Sustainable genetics project

Script

Ane Liv Berthelsen

2023-11-20

Contents

Description **1**

Methods **1**
Missing data 3
Error rate calculation 3

Results **7**

Description

Scientific research is a massive endeavor, consuming vast quantities of single-use plastics. The sustainable genetics project aim to empirically test the re-usability of lab plastic wear, specifically the plastic racks used for PCRs and detection. The project was based around the protocol for Antarctic fur seal genotyping using microsatellites. In this protocol, extracted DNA is first amplified using PCR and since transfered to detection plates for sequencing. The plates used for detection contain only a small diluted concentration of DNA, while the PCR plates have been used in the PCR process and therefore contain a high concentration of DNA. Therefore, we assume that the potential risk of contamination from a previously used plate would be higher in the PCR step.

Methods

The 285 samples were collected from Bird Island, South Georgia (54°00024.800 S, 38°03004.100 W) during the austral summer of 2020-2021 (The 3 repeats on each rack are from 2015-2016, and 2 samples on R154 are from 2006-2007). Tissue samples were collected from the flipper or umbilical cord of Antarctic fur seal pups and stored in 20% dimethyl sulphoxide saturated with sodium chloride at -20°C. The DNA was extracted using a standard chloroform-isoamylalcohol protocol and genotyped at 39 microsatellite loci separated into 5 mixes for improved scoring. PCR amplification was performed using a Type It Kit (Qiagen) with the following program:

Table 1: 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 mastermix-specific.

The PCR product was transferred to and diluted on detection plates (Applied Biosystems™ MicroAmp™ Optical 96-Well Reaction Plate, Thermo Fisher Scientific, Waltham, MA, USA) before being resolved by electrophoresis on an ABI 3730xl capillary sequencer (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA). Each plate contained three positive controls to ensure standardisation of microsatellite scoring across plates. Allele sizes were scored automatically using GeneMarker v. 2.6.2 (SoftGenetics, LLC., State College, PA, USA) and manually inspected and corrected when necessary.

To assess the re-usability of both PCR plates and detection plates, the samples were placed on a master plate from which the samples were transferred to PCR plates (PCR trays ROTILABO® 96 well, Standard, half frame, ROTH SELECTION, Germany). The samples were distributed on a total of 3 master plates. The samples were subjected to four treatments: standard procedure, internal control, re-used PCR plate and re-used detection plate. This is illustrated by Figure 1. The re-used plates all originated from the 'standard procedure' round. Thereby, we could ensure knowledge of the sample previously contained within a specific well. Following standard procedure, any locus with more than 30% missing data is gapfilled separately to minimize missing data.

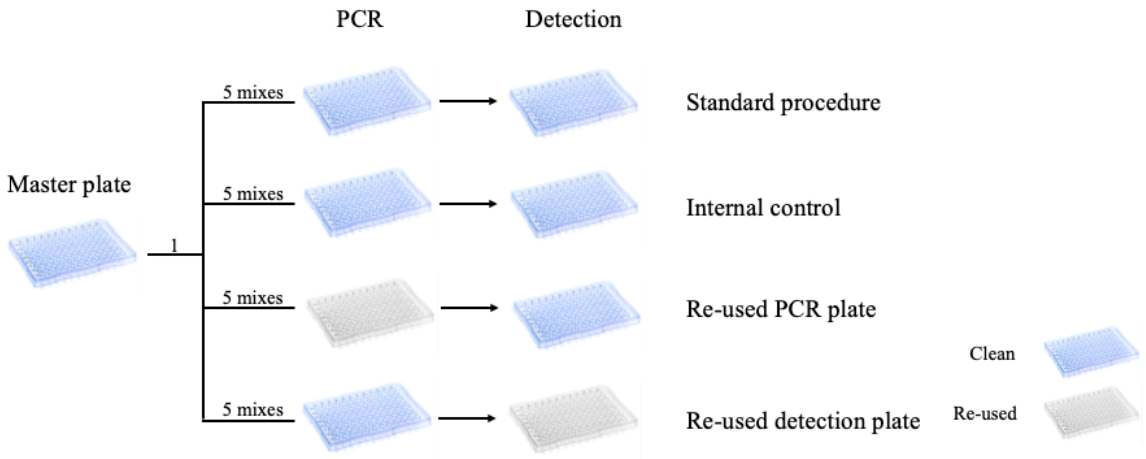

Figure 1: Schematic overview

Washing protocol

After the initial use, the racks were washed using the following steps. First, each plate was individually rinsed with distilled water and emptied 10 times, before submerging in soap water for two hours. After soaking, each plate was again rinsed and emptied before left on a paper towel over night to dry.

