Sustainable genetics project

Ane Liv Berthelsen

2023-11-20

Contents

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 274 samples were collected from Bird Island, South Georgia (54°00024.800 S, 38°03004.100 W) during the austral summer of 2020-2021. 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 before being resolved by electrophoresis on an ABI 3730xl capillary sequencer (Applied Biosystems, 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. 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.

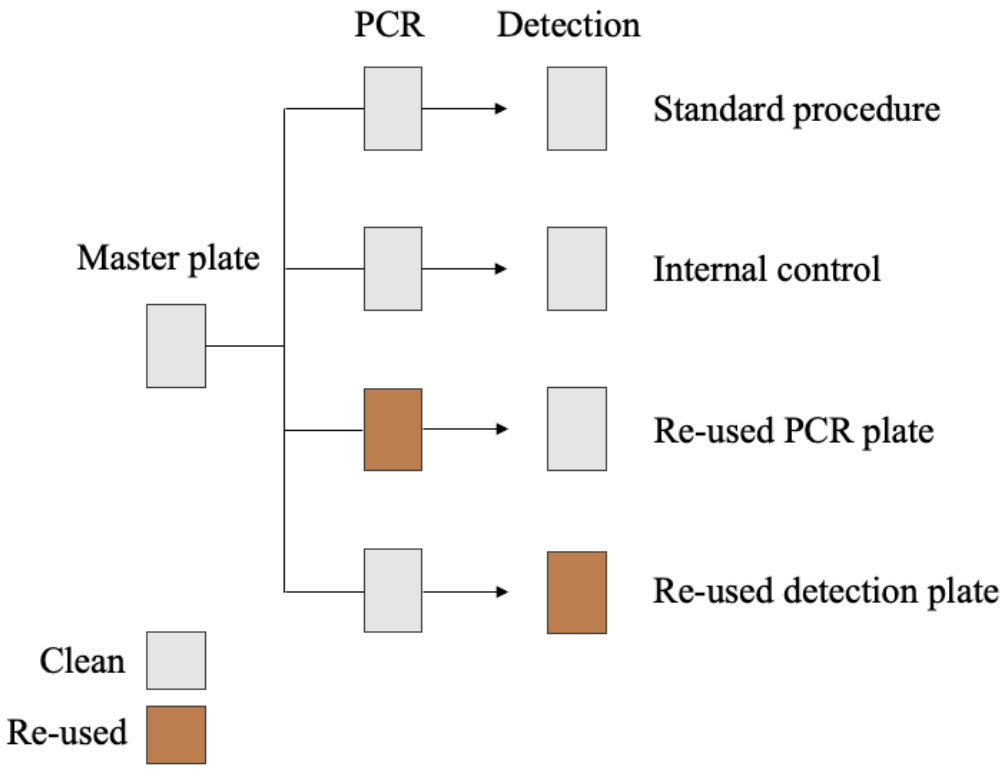

Figure 1: Schematic overview

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 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).

Before starting the error rate evaluation, the no. of genotypes that could no be scored for each round and mix per rack is tabulated (See Table 2).

Table 2: Failed genotypes

Mix	R154_R1	R154_R2	R154_R3	R154_R4	R155_R1	R155_R2	R155_R3	R156_R1	R156_R2
Mix1	108	88	58	1014	8	26	50	40	44
Mix2	60	42	61	164	14	42	12	46	46
Mix3	36	58	52	108	14	31	26	28	26
Mix4	40	32	42	548	30	22	10	42	1140
Mix5	146	110	94	228	62	266	18	153	64
Total	390	330	307	2062	128	387	116	309	1320

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 for error rates to have a comparable metric to evaluate if the reused plates inflate the error rate.

Table 3: Allelic mismatches

Rack 1	Rack 2	Treatment	No. of mismatches	Allelic error rate
R154_R1	R154_R2	Internal control	9	0.0012662
R155_R1	R155_R2	Internal control	18	0.0025080
R156_R1	R156_R2	Internal control	31	0.0050284
R154_R1	R154_R3	Re-used PCR	16	0.0022634
R155_R1	R155_R3	Re-used PCR	10	0.0013492
R154_R1	R154_R4	Re-used detection plate	137	0.0251931

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.

