www.GTseek.com
124 Main Ave. N, Suite 201, Twin Falls, ID 83301
+1 208 410 1872
nathan.campbell@gtseek.com

Oregon State – Albacore GT-seq panel design and validation

Summary
Sequences containing a set of 495 SNP loci and one sex-linked marker were provided for GT-seq primer design. Primer sets meeting the physical properties for GT-seq and filtered for expected performance in multiplex PCR (N = 332) were ordered for testing. The initial test library returned an average on-target rate of 31% and 28 loci were identified that contributed to most of the off-target sequences. Primers for the 28 loci producing primer artifacts were omitted from a new primer pool and a validation library was prepared using the remaining 304 primer sets. The overall on-target rate for the validation library was 61% among the samples that performed well. Further analysis of the dataset identified a few other primers contributing to the off-target signal.

Primer Design
In order to retain as many of the few high F_{ST} markers selected for the GT-seq panel, these markers were preferentially retained when our multiplex PCR primer design algorithm found a conflict with another designed primer. Other marker types labeled as “spacial” and “neutral” represented a larger proportion of the selected markers and were considered more expendable when there were conflicts. Of the initial 495 target markers, 369 primer sets were ordered for testing.

Library Testing and Validation
The initial test library was created using all 369 target specific primers that were ordered. This first test library only included a read 1 data and the MiSeq run was not set up to collect read 2. As such, our initial analysis of the first test library did not included a primer-analysis step and we relied on internal probe sequences in the amplicons to determine an on-target rate for all the primer sets. Analysis of the initial test library identified 28 loci that were contributing to the majority of the off-target sequences in the dataset. The validation library was created using a primer pool without the primers from the offending 28 markers identified in test 1. Analysis of this library showed a much improved on-target rate of 61%. Further analysis was conducted to identify the sources of the remaining off-target sequences. The total number of markers that seem to be producing high quality genotype data in the validation library is ~237.

Sex ID marker
The initial results for this marker are that all samples produce a heterozygote signal using these primers. Assuming all samples are not of the same genotypic sex, the primer set doesn’t seem to discriminate by sex. However, multiple samples are producing a co-amplified product that does seem to present in only about half the samples. Further analysis will be required to determine if this can be made into a viable sex determination marker.