Quantification of genotyping errors

To estimate the genotyping error rate, the 'standard procedure' treatment was first compared to the 'internal control' treatment to establish a baseline error rate. Any discrepancy between the two sets of genotypes was flagged as an error. The error rate was evaluated both for each allele and for the genotype on a locus-specific level and across all loci. After establishing a baseline error rate, the same comparision was preformed between the 'standard procedure' treatment and the two re-use treatments independently. In these cases, any discrepancy was first explored for the possibility of contamination, before flagged as an error, if no contamination was detectable.

Identification of contamination

The samples were each repeated for all four treatments. In the case of a discrepancy between the original genotype and the 're-used PCR plate' or 're-used detection plate' treatment, the new genotype was compared to

Table 2: Failed genotypes

(a) Rack 154

Mix	R154_R1	R154_R2	R154_R3	R154_R4
Mix1	108	88	58	72
Mix2	60	40	60	56
Mix3	36	58	52	108
Mix4	40	32	42	58
Mix5	64	78	76	166
Total	308	296	288	460

(b) Rack 155

Mix	R155_R1	R155_R2	R155_R3	R155_R4
Mix1	8	26	50	66
Mix2	14	40	12	52
Mix3	14	31	26	112
Mix4	30	22	10	96
Mix5	10	132	16	120
Total	76	251	114	446

(c) Rack 156

Mix	R156_R1	R156_R2	R156_R3	R156_R4
Mix1	40	44	52	66
Mix2	46	44	60	102
Mix3	28	26	50	84
Mix4	42	34	44	52
Mix5	52	52	70	84
Total	208	200	276	388

the genotype of the individual previously genotyped in the same well. If the genotype matched the genotype of the original individual, the mis-matched was marked as contamination.

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 PCR plate and 4 = Re-used detection plate).

```
<environment: R_GlobalEnv>
```

Missing data

Before starting the error rate evaluation, the no. of genotypes that could not be scored for each round and mix per rack is tabulated (See Table 2). From this, we know that a total of 3311 alleles could not be scored. In Figure 2 and Figure 3, the location of the missing data on the plate is visualized. More purple hues correspond to more missing loci for that particular samples, and we can conclude that certain samples seem to have failed almost completely. This could hint at poor DNA quality. The data is arranged in the plot to match the layout of the actual plate. Through visual inspection, there might be a slight tendency for the missing loci to be located more towards the edges of the plates compared to the center.

Error rate calculation

The error rate is calculated by comparing all scored alleles and genotypes from the standard procedure within one rack with the other treatments starting with the internal control. The internal control is included as a baseline

