Balancing Mitochondrial and Genomics Sequencing Coverage in Targeted GBS Applications

Angela Burrell[1], Haktan Suren[1], Prasad Siddavatam[1], and Rick Conrad[1]
[1]Thermo Fisher Scientific, 2130 Woodward Street, Austin, TX, USA, 78744.

INTRODUCTION

Mitochondrial SNPs are frequently used in targeted genotyping-by-sequencing (GBS) for the identification of individuals in forensic applications. However, having hundreds of copies of mitochondrial DNA (mtDNA) per cell as compared to only two copies of genomic DNA (gDNA) poses a challenge in next generation sequencing. The relatively high abundance of mtDNA can lead to over-representation during sequencing and poor coverage of gDNA targets. We have developed a strategy for combining mtDNA and gDNA targets in the same SNP panel while maintaining good coverage balance across all amplicons.

ABSTRACT

Two strategies were tested for optimizing co-detection of mtDNA and gDNA SNPs during sequencing reactions. We tested serially diluting either the mtDNA AgriSeq™ primer panel or the mtDNA amplicons between 1:4X and 1:200X relative to the gDNA. Libraries were then prepared using the standard AgriSeq™ HTS Library Kit and sequenced on the Ion GeneStudio S5 Sequencer using a 540 chip. Mean coverage was compared across all amplicons to find a dilution that resulted in balanced coverage between the mtDNA and gDNA targets. The best results were seen when using a 1:8X dilution of the mtDNA AgriSeq™ primer panel. Diluting the primer panel instead of the resulting amplicons also had the advantage of allowing all amplification reactions to occur in a single well instead of separate reactions for mtDNA and gDNA amplification.

The optimized protocol was tested with a 118-SNP panel (14 mtDNA and 104 gDNA targets) against 102 canine samples in duplicates. The mean call rate was >99% and the replicate genotype concordance was >99.8%. Coverage was very well balanced between mtDNA and gDNA targets with <0.2 fold change in mean coverage between them. In conclusions, we have developed an optimized workflow to balance coverage between mitochondrial and genomic targets for optimal performance from both DNA sources.

MATERIALS AND METHODS

We utilized an 118 marker customer AgriSeq™ panel for testing. The panel contained 104 genomic DNA SNP targets and 14 mitochondrial DNA SNP targets.

Initial optimization tests were performed using eight (8) DNA samples tested in replicates of n=36 to determine the optimal protocol and dilution factor. We tested two methods to balance coverage between mtDNA and gDNA targets using the AgriSeq™ HTS Library Kit workflow (Figure 1).

Option 1 Amplicon Dilution: We separated the amplification of the mtDNA and gDNA targets into separate reactions and then diluted the mtDNA amplicons between 1:10X and 1:200X . The mtDNA amplicons were the combined with the gDNA amplicons and the library prep was completed.

*Option 2 Panel Dilution: We d*iluted only the mtDNA primers between 1:4X and 1:200X prior to addition into the initial amplification reaction. Amplification was performed with the gDNA targets and library prep was completed.

For both methods, we separated the primers for the mtDNA and gDNA targets into separate primer pools so that we could adjust concentrations separately or separate the amplification into different reactions. Option 2 had the advantage of using less reagents since the initial amplification reaction can be performed in a single reaction instead of separating out into two reactions.

As controls, we ran both completely separate library preps where the mtDNA and gDNA libraries preps occurred in independent reactions. We also ran a set of libraries processed using the standard workflow where there was no dilution of the mtDNA amplicons or panel.

Following library preparation, all libraries were pooled into a single sequencing reaction and sequenced on a Ion 540 chip (Figure 2).

We determined the optimal dilution factor was 1:8X as it provided the best balance between mitochondrial and genomic target amplification. Once the optimal protocol and dilution factor were determined, we confirmed our results by testing against n=102 canine blood DNA samples in replicates of n=2.

Figure 1. AgriSeq™ Library Prep workflow

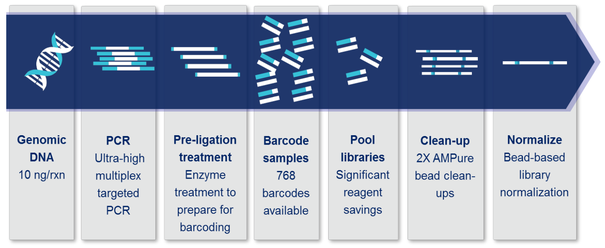


Figure 1. After the initial amplification described in Options 1 and 2, each sample was treated with a Pre-ligation Enzyme to remove residual primer dimers allowing for more efficient sequencing. Samples were ligated with unique barcoded adapters allowing them to be pooled for subsequent clean-up and sequencing while retaining traceability to the original sample during analysis. Libraries were cleaned-up by a two-round AMPure purification. A final bead-based normalization step was performed to ensure each library was at a consistent final concentration suitable for direct input into template prep on the Ion Chef™ instrument.

