RADseq Works in Primates, Dammit.

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

December 7, 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?
- PRESENT STUDY
  - We did RADseq on 6 Cercopithecoids
  - Assessed how well it worked
  - Did simple phylogeny?

2 Methods

Library Preparation and Sequencing

- 6 animals, table with sources and other info
- Etter et al. 2011, though with modifications

- How we picked enzyme, PspXI
- Adapter sequences. Barcodes with at least 3 mismatches. PE adapter
- What did it look like on BioA? Size? Concentration?
- Sequenced on Illumina MiSeq. 150PE
- NYU Langone Medical Center's Genome Technology Center (name right?)
- How much actually loaded?
- 30% spike in with PhiX control DNA to control for low diversity library

Analysis Pipeline - Mapping to Reference Genomes

- Demultiplex. Must have barcode and restriction site intact.
- Analyze reads with FastQC
  - 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
- Convert to BAM, sort and index with samtools
- 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
  - Remove reads with low mapping quality with bamtools

Analysis Pipeline - Variant Calling

- Local realignment with GATK
- Fix paired end data with Picard
- Call SNPs with samtools
- Summarize SNP stats with vcf-stats

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

Analysis Pipeline - Inferring Phylogeny

- Using Stacks?
- Using method like cichlid people?
- Using method like Rubin et al

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
- Phylogenetic tree

4 Conclusions

Conclusions

5 Acknowledgements

Acknowledgements