2010 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

June 28, 2011

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

Re: WGI project g1012 South Carolina BB SC3

Dear Ms. Ruth:

I have enclosed genetic results for 409 black bear hair samples and 2 black bear tissue samples that we received from you on February 22nd, 2011. The results have been merged with existing results from projects g0600 and g0741, and are presented in the attached MS Excel workbook using the same formatting as in the past. The following notes should provide the information needed to understand and defend this project, but please feel free to contact us for further detail.

Sample Classes

We received 2 tissue samples from an illegally killed bear, and we extracted and successfully genotyped the skin sample. The 414 database records from SC03 were classified as follows:

Xnotsent (n/a): 5 spreadsheet records with no matching samples.

Xinadequate (5%): 20 samples that lacked suitable material for extraction.

Xspecies (1%): 3 samples that did not look like bear hair.

Xsubselect (62%): 253 samples that were eliminated by subselection rules.

Xbomb (8%): 34 samples that failed during genetic analysis.

sample (24%): 99 samples that were assigned individual ID.

The 99 good hair samples were assigned to 17 individuals (11M:6F), of which 13 (9M:4F) were recaptures from previous projects.

Sample Selection and Database Management

There were 5 records listed on your spreadsheet for which we did not find physical samples. These records are classified '*Xnotsent*' in the results file. On the other side of the ledger, we manually created database records for 4 physical hair samples from SC03 and 2 tissue samples that were not listed on your spreadsheet. The tissue samples (identified by you as being from the same illegally killed bear) were given sample IDs of "illegalkilltooth" and "illegalkillskin". In one case a hand-entered sample had a non-unique sample ID (SC03-02-06-41-04), so we changed the envelope ID to SC03-02-06-41-05.

We did not find sample envelopes to go with records SC03-02-02-32-07 to SC03-02-02-32-09, but did find sample envelopes labeled SC02-02-02-32-01 to SC02-02-02-32-03. We re-named the electronic records SC02-02-02-32-01 to SC02-02-02-32-03 to match the sample envelopes.

As per projects g0600 and g0741, we tried to extract 1 sample from each of 153 site/period combinations, biasing towards high quality samples. There were 20 cases where no suitable samples were found for a given site/period combination.

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. An estimate of the amount of the leftover hair (see 'Left' column) was made using three classes: no guard hairs (C); 1 to 4 guard hairs (B); and more than 4 guard hairs (A). 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.

Success Rates

Looking back on the extraction notes, we see a marked decrease relative to previous years in the amount of hair per extracted sample. For example, there were an average of 3.6 guard hair roots per extracted sample this year, compared to 6.9 in g0600 and 8.2 in g0741. Unfortunately, this decrease was reflected in success rates, with just 74% of extracted samples being successfully assigned individual ID, compared to 90% in g0600 and 82% in g0741.

When success rates were compared across projects with a correction for sample quality, there was no real trend. For example, samples extracted from > 2 guard hair roots ranged from an 88% success rate this year to a 92% success rate in g0600.

As we have seen in past years, field season appeared to be a factor in success rate, with a success rate of only 54% for period 1, compared to 88% and 83% for periods 7 and 8, respectively.

Another factor that may have contributed to the decreased success rate was that many of the guard hairs were missing the root bulbs (i.e. only had "light ends"; see extraction comments). We have been told that the barbs on barbed-wire tend to loosen with use, acting more to comb out loose hair than to pluck hairs that are firmly attached, as when the wire is fresh; I wondered if this might be an issue in your project.

Another comment about sample quality that may or may not relate to success rates is that the samples were stored with chalky desiccant in plastic bags, which we have associated with lower success rates in other projects. We recommend that sample envelopes be stored in a breathable container, such as a cardboard box. If you are concerned about moisture, then samples can be placed in a sealed bag, along with silica desiccant that is contained in a separate breathable container (such as a sock), but we ask that desiccant and sample envelopes not be allowed to come into direct contact with each other. While this advice is not based on experimental results, we have seen the best success rates in projects that did not use desiccant or plastic bags.

Routine Microsatellite Genotyping

As per your quote of March 15, 2011, the analysis of individual identity used 6 microsatellite markers from g0600 and g0741, plus a ZFX/ZFY gender marker.

This analysis followed a 3-phase approach, starting with a first pass of all 134 extracted samples using all 7 markers. After first pass, we set aside 34 samples with high-confidence[1] scores for ≤ 3 of 7 markers, eliminating the most time consuming and error-prone samples from the remainder of the analysis.

[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.

The first pass was followed by a cleanup phase in which we re-analyzed 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 re-analysis were required before data points could be upgraded to high-confidence scores, but in the end all 100 samples that were not culled after first pass (99 hair and 1 tissue) had high-confidence scores for all 7 markers that we were analyzing.

