PARENTAGE ASSIGNMENT IN SALMON USING HIGH DENSITY SNP PANELS: A SIMULATION STUDY

Grashei K. E.[1,2]
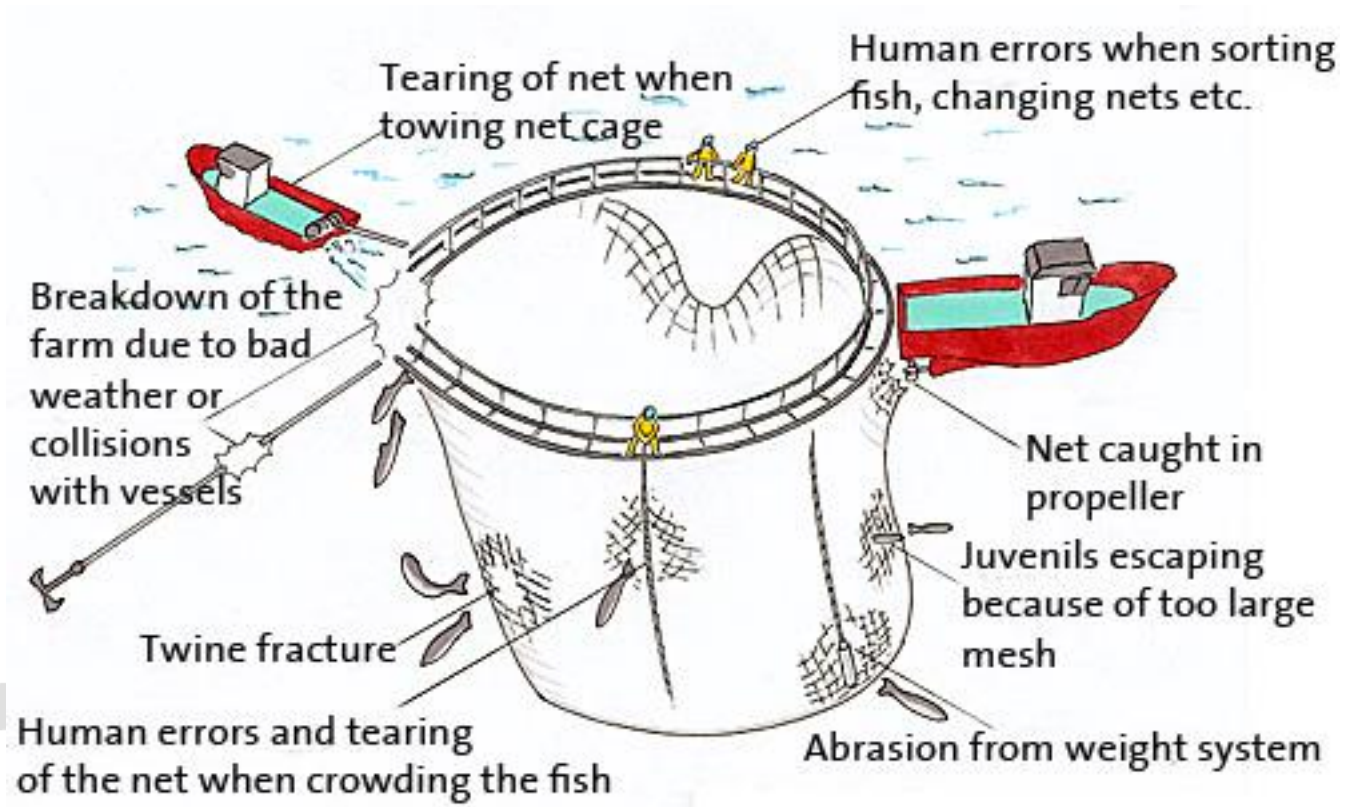
Kjøglum S.[1], Moen T.[1], Ødegård J.[1,2]

AquaGen AS[1]

Norwegian University of Life Sciences[2]