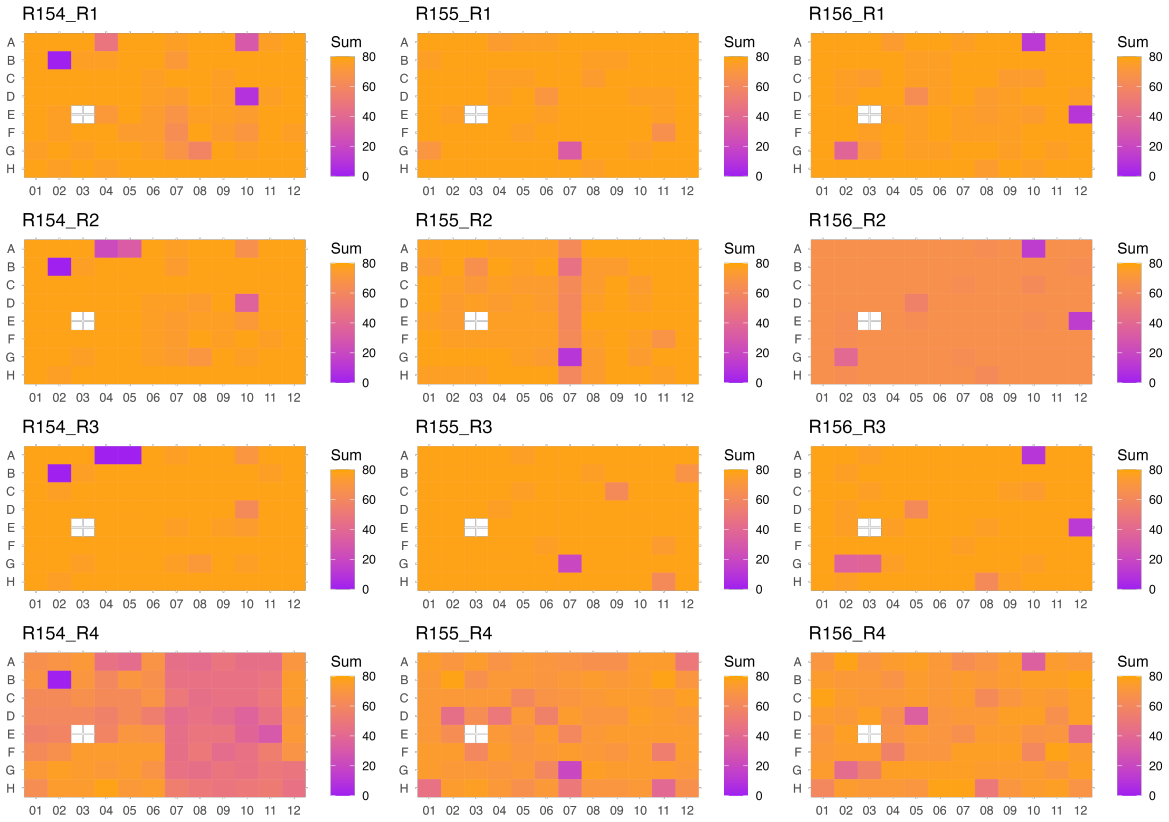

Figure 2: Heatmap of missing data before gapfilling

for error rates to have a comparable metric to evaluate if the reused plates inflate the error rate. The error rate is calculated both as per-allele genotyping error rate and per locus genotyping error rate.

Per-allele genotyping error rate

Inspect mismatches for contamination

The comparison between the different treatments and the standard procedure yielded a list of mismatches. To identify whether these mismatches in the case of the 're-used PCR plate' and 're-used detection plate' treatment was due to contamination from the samples previous found in the same well, the new genotype was compared to the original genotype scored from that well. If the genotype matched the original genotype, the mis-match was marked as potential case of contamination.

```
     Rack Control Cases                Treatment
1 R154_R3 R156_R1     5              Re-used PCR
3 R155_R3 R154_R1     1              Re-used PCR
5 R156_R3 R155_R1     7              Re-used PCR
2 R154_R4 R156_R1    39 Re-used detection plate
4 R155_R4 R154_R1    90 Re-used detection plate
6 R156_R4 R155_R1    49 Re-used detection plate
```

