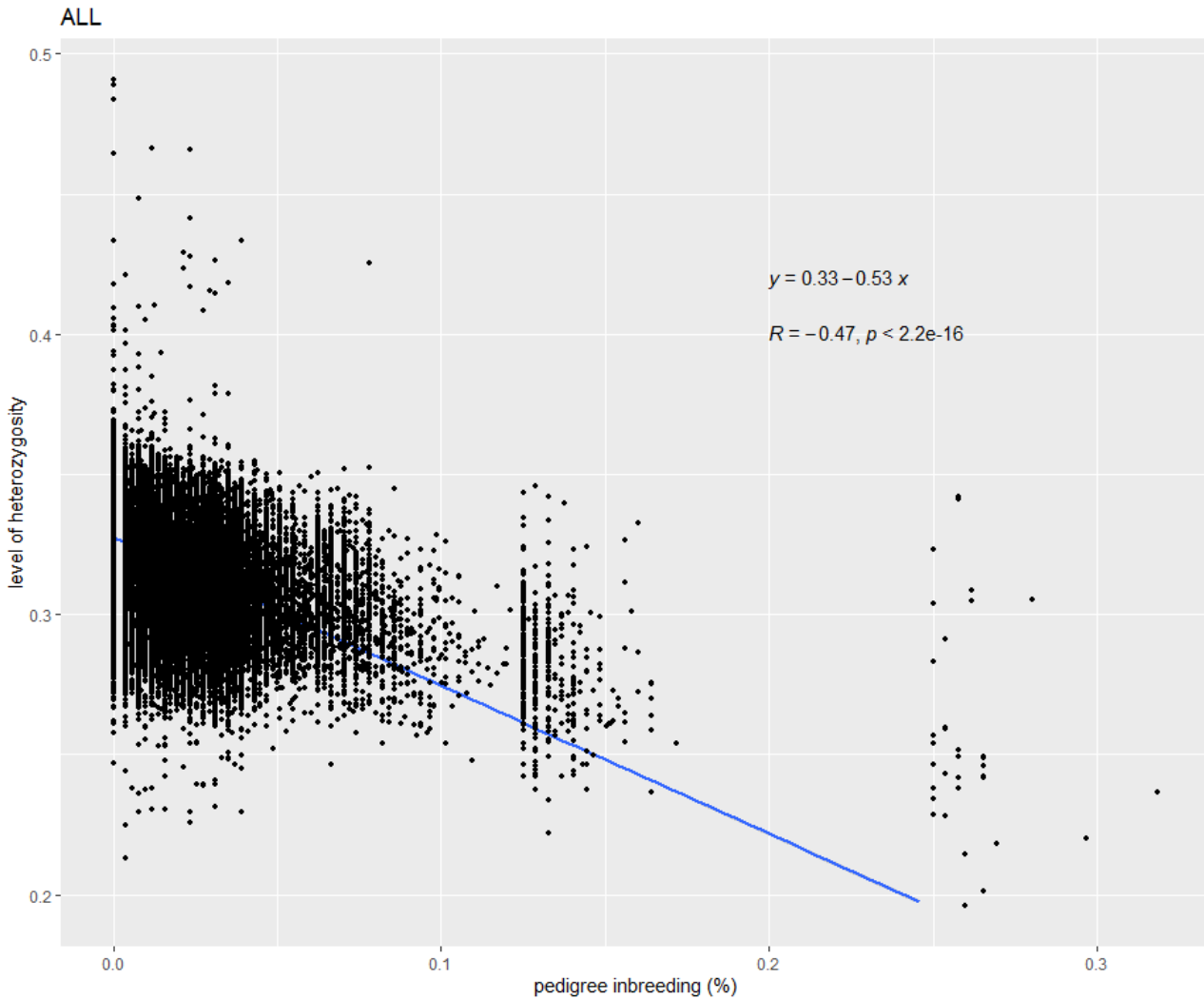


Figure 1. level of heterozygosity plotted against inbreeding coefficient for all the Danish genotyped animals.



STØTTET AF

Mælkeafgiftsfonden

Table 1. Number of animals, mean and std of level of heterozygosity (number of heterozygotes/total number of snps in the genome, mean and std of inbreeding coefficient (%) based on the pedigree traced back for 5 generations, correlation between level of heterozygosities and inbreeding and number of NA records of inbreeding for each Danish breed. The XXX animals in the purebred evaluations are not included.

| Breed | nAnimals | mean of Het (std) | mean of inbreeding (% std) | max of inbreeding (%) | correlation(Het, inbreeding) | NA records of inbreeding |
|-------|----------|-------------------|----------------------------|-----------------------|------------------------------|--------------------------|
| HOL | 176172 | 0.326(0.01) | 0.70(1.06) | 26.6 | -0.28 | 412 |
| JER | 90378 | 0.306(0.01) | 1.97(1.45) | 31.8 | -0.30 | 252 |
| RDC | 40235 | 0.331(0.01) | 1.00(1.25) | 26.6 | -0.36 | 897 |
| XXX | 7480 | 0.349(0.01) | 0.12(0.40) | 7 | -0.40 | 12 |

The number of snps is 46342 in HOL, 41897 in JER, 46914 in RDC and 47586 in XXX. It is the same as in Routine evaluation.

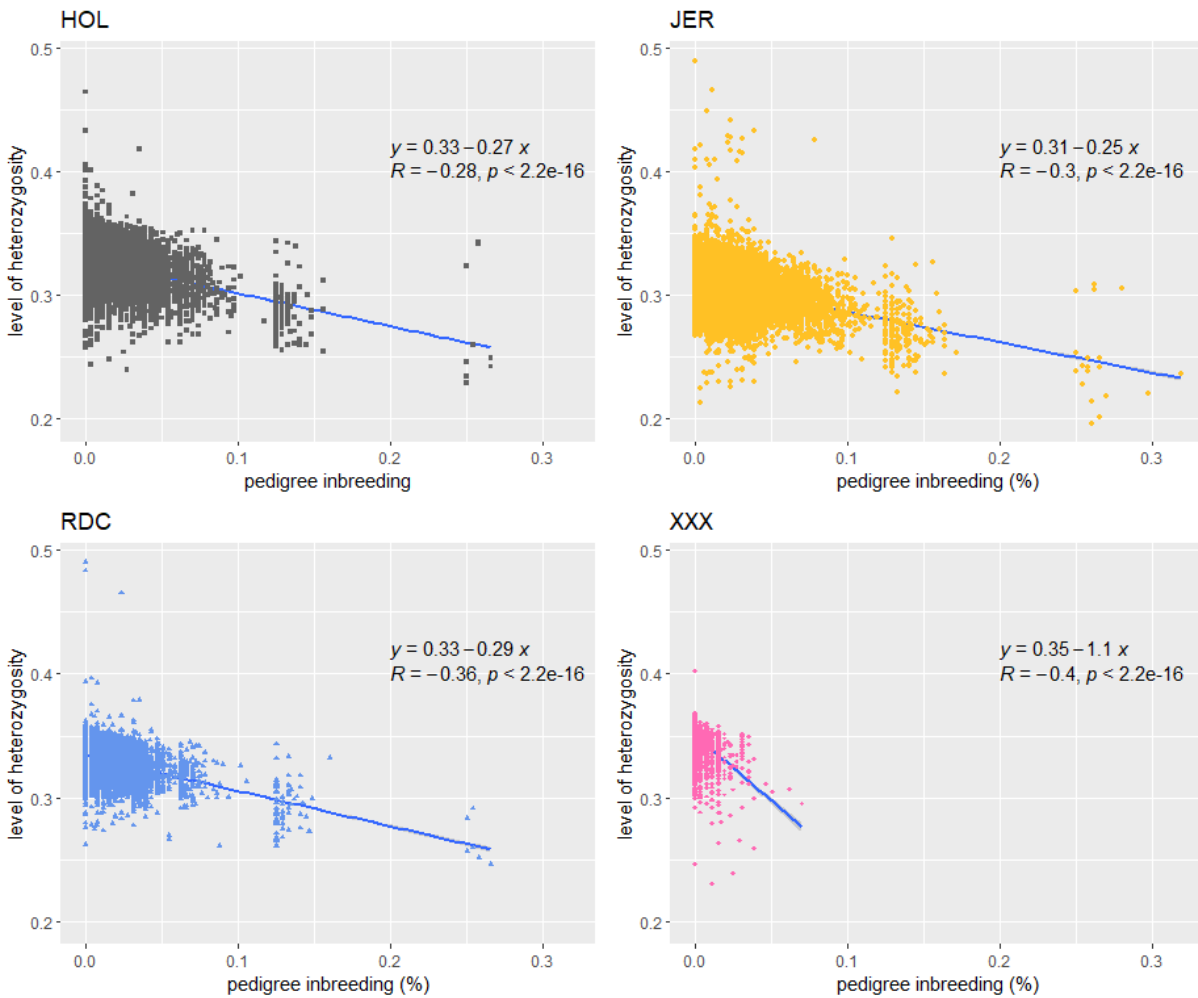


Figure 2. level of heterozygosity plotted against inbreeding coefficient for each breed. (The Rcode for this plot is in C:/MatingPlan/DATAcomparison)

Some test runs showed that the time spent on running 5 million pairs is 27 minutes.



STØTTET AF

Mælkeafgiftsfonden

Steps to compute level of heterozygosity

- The mating list to be filled out is called Heterozygoti_export.csv. In total there were 15656821 mating pairs.
- The ID of all the animals were extracted from this mating list. In total there were 112778 unique animals, where 139 unique animals in sire column and 112639 unique animals in dam column. The distribution of birth years in males and in females were given in the frequency Table 2 and Table 3 below. The males went to the mating list need to be double checked. Some of these animals are too old and some of them were not in the Nav_ped.
- The number of genotyped males and females for each breed:
 - ✓ HOL: 27 males and 62338 females.
 - ✓ JER: 22 males and 34492 females. * *see note*
 - ✓ RDC: 15 males and 12414 females.
 - ✓ XXX: 0 males and 3370 females.
 - ✓ In total 64 (46%) males and 112614 (99.9%) females were genotyped.
- The genotypes of these animals were extracted from the Routine evaluation. These genotypes are imputed genotypes using FImpute v2.2.
- The maps of the three breeds were merged, and there are 41238 SNPs in common (meaning that these SNPs appeared in all the breeds).
- We compared level of heterozygosity of all HOL sire and HOL dam mating pairs calculated by using 46342 SNPs (HOL set) and by using 41238 SNPs (common set). The correlation is 0.998. Therefore, it is OK to use the common SNP set.
- The expected heterozygosity was calculated SNP by SNP for the genotyped parents. If one of the parents or both parents are heterozygotes for the SNP, then the expected hetero is $\frac{1}{2}$. If the parents are alternative homozygosities for the SNP, then the expected heterozygosity is 1. The expected level of heterozygosity for a genotyped pair is then calculated as the sum of expected heterozygosity of all SNPs divided by the total number of SNPs (41238).
- 7207488 (equals 64 x 112617 **see note below*) mating pairs got a heterozygosity level. It means 46% of the mating pairs in the mating list got a level of heterozygosity.
- Level of heterozygosity for genotyped pairs can be found in:
/usr/home/nav/denmark/navhuli/mating_OnlyWithGenome.txt
Level of heterozygosity for all pairs (the same dimensions as the original mating list; -99 for missing) can be found in:
mating_fullTable_withHeader.txt

*Note:

Actually 25 JER males were genotyped, but 3 of them were misplaced in the dam column in the mating list:

JERDNKM000001851107377

JERDNKM000001851110649

JERDNKM000004517105490

And they do not have a combination with any dams in the mating list.

Therefore, there are 112616 - 2 females+3 males (112617 animals) in dam column in the mating list.



STØTTET AF

Mælkeafgiftsfonden

Next step: Converting the expected level of inbreeding to expected level of heterozygosity for mating pairs with at least one un-genotyped parent.

The FREQ Procedure

| year | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|------|-----------|---------|-------------------------|-----------------------|
| 1945 | 1 | 0.85 | 1 | 0.85 |
| 1946 | 2 | 1.69 | 3 | 2.54 |
| 1948 | 5 | 4.24 | 8 | 6.78 |
| 1949 | 5 | 4.24 | 13 | 11.02 |
| 1950 | 13 | 11.02 | 26 | 22.03 |
| 1951 | 11 | 9.32 | 37 | 31.36 |
| 1952 | 13 | 11.02 | 50 | 42.37 |
| 1983 | 1 | 0.85 | 51 | 43.22 |
| 1987 | 5 | 4.24 | 56 | 47.46 |
| 2019 | 23 | 19.49 | 79 | 66.95 |
| 2020 | 39 | 33.05 | 118 | 100.00 |

Table 2. The distribution of birth years in males.



STØTTET AF

Mælkeafgiftsfonden

The FREQ Procedure

| year | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|------|-----------|---------|-------------------------|-----------------------|
| 2001 | 1 | 0.00 | 1 | 0.00 |
| 2003 | 2 | 0.00 | 3 | 0.00 |
| 2004 | 7 | 0.01 | 10 | 0.01 |
| 2005 | 19 | 0.02 | 29 | 0.03 |
| 2006 | 40 | 0.04 | 69 | 0.06 |
| 2007 | 65 | 0.06 | 134 | 0.12 |
| 2008 | 178 | 0.16 | 312 | 0.28 |
| 2009 | 325 | 0.29 | 637 | 0.57 |
| 2010 | 1333 | 1.18 | 1970 | 1.75 |
| 2011 | 2797 | 2.48 | 4767 | 4.23 |
| 2012 | 4511 | 4.01 | 9278 | 8.24 |
| 2013 | 6285 | 5.58 | 15563 | 13.82 |
| 2014 | 8857 | 7.86 | 24420 | 21.68 |
| 2015 | 4008 | 3.56 | 28428 | 25.24 |
| 2016 | 5622 | 4.99 | 34050 | 30.23 |
| 2017 | 9198 | 8.17 | 43248 | 38.40 |
| 2018 | 12957 | 11.51 | 56205 | 49.91 |
| 2019 | 15735 | 13.97 | 71940 | 63.88 |
| 2020 | 20025 | 17.78 | 91965 | 81.66 |
| 2021 | 20654 | 18.34 | 112619 | 100.00 |

Table 3. The distribution of birth years in females.



STØTTET AF

Mælkeafgiftsfonden