AquaGen

Farmed fish escapes

The illustration shows common escape factors with special focus on holes in the nets and sea-cage operations. (Illustration: Mats A. Heide, Sintef)
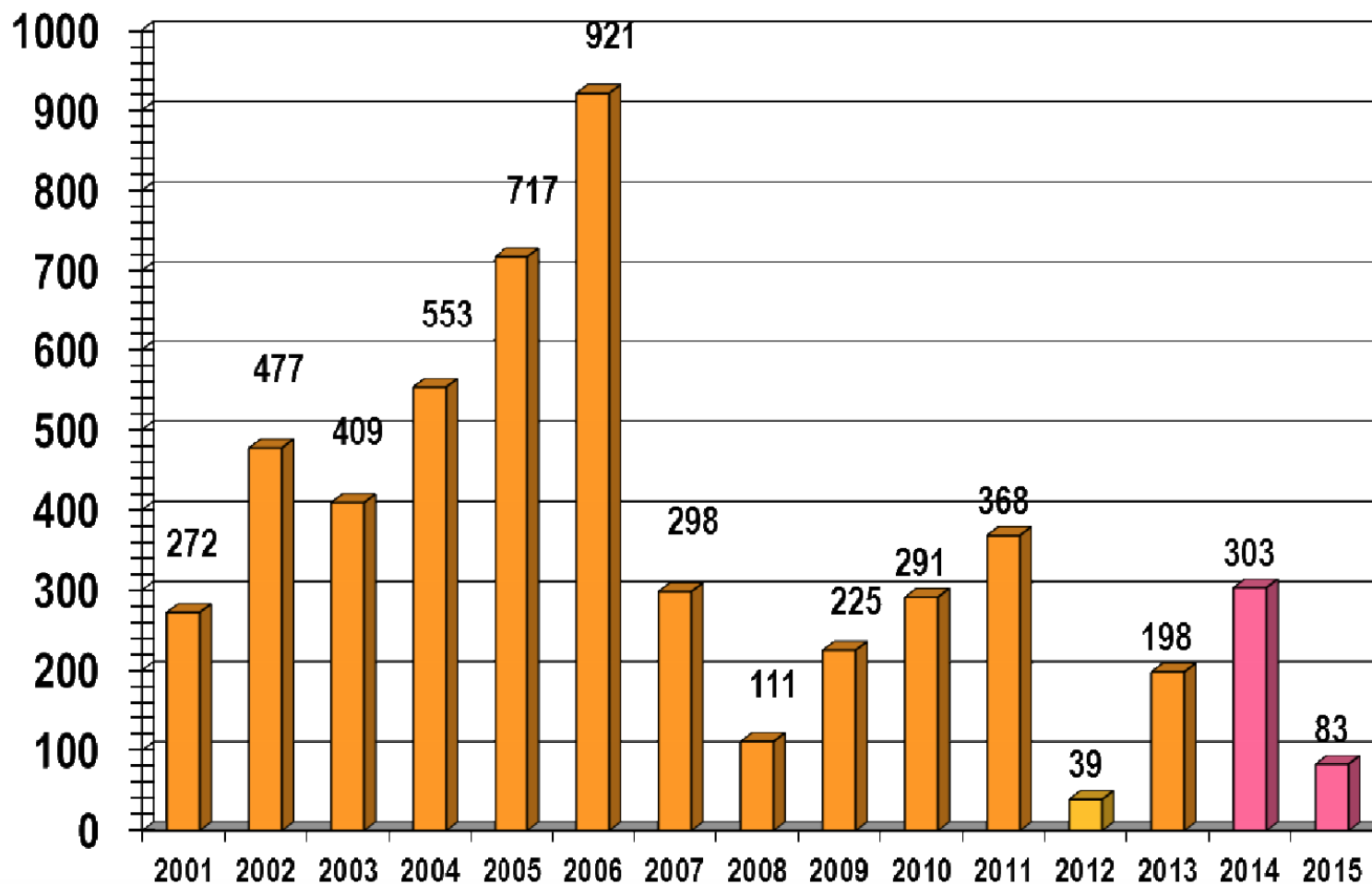
Rømming av laks 2001-2015
Oppdretternes innmeldte rømmingstall pr. 28.02.2015

TRACK project/product

- Unique parents for all offspring of each fish farm

- Genotype the parents with dense SNP chip

- Escaped fish genotyped and assigned parents

- > 45 million fertilized TRACK-eggs sold

AquaGen

Current methods

- Exclusion based:
  - Mendelian, easy to understand
  - Must adjust model when number of SNPs change
  - What is the "correct" number of exclusions between offspring and real parents?
- Probability based
  - LRs, POR, LOD score, etc.
  - Reasonably high MAFs
  - Models needed for typing errors and marker dependency assumptions
  - Multiplicative

