

Sustainable genetics project

Script

Ane Liv Berthelsen

2023-11-20

Contents

Materials and Methods	1
Identify mismatches	2
Model	2
Results	4
Tables	4
Data	7
PCR Program	7

#Aims We set out to explore the re-useability of 96-micro well plate with the aim to test the following hypothesis:
1) 96 microwell plates can be cleaned and reused within the context of microsatellite genotyping, 2) the genotype error rate will be higher for plates reused for PCR amplification.

Materials and Methods

We designed an experimental setup to assess the re-usability of 96-micro well plates within the context of microsatellite genotyping. The setup contained 4 treatments: standard procedure, internal control, re-used PCR plate and re-used detection plate (See Figure 3).

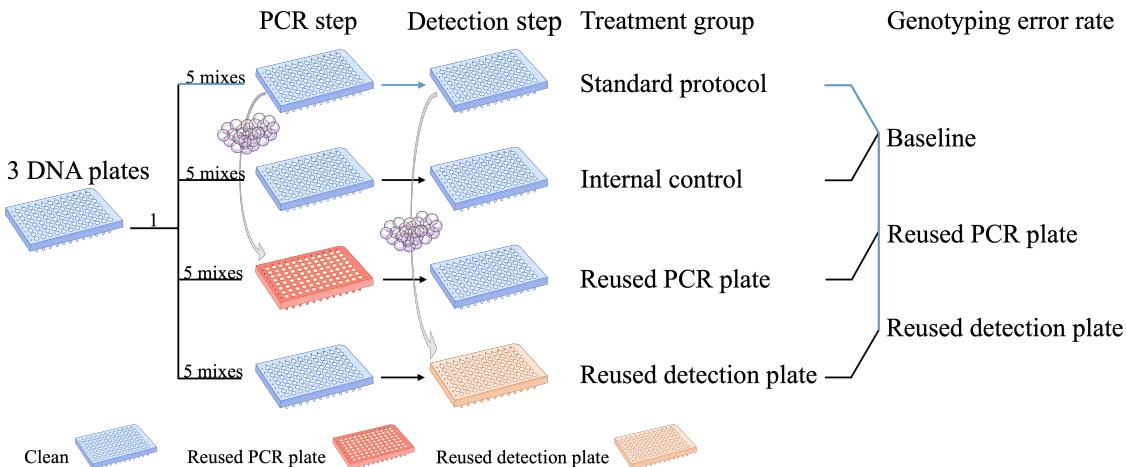


Figure 1: Schematic overview

Data evaluation

We calculated the allelic and genotype error rate between the different treatments using the ‘standard procedure’ treatment as reference. The ‘internal control’ treatment was included in the experimental setup to compute a baseline for error rates. The comparison between the different treatments and the standard procedure yielded a list of mismatches.

For each sample we noted whether the genotype matched with the ‘standard procedure’ per loci and for all treatments whether the genotype was scored or not. This yielded two datasets referred to as ‘match’ and ‘missing genotype’ data respectively.

Identify mismatches

Table 1: Error rates

Rack 1	Rack 2	Treatment	No. of mistyped alleles	No. of mistyped reactions	No. of reactions	Allelic error rate	Genotype error rate
R154_RR154_R	Internal control		9	7	3467	0.0012980	0.0020190
R155_RR155_R	Internal control		20	17	3567	0.0028035	0.0047659
R156_RR156_R	Internal control		45	33	3507	0.0064157	0.0094098
R154_RR154_R	Re-used detection plate		15	12	3478	0.0021564	0.0034503
R155_RR155_R	Re-used detection plate		11	10	3636	0.0015127	0.0027503
R156_RR156_R	Re-used detection plate		26	20	3485	0.0037303	0.0057389
R154_RR154_R	Re-used PCR plate		139	114	2898	0.0239821	0.0393375
R155_RR155_R	Re-used PCR plate		113	96	3261	0.0173260	0.0294388
R156_RR156_R	Re-used PCR plate		68	55	3231	0.0105231	0.0170226

