2011 + 2012 samples

Wildlife Genetics International

200-182 BAKER STREET (COURIER), P.O. BOX 274 (MAIL), NELSON, BC V1L 5P9
Phone: 250-352-3563 Facsimile: 250-352-3567 www.wildlifegenetics.ca

November 7, 2012

Deanna Ruth
SC Department of Natural Resources
420 Dirleton Rd.
Georgetown, SC 29440-9022

Re: WGI project g1173 South Carolina BB

Dear Ms. Ruth:

I have enclosed genetic results for 182 black bear hair samples that we separated into two subprojects: g1173a for the 125 samples that we received from you on March 6th, 2012 (Bear Data 2011); and g1173b for the 57 samples that arrived on September 21st (Bear Data 2012). The results were merged with those from previous projects, and are presented in the attached MS Excel file using the same formatting as in the past. Please don't hesitate to contact us with any questions that are not answered in the following review.

Sample Classes

The 125 database records from g1173a were classified as follows:

Xinadequate (20%): 25 samples that lacked suitable material for extraction.

Xspecies (2%): 3 samples that did not look like bear hair.

Xbomb (42%): 53 samples that failed during genetic analysis, including 18 of 20 analyzed samples from site 13.

sample (35%): 44 samples that were assigned individual ID.

The 44 g1173a samples that were genotyped successfully were assigned to 7 individuals (all male), of which 6 were recaptures from previous projects.

The 57 database records from g1173b were classified as follows:

Xinadequate (23%): 13 samples that lacked suitable material for extraction.

Xspecies (5%): 3 samples that did not look like bear hair.

Xbomb (21%): 12 samples that failed during genetic analysis.

sample (51%): 29 samples that were assigned individual ID.

The 29 g1173b samples that were genotyped successfully were assigned to 8 individuals (5M:3F), of which 6 (4M:2F) were recaptures from previous projects.

Combining the two subprojects, 12 (9M:3F) individuals were detected in project g1173, 3 (2M:1F) of which were new to the dataset, and 3 (all male) of which were 'caught' in both subprojects.

Sample Selection and Database Management

As per previous projects, we initially extracted 1 sample per site/period combination, biasing towards high quality samples. This led to extraction of only 19 samples from g1173a. On June 12th we consulted you, and agreed to revisit the 98 samples from g1173a that had initially been subselected. The final result was extraction of 138 samples (97 from g1173a and 41 from g1173b).

Record SC1-02-01-01-01 was present three times in your g1173b spreadsheet, so two repeats were removed. Sample SC1-02-01-01-02 was missing from the spreadsheet, so we created an entry for it. There were also 34 cases where the same sample ID was used in both subprojects, which we dealt with by adding the suffix "b" to the sample ID from g1173b.

DNA Extraction

DNA was extracted using QIAGEN's DNeasy Tissue kits, and following the manufacturer's instructions. We aimed to use 10 guard hair roots (see '#G' column) where available. When underfurs were used, the number recorded (see '#U' column) was an estimate because entire clumps of whole underfur were extracted rather than clipping individual roots. The amount of material leftover for potential re-extraction was classified as A (excellent) through C (none). Samples that did not contain at least 1 guard hair with a root, or 5 underfur, were not analyzed (*Xinadequate*) because their success rate is expected to be low. We also avoided some banded hair samples that seemed unlikely to be black bear samples (*Xspecies*).

Routine Microsatellite Genotyping

As per our quotes of January 9, 2012 (g1173a), and August 10, 2012 (g1173b), and following the template of previous projects, the analysis of individual identity used 6 microsatellite markers, plus a ZFX/ZFY gender marker.

This analysis started with a first pass of 138 extracted samples using 7 markers, after which we set aside 62 samples with high-confidence[1] scores for ≤ 3 of 7 markers, limiting the remainder of the analysis to samples with potential to yield complete, accurate genotypes.

The first pass was followed by a cleanup phase in which we reanalyzed data points that were weak or difficult to read the first time (i.e. that were scored with low-confidence, 2-digit allele scores). In some cases multiple rounds of reanalysis were required before data points could be upgraded to high-confidence (3-digit) scores. An additional 3 samples were set aside after cleanup.

The last phase of analysis was error-checking, following our published protocol of reanalyzing the mismatching data points in highly similar pairs of genotypes. Given that this project only added 3 multilocus genotypes to the dataset, it was unsurprising that we found no errors in the genotypes from this project.

Identification of Individuals

Once the genotypes were completed and checked for errors, we defined individuals for each unique multilocus genotype, taking ID numbers from the first sample to be assigned to each individual. This information is cross-referenced in the "Individual" column of the "Samples" spreadsheet, and the "List of Samples" column of the "Individuals" spreadsheet. Individuals that were first identified in g0600, g0741, and g1012 retain their name existing name, whereas the newly identified individuals have names derived from a g1173 sample.

Starting with 753 successfully genotyped samples from 4 projects, we identified 85 individuals (53 M:32 F). As per the summary above, the 12 individuals detected through the 73 successful samples from g1173 included 9 recaptures from previous projects, and biased 9 to 3 in favor of males. While the gender bias for g1173 is stronger than in the dataset as a whole, we noted that samples

[1] We use a combination of objective (peak height) and subjective (appearance) criteria to classify genotype scores. Low-confidence scores are identified by removing the leading digit from the allele score, and should be treated as equivalent to missing data.

collected in the first half of June in projects g0741 and g1012 were biased 12 to 5 in favor of males, perhaps suggesting some seasonality in gender bias.

One of the advantages of having sampled this study population for 5 years is that most of the genotypes in the dataset have been replicated in another sample. Given the number of markers in the analysis, and the number of potential errors that could be made at any given marker, the odds of our having recorded an inaccurate genotype in a g0600 sample, for example, and then recording the same inaccurate genotype in g1173, are extremely low. This type of data replication can only occur when genotyping errors are rare. Looking across the 753 samples to which we are currently assigning individual identity, 725 (96%) have had their multilocus genotype replicated in at least 1 other sample. Similarly, 64 of 73 successful samples from g1173 produced a genotype that matched to one from a previous year in this study area.

Success Rates

Reviewing the extraction notes, there is a correlation between guard hair roots per extracted sample and success rate. An average of 6.9 guard hair roots (90% success rate) was seen in g0600, 8.2 (82%) in g0741, 3.6 (74%) in g1012, 3.1 (45%) in g1173a, and 4.0 (71%) in g1173b.

Presumably a major driver of the reduced amount of hair per DNA extract in this project is the relaxed subselection rule, since we were able to target the best available samples in years where we limited ourselves to 1 sample per collection event.

As in past years, field season may have been a factor in success rate. In g1012 we saw a terrible 54% success rate for period 1 (June 2nd, 2010), compared to 88% and 83% for periods 7 (August 13th, 2010) and 8 (August 19th 2010). In g1173, all of the samples were collected in period 1 (June 10th, 2011 and June 14th, 2012), and the success rate was poor (45% for 2011, 71% for 2012).

In addition to time of year, there may be some meaningful variation by site, particularly for g1173a (2011), where samples from site 13 made up 20 of 97 DNA extracts (21%), but 18 of 53 'bombs' (34%). Removal of site raises the success rate from 45% to 55% for 2011 and from 71% to 74% for 2012.

The age of the barbed-wire used for collection is another parameter that has been associated with low success rates in other projects, since fresh wire seems to pluck hairs more aggressively.

Despite the poor headline numbers, the decision to extract multiple samples per site/period was productive. Not only did we identify bears at site/periods from which the first (pre-June 12) extract had 'bombed', but there were detections of multiple individuals at some site/periods. For example, we extracted 4 samples from the 2012 collection event SC1-02-01-27, identifying 3 individuals (the 4th sample appeared to contain a mix of DNA from 2 of the other individuals).

Marker Power

The addition of 3 individuals to an existing dataset of 82 is not expected to have a meaningful impact on considerations of marker variability or match probability. Consistent with this expectation, the 3 new multilocus genotypes were not in any 1MM- or 2MM-pairs. I continue to have confidence in the ability of the selected markers to uniquely identify each individual that you sample from this area.

Changes Relative g1012 Results File

Four samples from previous projects were used as positive controls since you last received results from us. In 3 cases, samples from g1012 were analyzed at 5 extra markers (markers other than those used for individual identity), so the change is just the addition of data for those markers. In the 4th case, sample SC1-02-09-21-02 from g0600 was reanalyzed at several extra markers, during which process we discovered an error in our original result at marker *REN145P07* ('*P07*'). This error had no impact on individual identification, but it did change an old result, rather than adding data as in the other 3 cases. These changes and additions are highlighted in the results file, and you should update your files accordingly.

Various and Sundries

It is my intention to communicate these documents in electronic form only, but I'd be happy to send hardcopies through the post if you need them. An invoice for US $4,365 for the g1173a samples was e-mailed to you on June 15, 2012 (paid; thank you). A second invoice to cover g1173b is attached, totalling US $1,845. Unless you tell me otherwise, I'll count on you to shepherd this invoice to the appropriate desk for processing.

We are willing to archive leftover DNA and hair for up to 5 years if there is a prospect for additional work in the region. However, if you do not foresee building on this dataset, we ask that you make other arrangements for the disposal or long-term storage of leftover materials.

Thank you for your continued patronage, and, once again, please feel free to call with questions or concerns.

Yours sincerely,

David Paetkau, Ph.D.
President

encl.: g1173 Results.xls; g1173b Invoice.pdf