**Lesson 3 Exercises**

1. Pick a paper on ancient DNA published after Haak et al. (2015). Read through their methods and write a short paragraph explaining the DNA extraction, screening, quality control, contamination, and processing done. Where appropriate, explain why you think they used a certain method.
2. Read the methods section of the paper for your individual – write briefly what type of data is available for your individual, as well as contamination, coverage, etc. **Next class be prepared to share this information.**
3. The command below aligns reads in the fastq file (student.fastq) to the reference human mtDNA sequence and creates an indexed BAM file.

bwa aln -t N /mnt/solexa/Genomes/human\_MT/bwa-0.5.10 /public/adna/student/data/student.fastq | bwa samse /mnt/solexa/Genomes/panda/bwa-0.5.10 - /public/adna/student/data/student.fastq | samtools view -uS - | samtools sort -o test.bam - ; samtools index test.bam

1. Test out the code in your exercises/ folder. Can you research what each part of the command above (between each “|”) is doing? Describe each part briefly.
2. Next, run the command “bam-rmdup -o test.rmdup.bam test.bam”. How many duplicates were in the file? How many unique reads were present?
3. Below is a screenshot of a table with information prepared about the libraries prepared for a set of samples. Red cells indicate mtDNA contamination estimates failing the criteria of <95% for the point estimate, or a 95% CI that includes values <85%.
4. Explain what each of the column headings mean.
5. Each library passed, failed, or only damage-restricted was used. Explain how library decisions were determined to the best that you can.

Table for Exercise 3.3.