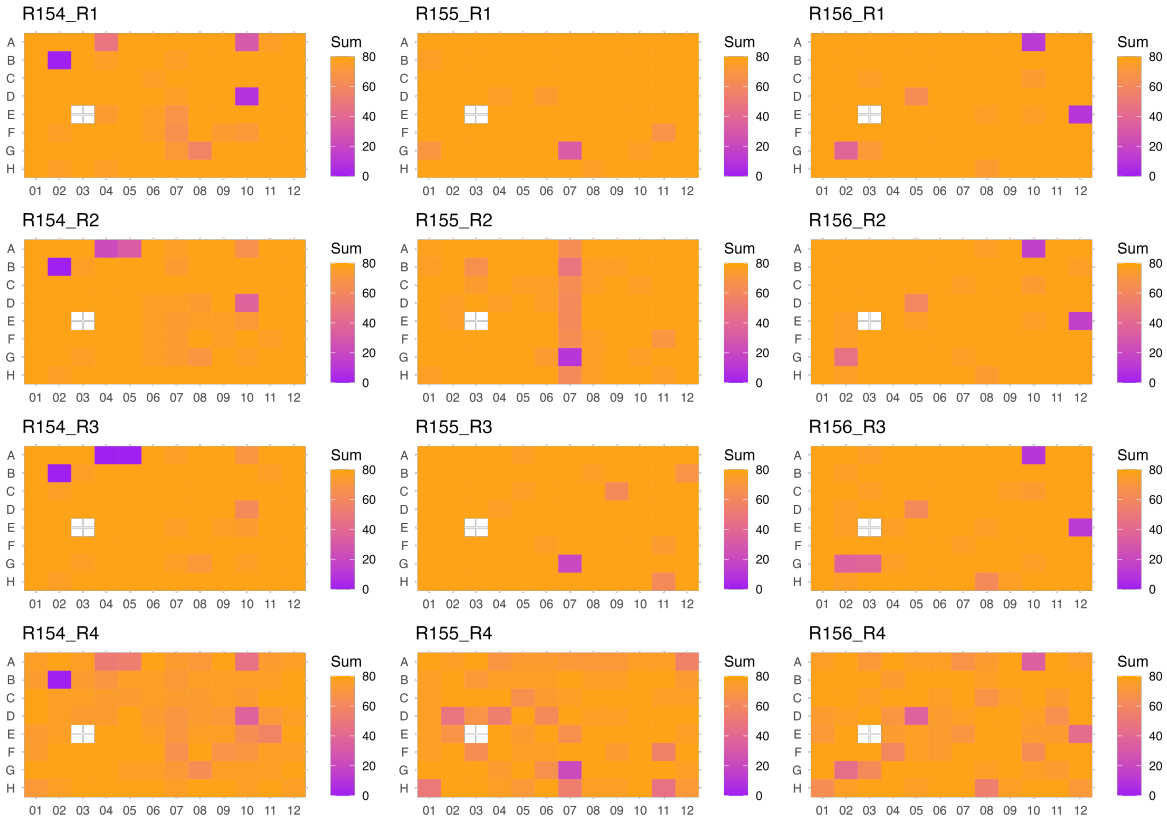

Figure 3: Heatmap of missing data after gapfilling

Analysis

```
Analysisformat <- function(DF) {
  Base <- DF[1:2]
  colnames(Base) <- c("Well", "ID")
  Base$Row <- substring(Base$ID, 1, 1)
  Base$Column <- substring(Base$ID, 2, 3)
  Base$Column <- as.integer(Base$Column)
  Base$ID <- gsub("^\\w{3}_", "", Base$ID)
  Base$Rack <- deparse(substitute(DF))
  Round <- gsub("R\\d{3}|_", "", deparse(substitute(DF)))
  if (Round == "R1"){
    Treatment <- "Standard procedure"
  } else if (Round == "R2") {
      "Internal control"
  } else if (Round == "R3") {
    Treatment <- "Re-used PCR"
  } else {
    Treatment <- "Re-used detection plate"
  }
  Base$Treatment <- Treatment
  return(Base)
}
```

Table 3: Allelic mismatches

Rack 1	Rack 2	Treatment	No. of mismatches	Allelic error rate	Genotype error rate
R154_R1	R154_R2	Internal control	9	0.0012980	0.0020190
R155_R1	R155_R2	Internal control	21	0.0029441	0.0050470
R156_R1	R156_R2	Internal control	45	0.0064157	0.0128315
R154_R1	R154_R3	Re-used PCR	17	0.0024453	0.0037399
R155_R1	R155_R3	Re-used PCR	10	0.0013759	0.0024766
R156_R1	R156_R3	Re-used PCR	26	0.0037389	0.0069025
R154_R1	R154_R4	Re-used detection plate	69	0.0102192	0.0201422
R155_R1	R155_R4	Re-used detection plate	110	0.0158547	0.0311329
R156_R1	R156_R4	Re-used detection plate	67	0.0098966	0.0183161

Table 4: Potential contamination cases

Rack	Control	Cases	Treatment
R154_R3	R156_R1	5	Re-used PCR
R154_R4	R156_R1	39	Re-used detection plate
R155_R3	R154_R1	1	Re-used PCR
R155_R4	R154_R1	90	Re-used detection plate
R156_R3	R155_R1	7	Re-used PCR
R156_R4	R155_R1	49	Re-used detection plate

```
test <- Analysisformat(R154_R1)
```

To determine if treatment had a significant effect on the following three measures; no. of mismatches, allelic error rate and potential cases of contamination, we performed ANOVAs followed by a post hoc test for pairwise comparisons for the three treatment groups measured against the standard procedure treatment.

```
            Df Sum Sq Mean Sq F value  Pr(>F)
Treatment    2   7442    3721   11.28 0.00927 **
Residuals    6   1979     330
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = `No. of mismatches` ~ Treatment, data = Mismatches)

$Treatment
                                              diff        lwr       upr
Re-used detection plate-Internal control  57.000000   11.50546 102.49454
Re-used PCR-Internal control              -7.333333  -52.82787  38.16121
Re-used PCR-Re-used detection plate      -64.333333 -109.82787 -18.83879
                                             p adj
Re-used detection plate-Internal control 0.0199402
Re-used PCR-Internal control             0.8764882
Re-used PCR-Re-used detection plate      0.0115670
```

The total no. of mismatches between the standard procedure and the additional treatments could be significantly determined by treatment. Further exploration using a Tukey pairwise comparison revealed that this treatment effect was driven by a significant difference in no. of mismatches in the 're-used detection plate' treatment when compared to both the internal control and the re-used PCR treatment. When comparing those two treatments to

each other, no significant difference could be found.