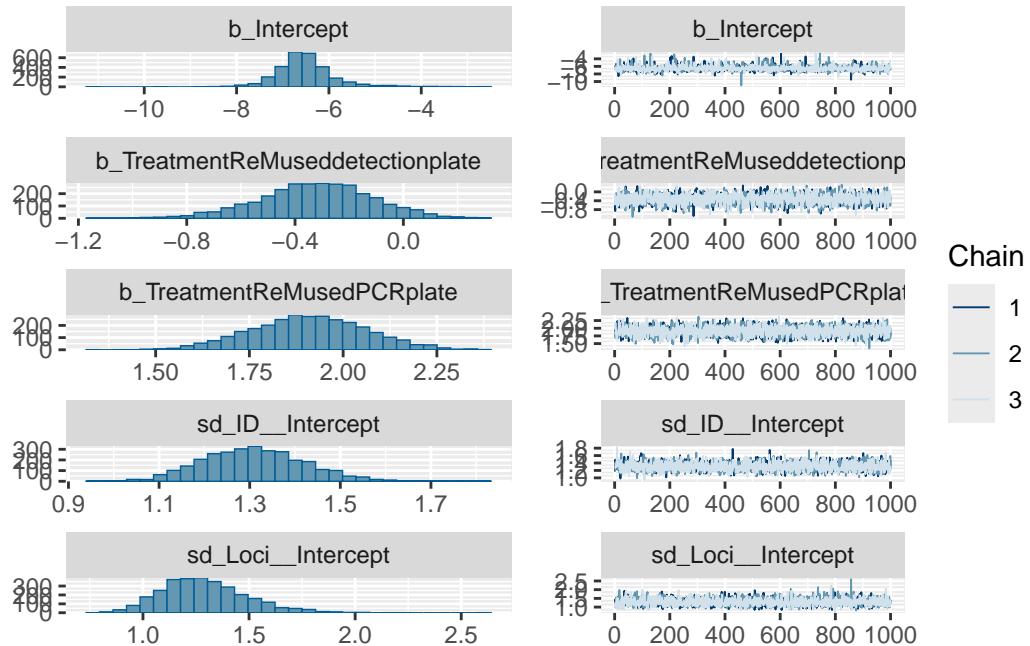
Model

We ran a Bayesian binary logistic regression mixed model with treatment included as a three level categorical variable to explore the effect on mismatched genotypes data. DNA plate, multiplexes and loci variables were included as random effects in the model. All statistical analyses were implemented in R version 4.2.1 with Rstudio version 2023.09.1+494.

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Warning: The `facets` argument of `facet_grid()` is deprecated as of ggplot2 2.2.0.
  i Please use the `rows` argument instead.
  i The deprecated feature was likely used in the bayesplot package.
    Please report the issue at <https://github.com/stan-dev/bayesplot/issues/>.

Warning: There were 6 divergent transitions after warmup. Increasing
adapt_delta above 0.99 may help. See
http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
```