Figure 2. Complete AgriSeq™ Sequencing Workflow

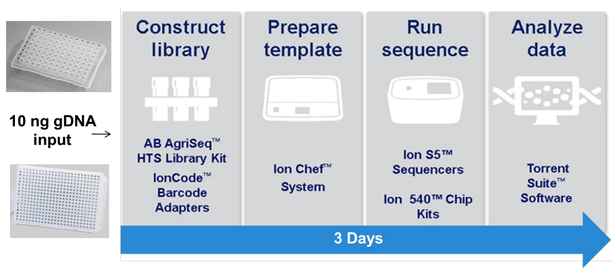


Figure 2. Following library prep, libraries were pooled into a single tube at a 1:1 ratio and run overnight on the Ion Chef™ instrument for template prep. The following day, libraries were sequenced on the Ion S5™ XL instrument and data was analyzed using the Torrent Suite™ Software v5.10. Genotypes for all markers were obtained from the Torrent Variant Caller plugin.

RESULTS

Figure 3. Mean coverage results for gDNA targets

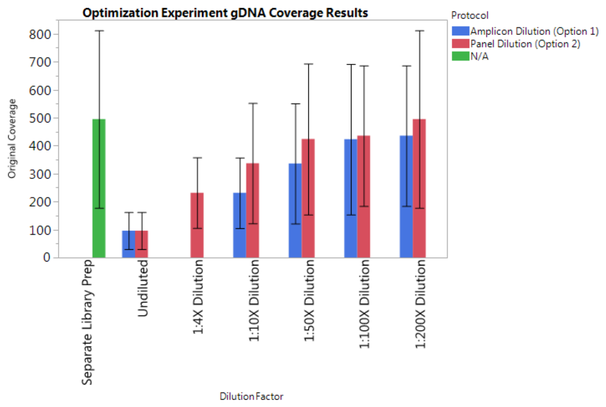


Figure 3. Mean coverage for gDNA amplicons increased as either the mtDNA amplicon or panel dilution increased. Mean gDNA coverage increased from 96X in the undiluted controls to >400X at the 1:200X dilution. Equivalent coverage results were seen between the amplicon and panel dilution conditions.

Figure 4. Mean coverage results for mtDNA targets

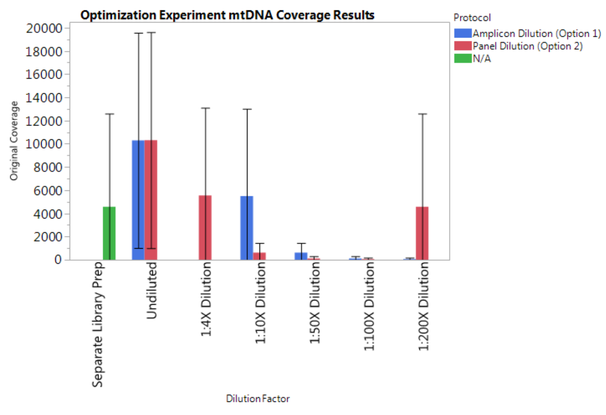


Figure 4. Mean coverage for mtDNA amplicons tended to decrease as the dilution factor increased. Mean mtDNA coverage was 10,000X in the undiluted reactions and ~100X in the 1:100X dilution. For an unknown reason, the mtDNA coverage of the panel dilution samples was significantly higher in the 1:200X dilution reactions.

Figure 5. Sample Call Rate Results

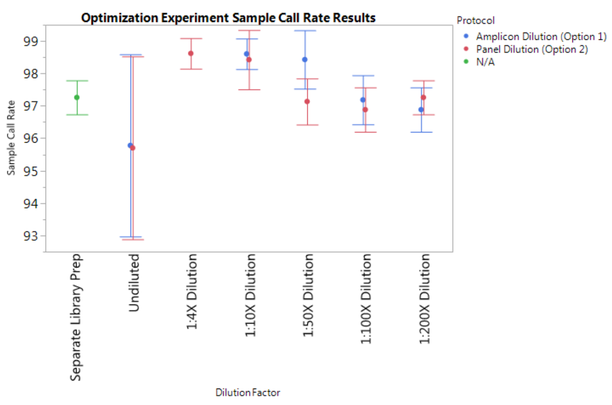


Figure 5. Sample call rates were highest between 1:4X and 1:10X dilution for both protocol options. Sample call rates fell off as the dilutions increased due to a fall-off of mtDNA target detection (see subsequent figure).

