|  |  |  |
| --- | --- | --- |
| **Pool #** | **PET-PCR Ct**  [average] | **Parasites per uL**  [estimate] |
| 1 | 22.20 | 30,000 |
| 2 | 24.00 | 10,000 |
| 3 | 25.70 | 3000 |
| 4 | 27.90 | 600 |
| 5 | 29.80 | 150 |
| 6 | 31.80 | 40 |
| 7 | 34.10 | 10 |
| 8 | 36.00 | 3 |
| 9 | 37.80 | 1.0 |
| 10 | 40.0 | 0.4 |

**Table 1. Pooling of samples based on parasitemia levels.** Ten sets of 10:1 ratio pools were made based on similar Ct (estimated parasitemia) values. Ct values with corresponding parasites per uL are shown.



**Figure 1. Sequencing outcome for *dhfr* and *mdr1* genes between individual and pooled sequenced samples.** Sequencing success (blue dots) or failure (orange dots), measured by high quality reads obtained. Genes are shown on the *x* axis and parasites per uL are shown on the *y* axis. Individual sequenced samples = 100 total samples; same samples, pooled at 10:1 ratio = 10 total sequenced samples.



Figure 2. **Read depth coverage comparison for *dhfr* and *mdr1* SNPs associated with anti-malarial drug resistance between individual and pooled sequenced samples.** SNP loci are shown on the *x* axis and read depth coverage on *y* axis. Orange, pooled sequenced samples; green, individual sequenced samples. Violin plot distribution values: white dot = median; thick black bar = interquartile range (25th, 50th, 75th); thin black bar = rest of data distribution; kernel density= shows distribution shape of data. Wider sections of plot represent higher probability that samples will take on given value and skinner sections a lower probability.

**Dhfr**

method

individual 0.25 106.5

0.50 280.5

0.75 427.0

1.00 733.0

pooled 0.25 70.0

0.50 88.0

0.75 130.0

1.00 294.0

**Mdr1**

method

individual 0.25 39.00

0.50 79.00

0.75 122.00

1.00 458.00

pooled 0.25 7.00

0.50 62.00

0.75 77.75

1.00 106.00



Figure 3. **Allele frequency comparison between individual and pooled sequenced samples for *dhfr* and *mdr1*.** Parasitemia (parasites/uL) are shown on *x* axis. Percent allele frequencies are shown on the *y* axis. Blue dots, individual sequenced samples allele frequency (expected allele frequency); orange dots, pooled sequenced samples allele frequency (observed allele frequency). Blue/orange line = best fit, blue/orange highlighted region = confidence bounds.

* Negative relationship between (dependent = allele freq, and independent = parasitemia);
* With lower parasitemia the expected allele frequency calls (based on individual sequenced samples) breaks down suggesting that at lower parasitemia levels the pooled method could underestimate, or completely miss particular alleles as compared to individual sequenced samples == likely due to lower overall coverage (evident from violin plot and data distribution).