```

Family: bernoulli
Links: mu = logit
Formula: Mismatch ~ Treatment + (1 | ID) + (1 | Rack) + (1 | Loci) + (1 | Mix)
Data: Analysis.data (Number of observations: 30834)
Draws: 3 chains, each with iter = 1e+05; warmup = 30000; thin = 70;
      total post-warmup draws = 3000

```

Multilevel Hyperparameters:

	Estimate	Est.Error	l-95%	CI	u-95%	CI	Rhat	Bulk_ESS	Tail_ESS
~ID (Number of levels: 280)									
sd(Intercept)	1.31	0.11	1.11	1.54	1.00	2707	3026		

~Loci (Number of levels: 39)

	Estimate	Est.Error	l-95%	CI	u-95%	CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	1.29	0.20	0.96	1.74	1.00	2969	2754		

~Mix (Number of levels: 5)

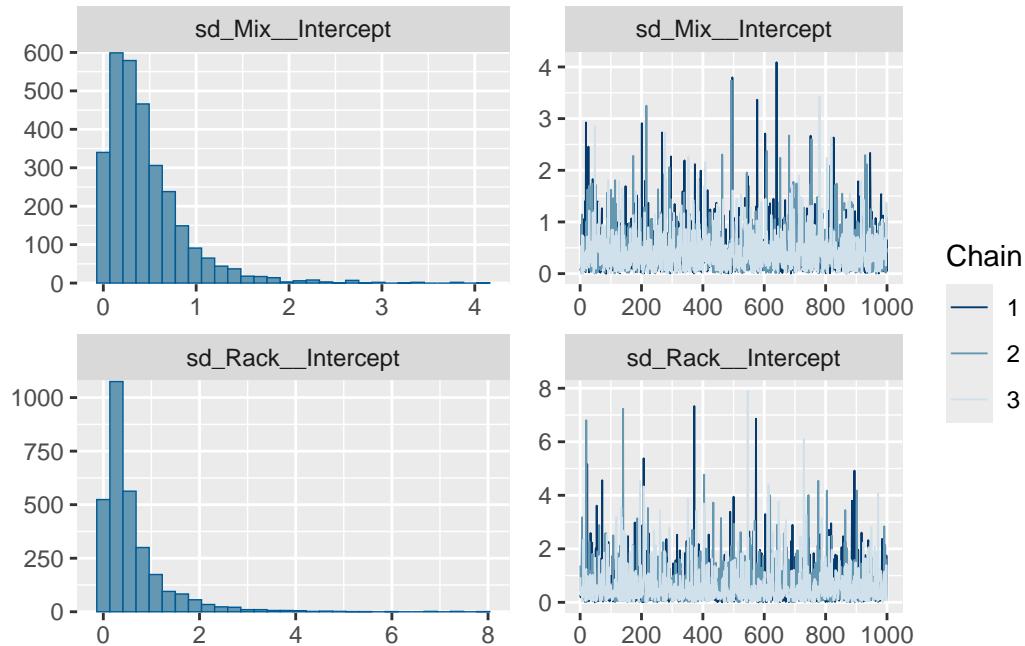
	Estimate	Est.Error	l-95%	CI	u-95%	CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.46	0.43	0.02	1.55	1.00	2995	3005		

~Rack (Number of levels: 3)

	Estimate	Est.Error	l-95%	CI	u-95%	CI	Rhat	Bulk_ESS	Tail_ESS
sd(Intercept)	0.60	0.72	0.02	2.56	1.00	3117	2754		

Regression Coefficients:

	Estimate	Est.Error	l-95%	CI	u-95%	CI	Rhat
Intercept	-6.51	0.66	-7.61	-4.83	1.00		
TreatmentReMuseddetectionplate	-0.32	0.21	-0.75	0.07	1.00		
TreatmentReMusedPCRplate	1.91	0.15	1.61	2.21	1.00		
			Bulk_ESS	Tail_ESS			
Intercept	3004	2878					
TreatmentReMuseddetectionplate	2900	2845					



Treatment ReMused PCR plate 2992 2963

Draws were sampled using `sample(hmc)`. For each parameter, `Bulk_ESS` and `Tail_ESS` are effective sample size measures, and `Rhat` is the potential scale reduction factor on split chains (at convergence, `Rhat = 1`).

Results

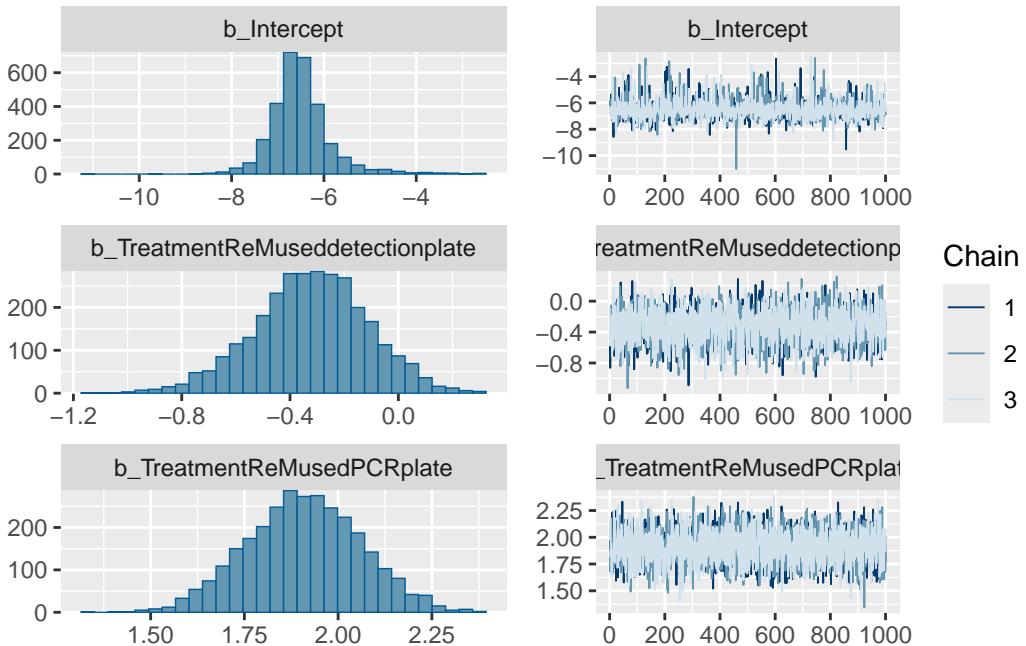
281 samples were included in the analysis. On rack 154 the two samples from 2007-2008 were excluded, due to lack of DNA for the last treatment. Likewise for two of the positive controls on R156.

Tables

