Demultiplexing ddRAD

Demultiplexing ddRAD data with Stacks exercise:
This exercise should be done on TACC.
First copy over the exercise directory:

<table>
<thead>
<tr>
<th>copy over directory</th>
</tr>
</thead>
<tbody>
<tr>
<td>start an idev session if you have not already idev</td>
</tr>
<tr>
<td>go to coure directory on scratch</td>
</tr>
<tr>
<td>cd rad_intro/</td>
</tr>
<tr>
<td>copy over the demultiplexing directory</td>
</tr>
<tr>
<td>cp -r /work/02260/grovesd/lonestar/intro_to_rad_2017/demultiplexing/process_ddRAD_stacks .</td>
</tr>
<tr>
<td>go to the demultiplexing with stacks exercise directory</td>
</tr>
<tr>
<td>cd process_ddRAD_stacks</td>
</tr>
</tbody>
</table>

Check out the reads to make sure they make sense:
# we have two fastq files: Lib01_R1.fastq and Lib01_R2.fastq
ls *.fastq

# These are paired end reads, two separate reads from either end of the same DNA fragment
# Double-check that the files have the same number of reads. (note that fastq generally files have 4 lines per read)
expr $(cat Lib01_R1.fastq | wc -l) / 4
expr $(cat Lib01_R2.fastq | wc -l) / 4
#7453585
# (Note these are for demo purposes and shorter than they should be)
# this ddRAD library was prepared so that forward reads were cut with the nlaIII restriction enzyme.
# Looking up nlaIII, we see it's cut site:
#
#     CATG'
#     'GTAC

# check if we see this cut site in the forward reads
grep CATG Lib01_R1.fastq | wc -l

# compare this with the total number of reads we got above

# do these numbers make sense?

# look at where in the reads the cut site is (may help to paste into a text editor and character search)
grep CATG Lib01_R1.fastq | head -n 30

process the reads using process_radtags (part of the Stacks package):

# make a directory to put the resulting sample fastq files into
mkdir sample_fastqs

# look at the documentation for process_radtags
./process_radtags -h

# execute the command to process the rad data
./process_radtags -i 'fastq' -1 Lib01_R1.fastq -2 Lib01_R2.fastq -o ./sample_fastqs/ -b barcodes_Lib1.tsv --inline_index -e 'nlaIII' -r --disable_rad_check

Check the results:
check results

# how many barcode combinations did we have?
# First look at the barcodes file:
cat barcodes_Lib1.tsv

# then count the lines
cat barcodes_Lib1.tsv | wc -l

# how r1 and r2 fastq files did we output (ignore the 'rem' files)
ls sample_fastqs/*AGCGAC.1.fq
ls sample_fastqs/*AGCGAC.1.fq
ls sample_fastqs/*AGCGAC.1.fq | wc -l
ls sample_fastqs/*AGCGAC.1.fq | wc -l

# are the paired end files still the same length?
cat sample_fastqs/sample_CATAT-AGCGAC.1.fq | wc -l
cat sample_fastqs/sample_CATAT-AGCGAC.2.fq | wc -l