Figure 6. Call rate by marker results

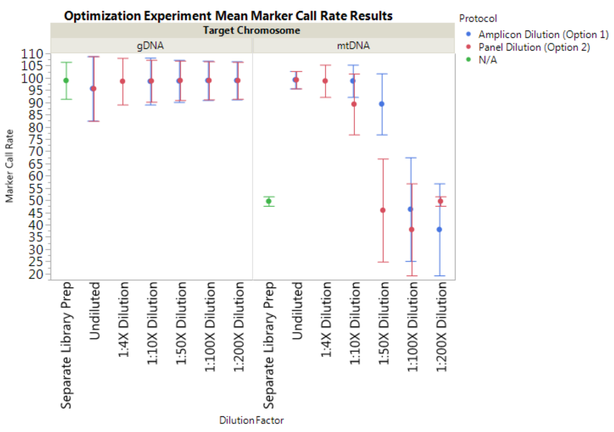
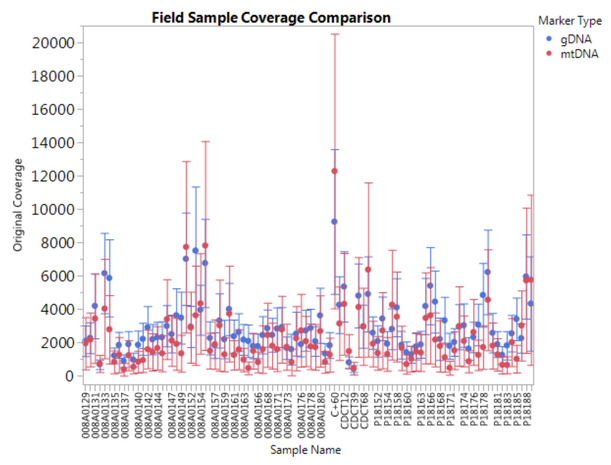


Figure 6. The highest call rates for both the gDNA and mtDNA targets were seen between 1:4X and 1:10X dilutions. For confirmation testing with 102 field samples, we used Option 2 (panel dilution) at a 1:8X dilution since this option did not require separating the initial amplification into two reactions. The call rate for the mtDNA control reaction (separate library prep) likely fell off due to an insufficient number of markers present in the reaction.

Figure 7. Field sample coverage balance



Means for Oneway Anova					
Level	Number	Mean	Std Error	Lower 95%	Upper 95%
gDNA	21420	3067.48	16.240	3035.7	3099.3
mtDNA	2856	2599.72	44.474	2512.5	2686.9

Std Error uses a pooled estimate of error variance

Figure 7. The difference in mean coverage depth between the mtDNA and gDNA targets was very small, <0.2 fold change between them. This demonstrates the excellent balance in coverage we were able to achieve with a 1:8X dilution of the mtDNA panel.

Figure 8. Field sample call rate results

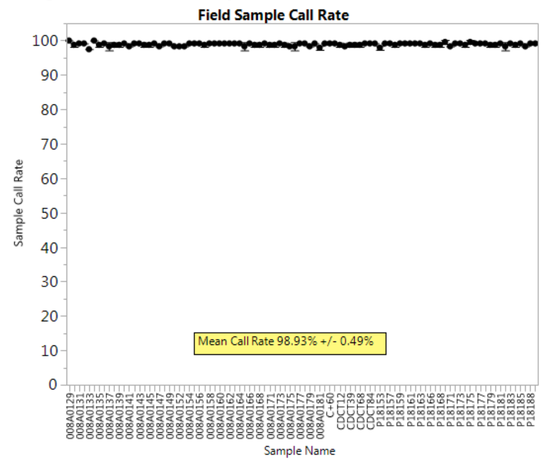


Figure 8. Mean sample call rate was >99% indicating the high level of performance we were able to obtain with the optimized workflow.

Figure 9. Field sample marker call rate results

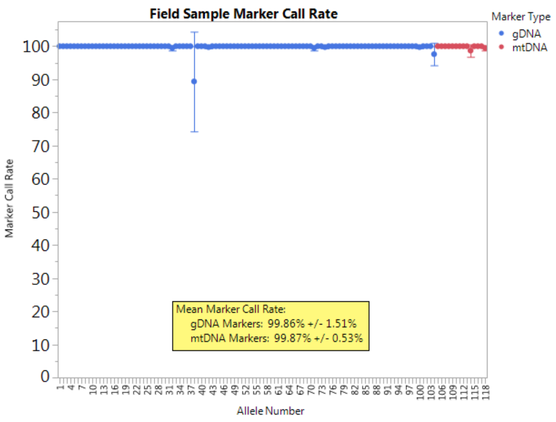


Figure 9. Mean marker call rate was >99% and there was not a significant different in marker call rate between the mtDNA and gDNA marker targets.

CONCLUSIONS

Genomic and mitochondrial targets can be genotyped in the same sequencing reaction without over-representation of the mitochondrial DNA by diluting the mtDNA AgriSeq™ panel prior to the initial amplification reaction. We found that a 1:8X dilution of the mtDNA panel prior to the initial amplification reaction was optimal for this panel to balance detection of all targets and optimize call rates.

Validating with a large panel of field samples, we achieved >99% call rates for both the gDNA and mtDNA marker targets with <0.2 fold change in coverage between the target types.

In conclusion, the AgriSeq™ workflow can be optimized to detect organelle targets with multiple genome copies simultaneously with diploid genomic targets without having to separate the library prep into multiple reactions.

REFERENCES

1. Cariou M, Duret L, Charlat S. How and how much does RAD-seq bias genetic diversity estimates? BMC Evolutionary Biology. 2016;16:240. doi:10.1186/s12862-016-0791-0.

2. Kidd, Kenneth K. et al. Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics . Forensic Science International: Genetics , Volume 12 , 215 – 224.

ACKNOWLEDGEMENTS

We would like to thank Oscar Ramirez Bellido at Vetgenomics SL for providing the field samples used during confirmation testing.

TRADEMARKS/LICENSING