AquaGen

# Alternative method – Genomic relationship likelihood (GRL) Part 1

- Trio assignment – Offspring and both parents
- Uses genomic relationship between individuals
- G-matrix relationships used in genomic selection (VanRaden 2008, method 2)

Alternative method – Genomic relationship likelihood (GRL) Part 2

- r_{ij} is the genomic relationship between animals *i* and *j*
- $E(r_{o,p1} \mid parents) = 0.5(r_{p1,p1} + r_{p1,p2})$
- $E(r_{o,p2} \mid parents) = 0.5(r_{p2,p2} + r_{p1,p2})$
- $E(r_{o,o} \mid parents) = 1 + 0.5r_{p1,p2}$
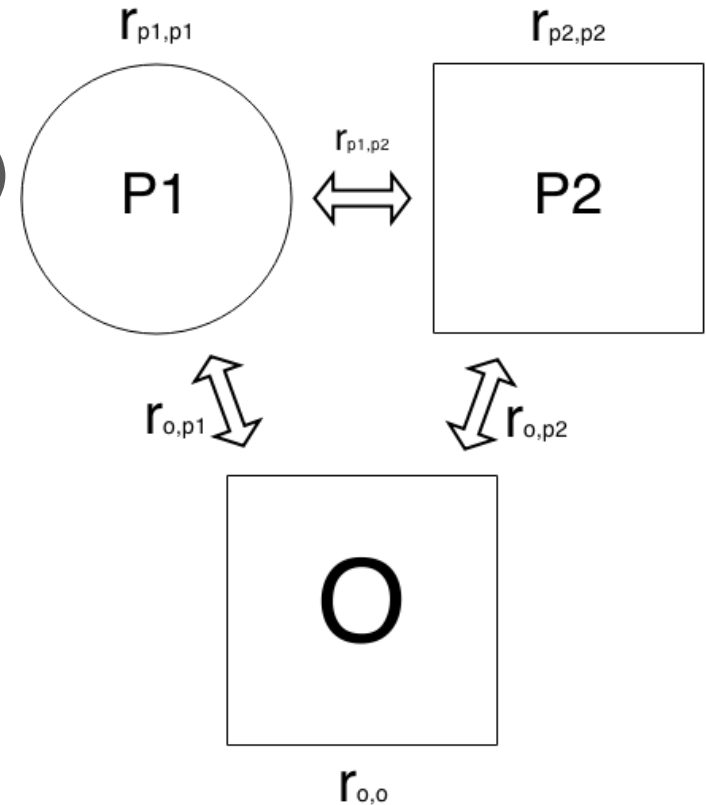
- Residuals:
- $e_{o,p1} = r_{o,p1} - E(r_{o,p1} \mid parents)$
- $e_{o,p2} = r_{o,p2} - E(r_{o,p2} \mid parents)$
- $e_{o,o} = r_{o,o} - E(r_{o,o} \mid parents)$

- Residuals for real parents used as training data to fit a multinormal distribution
  - Produces the genomic relationship likelihood for test data

AquaGen

Simulation

- QMSim (Sargolzaei, M. and F. S. Schenkel. 2009) used for simulation

- Historical population:
  - N_e=500 individuals over 5000 generations

- Population:
  - 25 males X 250 females for 10 generations
  - Family size: 4 (total 1000 offspring per generation)
  - Using genotypes from the last two generations