Percentage of per-reaction missing data for the internal control, reused detection plate and reused PCR plate treatment groups.

Per-reaction genotyping error rates for the internal control, reused detection plate and reused PCR plate treatment groups, calculated relative to the standard protocol treatment group.

Point beta estimates and 95% confidence intervals from the Bayesian logistic mixed effect model testing for the effects of the fixed effect “treatment” on genotype errors while controlling for the random effects sample ID, DNA plate, multiplexed reaction and loci.



Genotype errors

Fixed effects

Internal control -6.569 [-7.610, -4.830]

Reused detection plate -0.318 [-0.751, 0.071]

Reused PCR plate 1.910 [1.610, 2.210]

Random effects

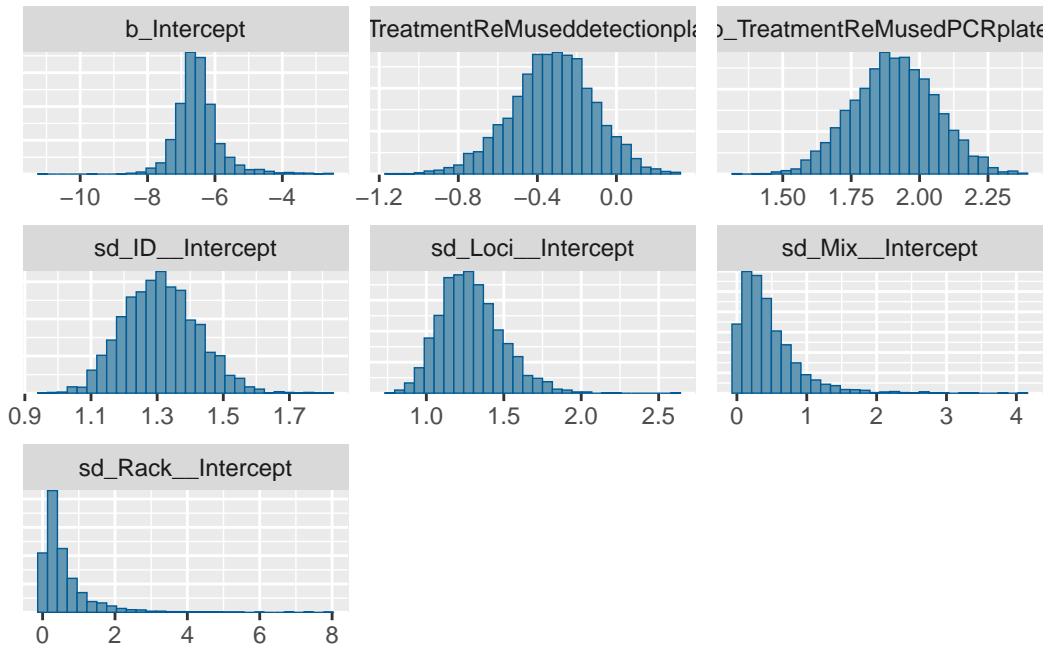
Sample ID 1.306 [1.106, 1.539]

DNA plate 0.374 [0.022, 2.558]

Multiplexed reaction 0.348 [0.017, 1.553]

Locus 1.273 [0.958, 1.739]

Number of observations 30,834

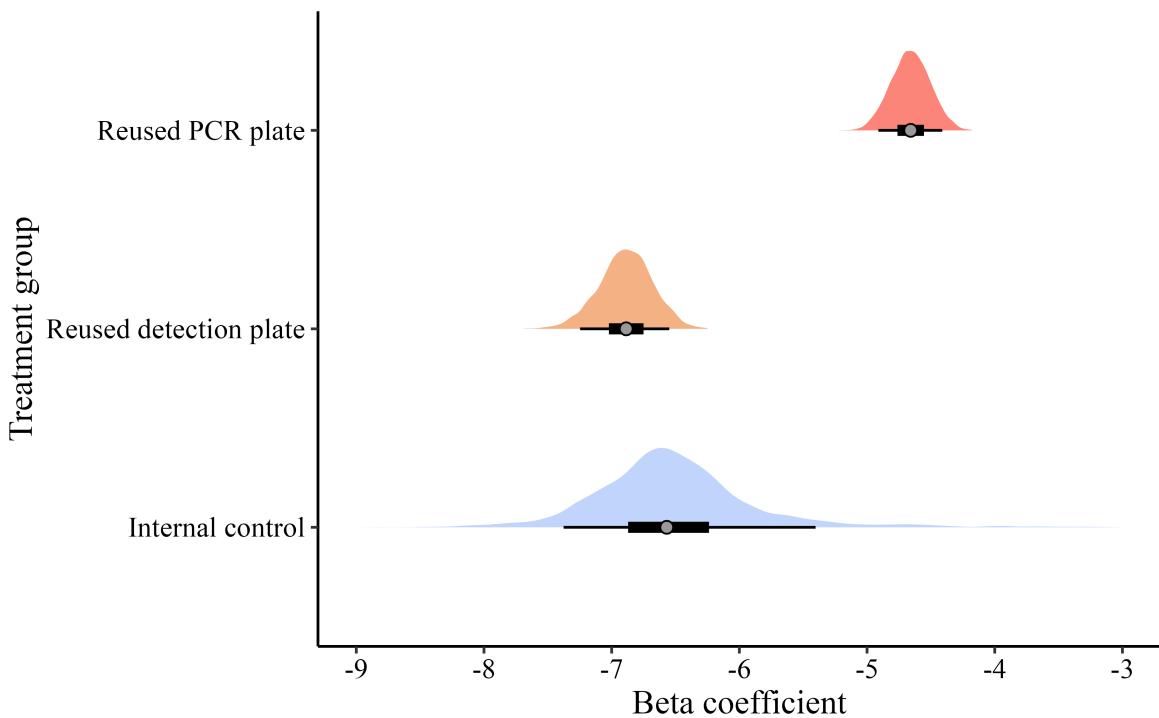


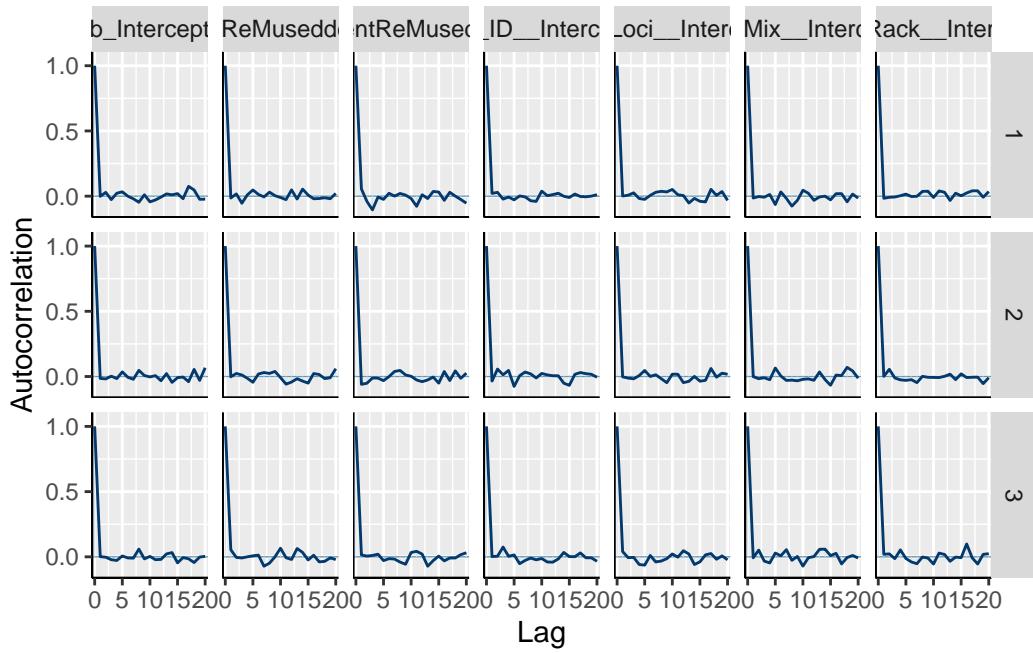
Figures Posterior distributions of the beta coefficients of the internal control (blue), reused detection plate (orange) and reused PCR plate (red) treatment groups on genotyping errors. 50% and 90% confidence intervals are represented with thick and thin black lines respectively.

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.

i Please use `linewidth` instead.

Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as of ggplot2 3.3.4.





Supporting Information

Data

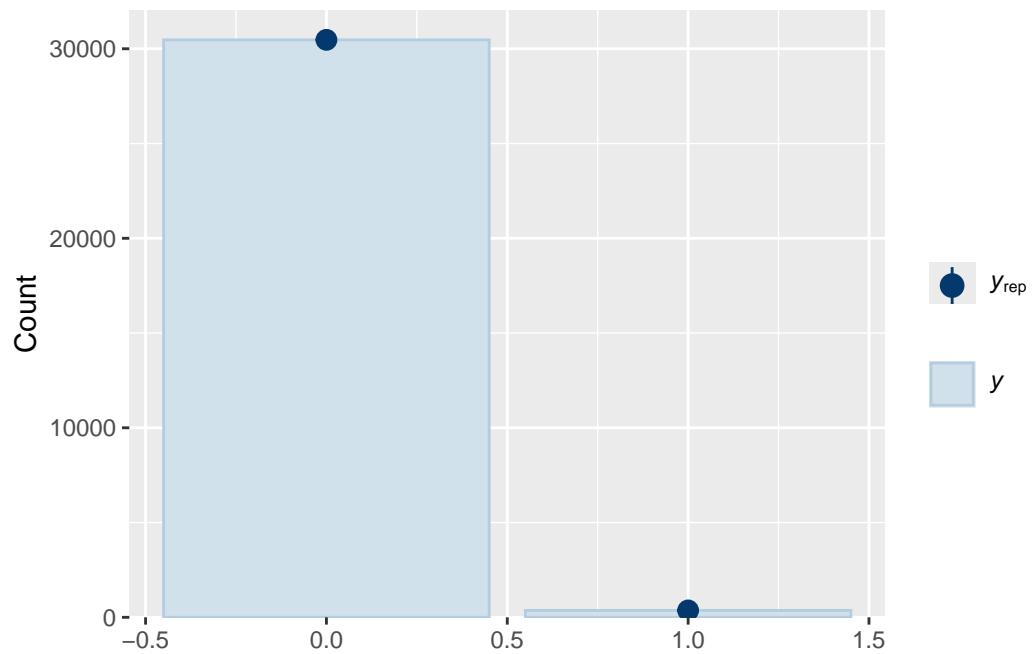
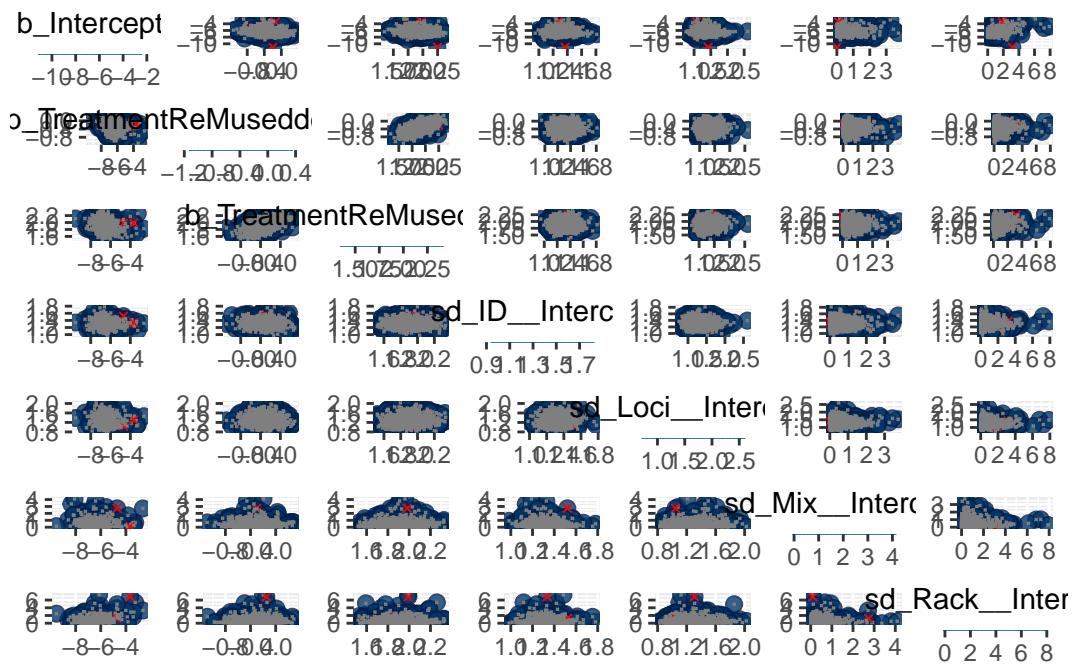
The data for this project are the raw sequencing reads from the ABI 3730xl capillary sequencer. Each files is identified with the following name structure: RackX_mixY_Z. Where X gives the rack