Error-Checking

The last phase of analysis was error-checking, following our published protocol of selective data re-analysis (Paetkau 2003). Sample SC1-02-06-17-01 from project g0600 was a 1MM at gender to your sample SC03-02-05-11-01, so we re-analyzed gender in both samples using both the ZFX/ZFY and the amelogenin marker. This process confirmed an error in SC1-02-06-17-01, which we corrected (highlighted in the results file).

Following the correction to the sample from g0600, there was 1 1MM-pair and 1 2MM-pair that had been created by the addition of your g1012 samples. Data for the mismatching markers in these pairs were confirmed through re-analysis, but as an extra precaution against genotyping error we extended the genotypes involved in these 2 pairs to 10 markers by analyzing 1 sample per individual at markers *MU50*, *G1A* and *G1D*. With the addition of the extra data, both pairs mismatched at ≥ 3 markers, confirming that the genotypes in question came from different bears. The differences underlying these similar pairs were also inconsistent with the most common types of errors, such as allelic dropout or a scoring shift to an adjacent allele. Under these circumstances, there is no reasonable probability that the number of individuals identified in the combined dataset was inflated through undetected genotyping error (Kendall *et al.* 2009 *JWM*).

Identification of Individuals

Once the genotypes were completed and checked for errors, we defined individuals for each unique 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" worksheet, and the "List of Samples" column of the "Individuals" worksheet. Individuals that were first identified in g0600 and g0741 retain their name from that project, whereas newly identified individuals will have names derived from g1012 samples.

Starting with 465 Lewis Ocean Bay (study area 02) samples from 3 projects, we defined 40 individuals (22M:18F). The 99 good hair samples from the current project were assigned to 17 individuals (11M:6F), of which 13 were recaptures of individuals that had been identified in SC1 and/or SC2 (refer to the capture matrix in the Individuals worksheet for details). The tissue sample from the illegal kill assigned to a new male individual.

The overall sex bias in this study region is small (22M:18F), but the average male had substantially more samples assigned to it (17 samples per individual) than the average female (5.1 samples per individual). It is unclear how these numbers translate into capture frequencies, since males tended to be caught in either 1 or 3 of 3 years, whereas most females were caught in 2 of 3 years.

One of the advantages of the sampling intensity that comes from 3 years of work 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 an SC1 sample, for example, and then recording the same inaccurate genotype in SC3, are extremely low: generally speaking, this type of data replication only occurs when the data are accurate. Looking across the 465 Lewis Ocean Bay samples to which we are currently assigning individual identity, 457 (98%) have had their multilocus genotype replicated in at least 1 other sample. That level of data replication speaks to a genotyping protocol with a very low rate of error.

Marker Power

If one conducts enough observations, one will occasionally encounter rare results (the concept behind Type I error), and in this study the observation of 2 1MM-pairs as compared to 1 2MM-pair stands out against hundreds of other mismatch curves that we have created for similar projects: we never see more 1MM-pairs than 2MM-pairs.

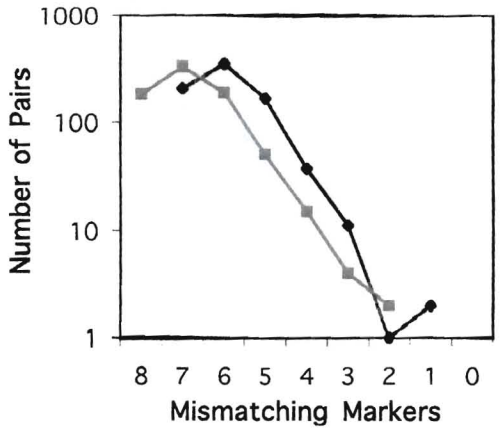
In cases where we know family relationships, we see that the right-hand tail of mismatch distributions is composed disproportionately of first-order relatives (as an aside, this illustrates why calculated match probabilities provide so little practical information: they vary dramatically with degree of relatedness). While we normally describe 10-fold decreases with increasing degree of similarity (e.g. 10 2MM-pairs for every 1MM-pair; Paetkau 2003), it is often the case that small populations in which the proportion of first-order relatives is comparatively high

deviate from this trend. The variability in your study population does not suggest dramatic isolation and 'inbreeding', but I'm curious whether you think that small population size and isolation might provide part of the explanation for the unexpected observation of 2 1MM-pairs.

Other than consanguinity, the other explanations for an overabundance of 1MM-pairs would be genotyping error and chance. As detailed above, we went to great lengths to rule out genotyping error in these cases, even confirming the results through the analysis of extra markers. Using the same 7 markers, the 26 individuals from study area 01 in SC1 and SC2 (not included in this results file) do not include any 1MM- or 2MM-pairs. These observations reinforce my impression that small datasets often deviate from expectation in one direction or the other, and that chance played a role in creating the unusual mismatch curve that we observed for these 40 black bears (Fig. 1).

Having thought about this issue longer than I probably should have, it is still my opinion that the 7-locus marker system used in this project (i.e. without *MU50*) is sufficient to ensure a low probability of sampling any pair of individuals with the same multilocus genotype. At the same time, the 1MM-pairs provide a note of caution, and a reminder to let us know if you see any matches that seem unlikely (e.g. matches that are widely separated in space) so that we can confirm them by analyzing additional markers.

Fig. 1. Distribution of genotype similarity for the 40 7-locus genotypes in the attached results file (diamonds), and as it would have looked with the continued use of *MU50* as an 8th marker (squares). The right-hand tail of the 8-locus curve certainly looks more typical, although both deviate from the straight log-linear relationship that we are used to.

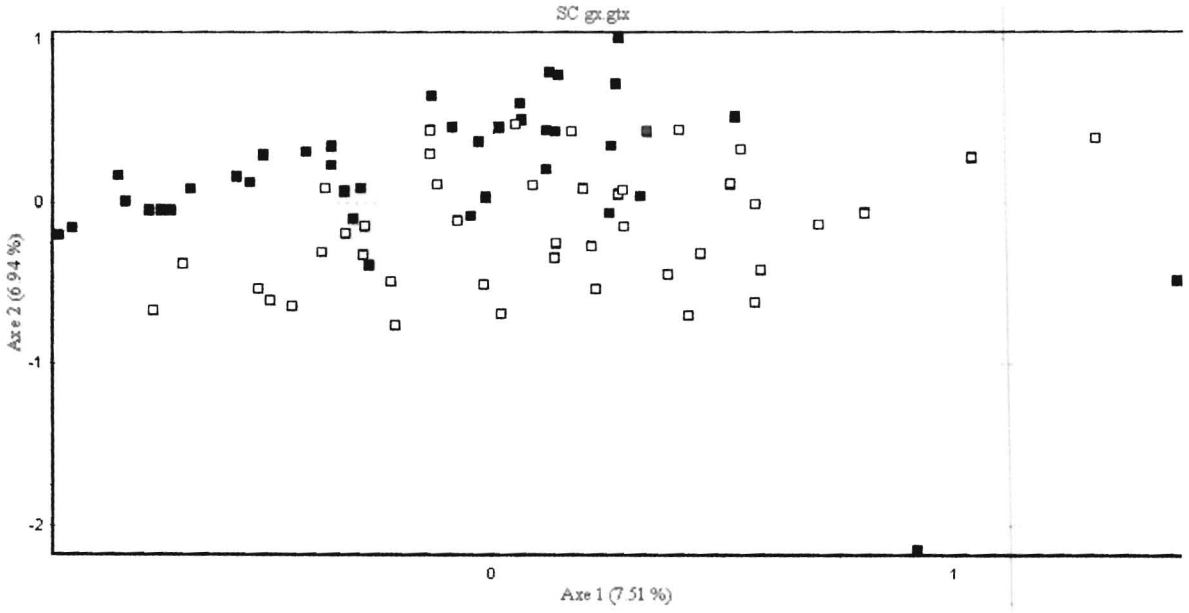


Clustering Analysis

In project g0741, we did a 22-locus comparison of 37 individuals from your region, and did not find any capacity to separate animals based on study area. This year we performed another PCA-style clustering analysis in the program Genetix, this time using data from 7 markers and 82 individuals (Fig. 2). One expects the clustering power to be low in an analysis based on just 7 markers, but we hoped that the better sample size might partially offset this weakness.

Interestingly, the 7-locus clustering analysis did show more separation between study areas 01 and 02 than was apparent in the 22-locus analysis. The new analysis also identified 2 outliers (SC1-02-03-18-01 and SC1-02-07-11-04), one of which was a dramatic outlier in the 22-locus version. The genotypes in question had been replicated in 4 and 17 samples, respectively, effectively ruling out this explanation. These outlying genotypes also had unusual alleles at multiple markers (highlighted in red in the 'Individuals' section of the results file), rather than a strange result at a single marker as expected when a data entry or amplification error causes an incorrect genotype to be recorded. Given sufficient interest, we could pursue this aspect of the study by expanding the number of individuals with genotypes for 20 or more markers, and by considering source populations for putative immigrants. For the second time in two pages, I also find myself being curious about the structure of the population system under study.

Fig. 2. 7-locus clustering results for 82 individuals color coded by origins: study area 01 (yellow); study area 02 (blue); miscellaneous (white). Note the two outliers at bottom right.



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 $6,030 was emailed to you on June 8, 2011, and a copy has been enclosed for your reference. Please tell me if you would like a copy of the invoice forwarded to someone else, otherwise I'll count on you to shepherd it to the appropriate desk for processing.

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

Yours sincerely,

David Paetkau, Ph.D.
President

encl.: g1012 Results.xls; g1012 Invoice.pdf