- Genome
  - 30 chromosomes of length 100 cM with 30-7000 SNPs
  - Genotype error: 0.1-0.5%
  - Training data for GRL is 500 animals for each SNP count
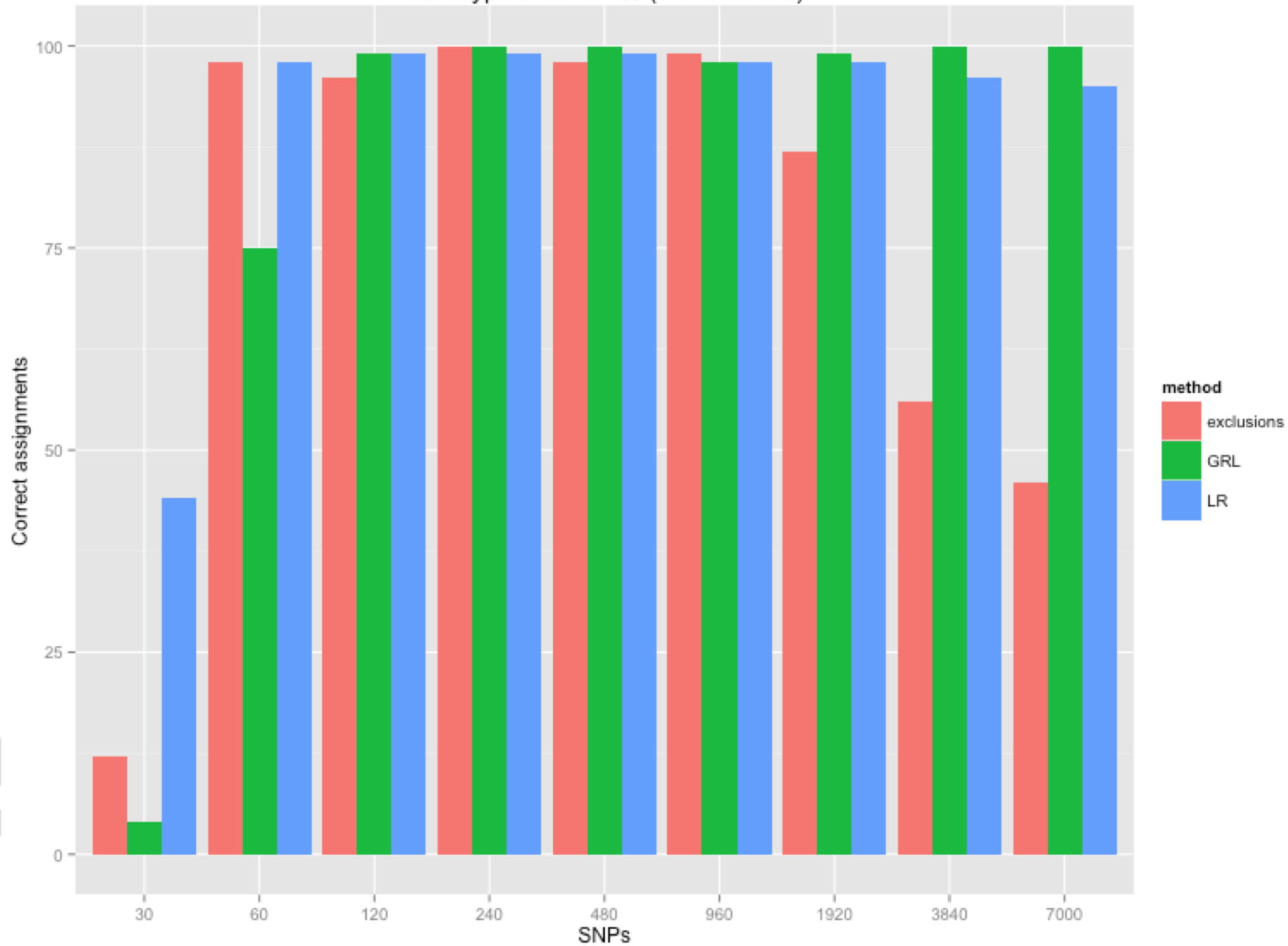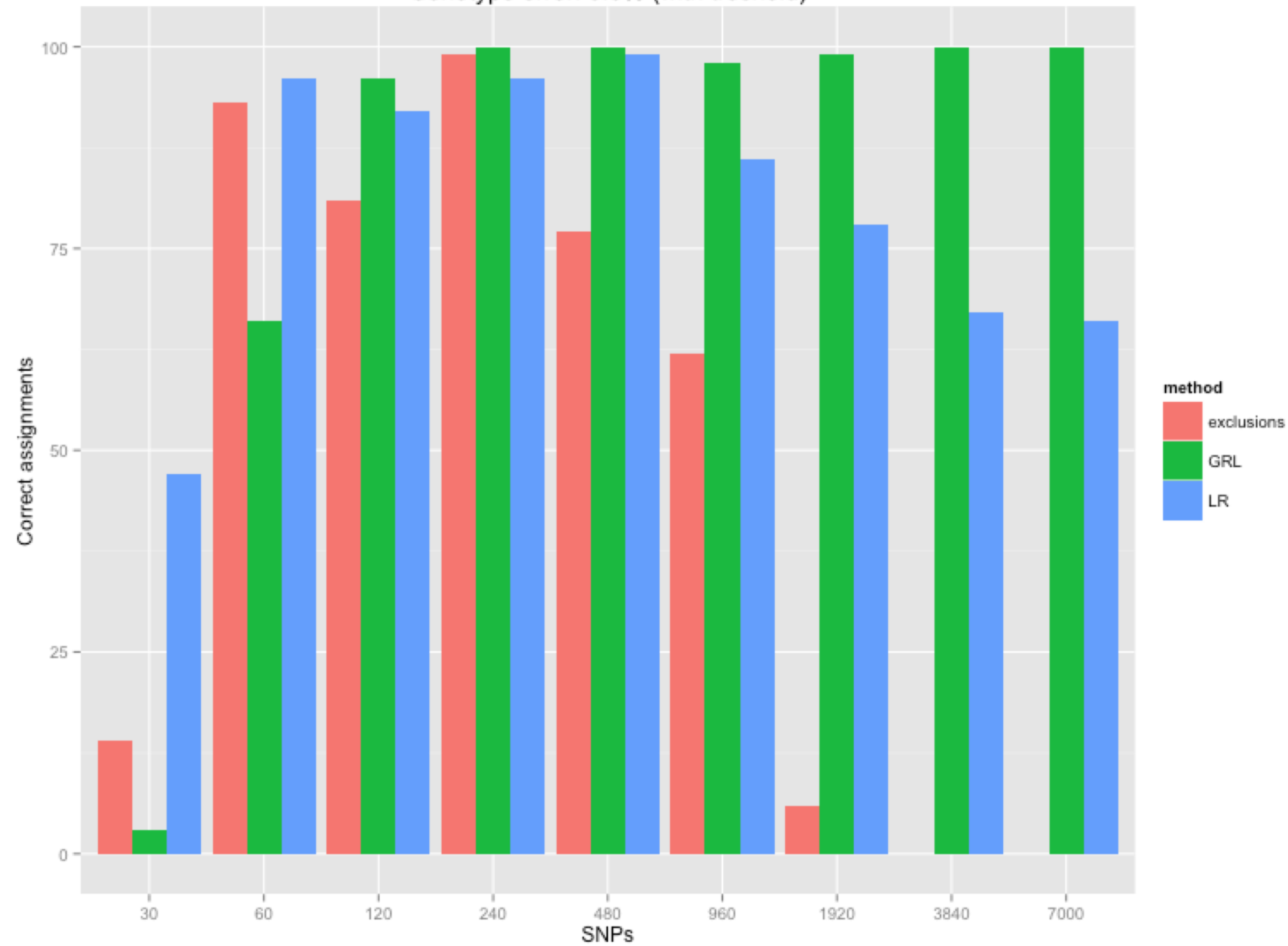
AquaGen
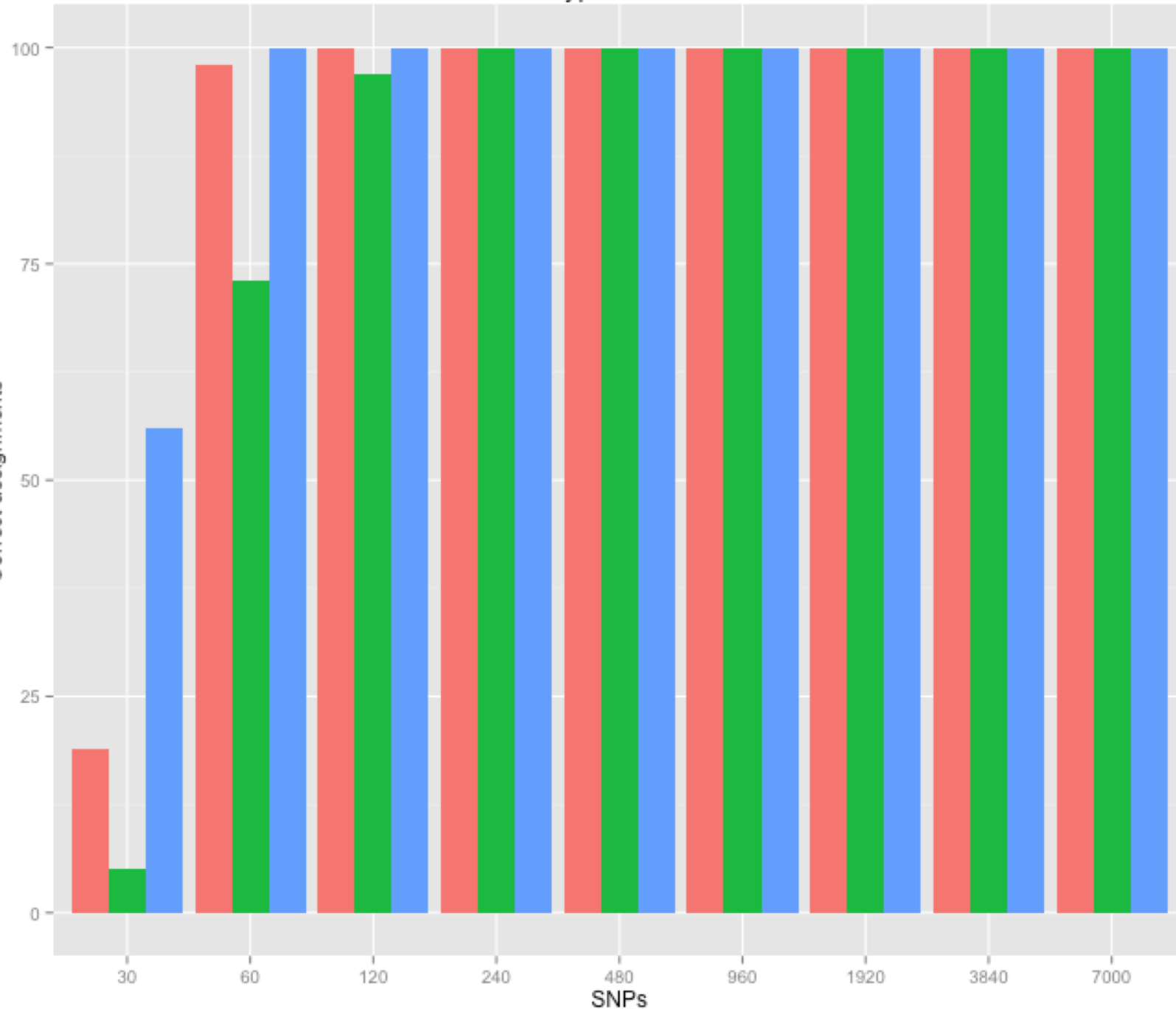
Genotype error: 0.2% (with treshold)