number, Y the mix and Z the treatment (1 = standard procedure, 2 = internal control, 3 = Re-used detection plate and 4 = Re-used PCR plate).

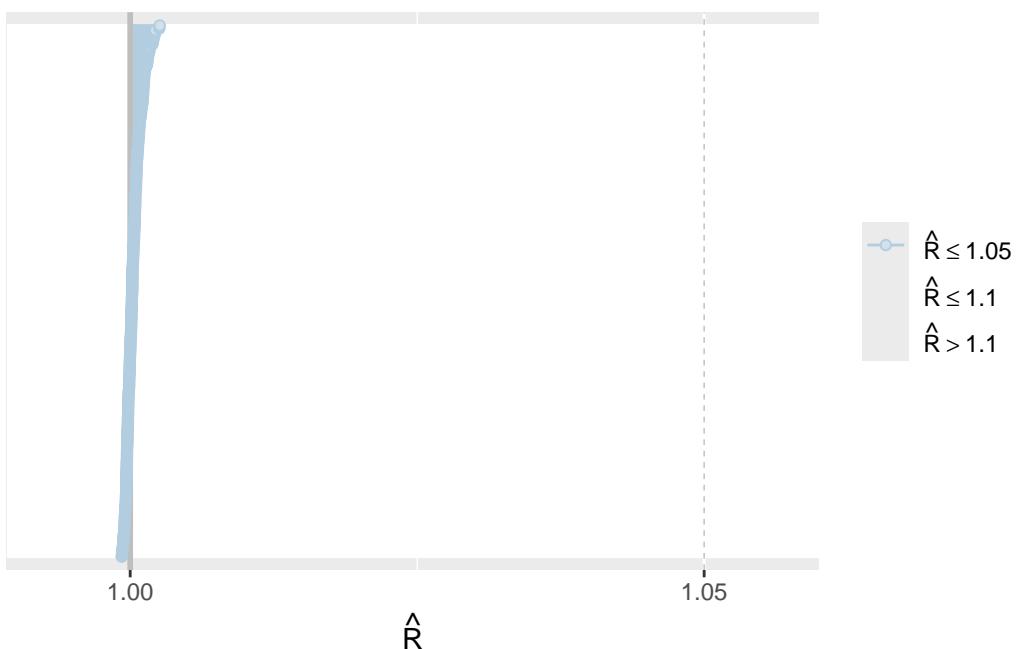
PCR Program

Table 2: PCR program

Stage	No. of cycles	Temperature (°C)	Duration	Process
1	1	94	5 minutes	Heat up
2	28	94	30 seconds	Denaturation
		60/53	90 seconds	Annealing
		72	30 seconds	Extension
3	1	60/53	30 minutes	Annealing
4	1	10	Hold	Cool down

Note: Annealing temperatures are mix-specific.





Treatment	No. of single-locus genotypes	No. of missing data	Missing data %
Internal control	10959	322	2.938
Reused detection plate	10959	248	2.263
Reused PCR plate	10959	1472	13.432

Figure 2: Missing data

Treatment	No. of single-locus genotypes	No. of mismatches	Genotype error rate
Internal control	10.541	57	0.005
Reused detection plate	10.599	42	0.004
Reused PCR plate	6.159	210	0.034

Figure 3: Error rates