Results

Will come later

Figure 4 shows 4 graphic panels. A visualize the total no. of mismatches observed when comparing the standard procedure treatment of each rack with the other treatments. B the total number of potential contamination cases, while C and D show the per allele and per locus genotype error rate, respectively.

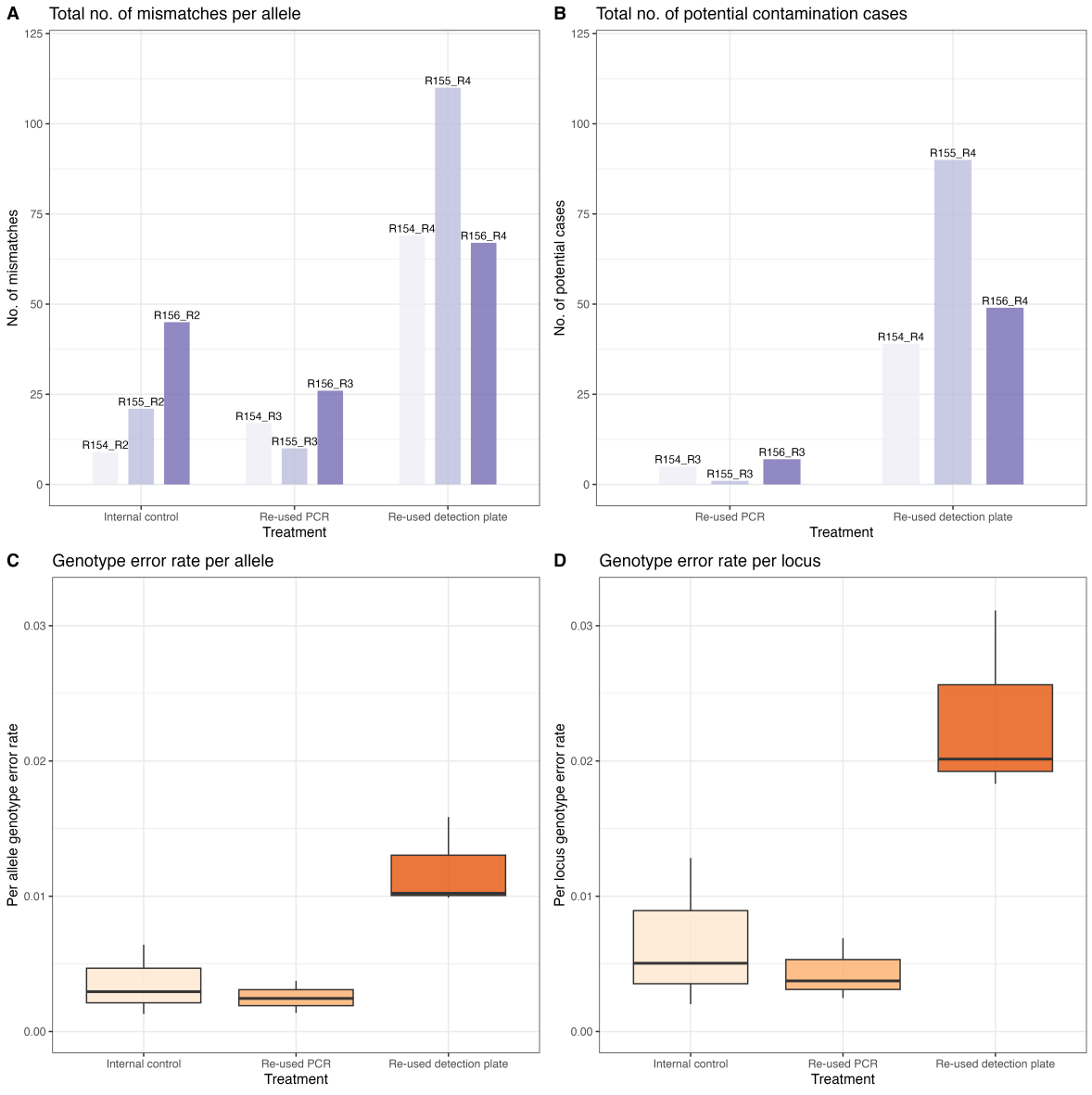

Figure 4: Mismatches and error rates

We found a total of 374 mismatches between all treatments. A subset of these (191) were potential contamination cases.