Table 4: Potential contamination cases

Rack	Control	Cases
R154_R3	R156_R1	6
R154_R4	R156_R1	104
R155_R3	R154_R1	2

Analysis

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  12088    6044   69.03 0.0031 **
Residuals    3    263      88
---
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  117.666667   72.51654 162.81679
Re-used PCR-Internal control               -6.333333  -42.02764  29.36097
Re-used PCR-Re-used detection plate      -124.000000 -171.88894 -76.11106
                                             p adj
Re-used detection plate-Internal control 0.0034145
Re-used PCR-Internal control             0.7590749
Re-used PCR-Re-used detection plate      0.0034794
```

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 eachother, no significant difference could be found. The same was true for the error rate.

```
            Df    Sum Sq   Mean Sq F value  Pr(>F)
Treatment    2 0.0004313 2.157e-04   83.29 0.00235 **
Residuals    3 0.0000078 2.590e-06
---
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 = `Allelic error rate` ~ Treatment, data = Mismatches)

$Treatment
                                                diff          lwr          upr
Re-used detection plate-Internal control  0.022258893  0.014494705  0.030023082
Re-used PCR-Internal control             -0.001127909 -0.007266039  0.005010221
Re-used PCR-Re-used detection plate      -0.023386802 -0.031621968 -0.015151636
                                             p adj
Re-used detection plate-Internal control 0.0025877
Re-used PCR-Internal control             0.7454382
Re-used PCR-Re-used detection plate      0.0026598
```

To explore the effect of treatment on the number of potential contamination cases, a non-parametric wilcoxon test was used.

Note: the test is performed on only one case of the 're-used detection plate' treatment and therefore not yet considered informative.