Genotype error: 0.5% (with treshold)

Method performance – 7088 SNPs

Offspring	Parent candidates	LR (seconds)	GRL (seconds)	LR / GRL (time ratio)
1	1 000	0	9	**0**
100	1 000	43	31	**1.39**
1 000	1 000	429	222	**1.93**
100	1	0.01	0.1	**0.1**
100	1 000	44	30	**1.47**
100	2 000	121	183	**1.51**

AquaGen

Method performance – 99 990 SNPs

Offspring	Parent candidates	LR (seconds)	GRL (seconds)	LR / GRL (time ratio)
100	500	219	53	**4.1**
100	1 000	956 (~16min)	255 (~4min)	**3.7**
100	1500	2 118 (~35 min)	394 (~7 min)	**5.4**
500	500	1087 (~18min)	128 (~2 min)	**8.5**
500	1 500	9660 (~2.7hrs)	772 (~13min)	**12.5**
500	2000	17071 (~4.7hrs)	1176 (~20min)	**14.5**
1 000	500	2183 (~36 min)	221 (~4 min)	**9.9**
1 000	1 000	8603 (~2.4hrs)	658 (~11 min)	**13.1**
1 000	2 000	34259 (~9.5 hrs)	1997 (~33 min)	**17.1**

AquaGen

Summary

- Alternative method using genomic relationship likelihood (GRL) to assign both parents to an offspring

- Initial results indicate good performance (but need more data)

- Reusability of data for genomic selection and possibly for other areas such as population assignment

- Seems like the GRL method is less affected by genotyping errors

Partly funded by:
(project num. 245519)

Collaborator: