RADseq Works in Primates, Dammit.

Christina M. Bergey Andrew S. Burrell
Luca S.J. Pozzi Todd, I suppose

December 8, 2012

Abstract

... Blah, blah, blah, RADseq, blah, blah, Cercopithecoidea. ...

1 Introduction

- Next-gen sequencing revolution promises gains in primatology
- Still expensive
- Many genomes, but still tough doing genomics on non-model organisms
- What is RADseq?
- Previous RADseq studies
- Why would it be good for primates
- PRESENT STUDY
  - We did RADseq on 6 Cercopithecoids
  - Assessed how well it worked
  - Show it has promise for primates

2 Methods

Library Preparation and Sequencing Genomic DNA from 6 animals was digested with *PspXI* (New England Biolabs) and used to create a multiplexed RAD tag library. Our library preparation method followed that of Etter et al, 2011 with the following modifications: the P1 adapter top(?) oligonucleotide was modified to have an overhang corresponding to the cut site of *PspXI*, and a longer P2 adapter suitable for paired end sequencing was used (P2_top: 5'-SEQUENCEHERE-3'; P2_bottom: 5'-SEQUENCEHERE-3'). Individual-specific barcodes contained in the P1 adapter differed by at least three nucleotides. We chose *PspXI* based on the results of *in silico* digestion of the human, rhesus macaque, and baboon reference genomes using custom Perl scripts (refs). We sequenced the prepared library as one 150-cycle paired-end run of an Illumina MiSeq at the NYU Langone Medical Center's Genome Technology Center using a spike-in of 30% PhiX DNA to control for low diversity in the library at the barcode and restriction sites. Sequences are available to download from the NCBI Short Read Archive (accession number SRAXXXXXX.X).

Analysis Pipeline - Mapping to Reference Genomes

- Demultiplex. Must have barcode and restriction site intact.
- Analyze reads with FastQC (No need to say this)
  - Total sequence bp
  - Maximum possible sequence depth
  - Other stats that FastQC gives you
- Aligned to rhesus genome using BWA aln
  - default parameters
- Combine paired-end reads with BWA sampe (No need to say this)
- Convert to BAM, sort and index with samtools (No need to say this)
- Analyze mapped reads with samtools utilities flagstat and idxstats and bamtools utility

- Post-alignment filtering steps
  - Fix mate pair info with Picard
  - Filter for mapped and paired.
  - Remove dups with Picard
  - Add read group info with Picard (No need to say this)
  - Remove reads with low mapping quality with bamtools

Analysis Pipeline - Variant Calling

- Local realignment with GATK
- Fix paired end data with Picard (No need to say this)
- Call SNPs with samtools. Min coverage = 3, max = 100
- Summarize SNP stats with vcf-stats (No need to say this)

Analysis Pipeline - Analysis of Degree of Overlap

- Calculate coverage of restriction site-associated regions
  - Info on targeted intervals
    - Total number possible targets in rhesus genome (compare to human too?)
    - Total possible target BP
  - How many targets did we hit?
    - BEDtools multiBamCoverage for this job
    - Number and percentage of targets with coverage ≥ 1
    - Number and percentage of targets with coverage $\geq N$
- Count orthologous SNPs shared between individuals
  - VCFtools vcf-compare for this job

3 Results

- Table:
  - Number of reads per animal
  - Number that passed filtration
  - Number of loci hit
  - Number of loci hit with coverage $\geq N$
  - Number of SNPs
- SNP Venn diagram?
- Table of overlapping region, orthologous SNP counts

4 Discussion

- RADseq is viable tool for researcher interested in primate phylogenetics, pop. gen.
- Enzyme choice allows control over coverage, number of individuals, number of loci.
- Potential problems with RADseq method
- Promise for primatology

5 Acknowledgements

Acknowledgements