```
    Wilcoxon rank sum exact test

data:  Cases by Treatment
W = 2, p-value = 0.6667
alternative hypothesis: true location shift is not equal to 0
```

Results

Will come later

Figure 2 shows the total no. of mismatches observed when comparing the standard procedure treatment of each rack with the other treatments.

We found a total of 221 mismatches between all treatments. A subset of these (112) were potential contamination cases.

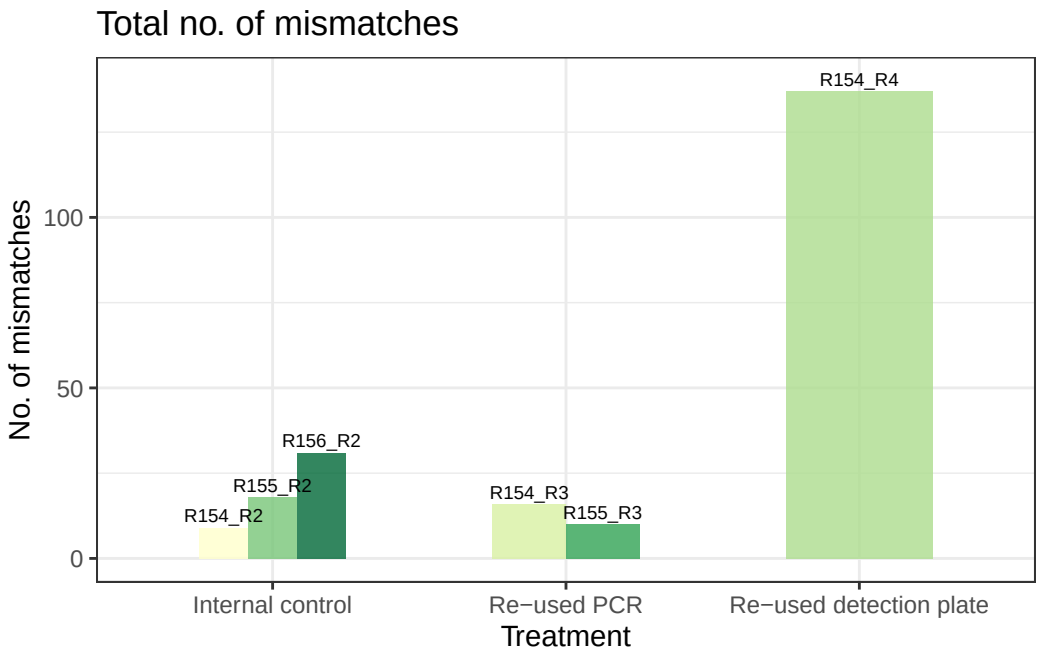

Figure 2: Allelic mismatches

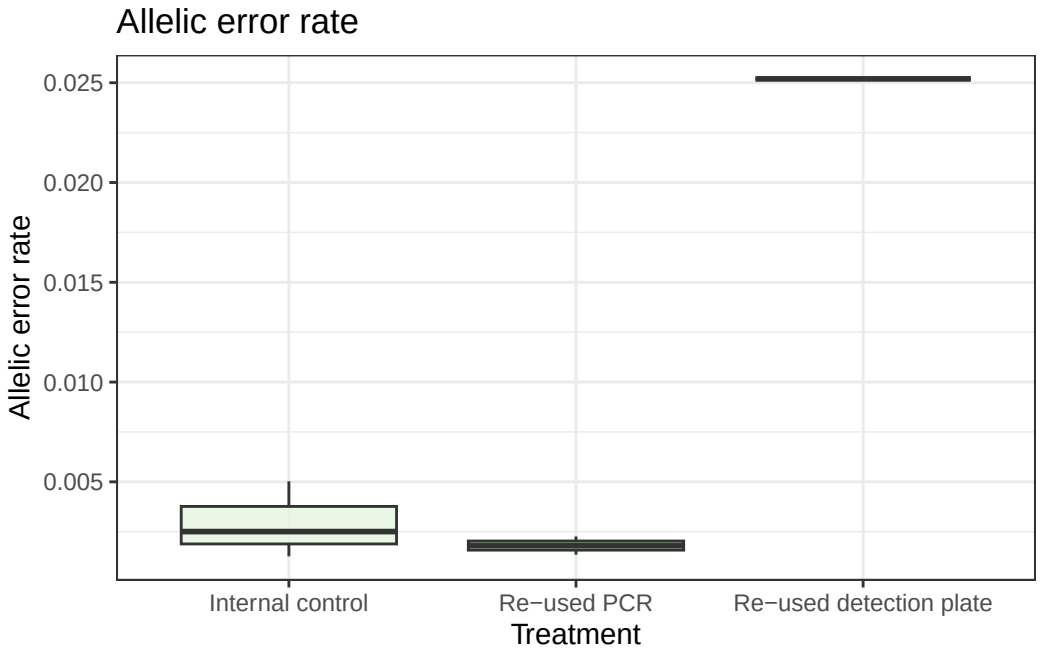

Figure 3: Allelic error rate

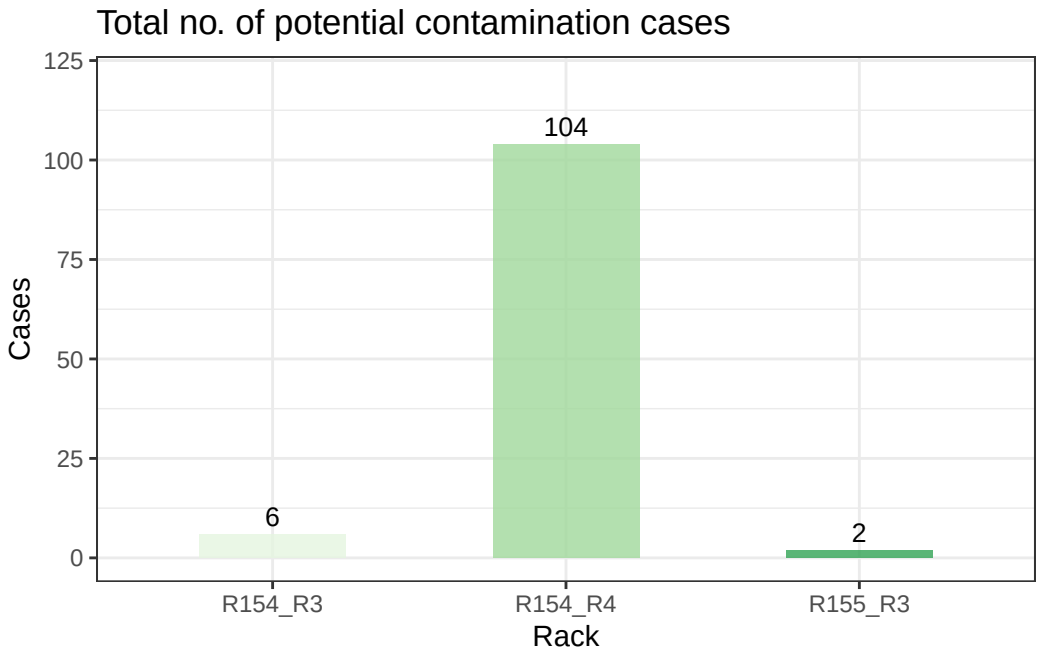

Figure 4: Potential contamination