0003 - Steelhead RADseq Project

Goal: Comparison of natural versus hatchery raised steelhead
Investigators: Mackenzie Gavery

2016/05/09

Fastqc Results:
Plate1
Plate2

De-multiplexing the plates into the individual samples using process_radtags from stacks (v1.23).

process_radtags script:

```
#!/bin/bash

STACKS=/share/bioinformatics/stacks-1.23/bin
IN=raw
OUT=process_radtags

${STACKS}/process_radtags \
    -f ./raw/Plate1.fq \
    -i fastq \
    -y fastq \
    -o ./${OUT} \
    -e sbfI \
    -r \
    -c \
    -q \
    -b 2015WinthropAdult_RADbarcode_Plate1.txt \
    -D \
    > pr.plate1.log 2>&1

mv ./${OUT}/process_radtags.log ./${OUT}/process_radtags.Plate1.log

${STACKS}/process_radtags \
    -f ./raw/Plate2.fq \
    -i fastq \
    -y fastq \
    -o ./${OUT} \
    -e sbfI \
    -r \
```

```
    -c \
    -q \
    -b 2015WinthropAdult_RADbarcode_Plate2.txt \
    -D \
    > pr.plate2.log 2>&1

mv ./${OUT}/process_radtags.log ./${OUT}/process_radtags.Plate2.log
```

process_radtags Results
Plate1
Plate2

Running bowtie (v1.1.2) to map samples to omykiss scaffolds

Bowtie Mapping Script:

```
#!/bin/bash

BASE=/scratch/ggoetz/omykiss-rad-mac

IN=${BASE}/process_radtags
OUT=${BASE}/bowtie

REF=${BASE}/ref/OncMyk.scaffolds

LOC=/share/bioinformatics/bowtie

cd ${IN}
for file in `ls -1 *.fq`
do
    NAME=`echo ${file} | awk -F "." '{print $1}'`
    echo ${NAME}
    ${LOC}/bowtie \
        --chunkmbs 200 \
        -n 3 \
        -k 10 \
        -p 6 \
        --best \
        ${REF} \
        ${file} \
        ${OUT}/${NAME}.bowtie \
        > ${OUT}/${NAME}.bowtie.log 2>&1
done
cd ${BASE}
```

Percent Mapping Results:
Bowtie - Percent Mapping Results

Running pstacks using the bowtie mapping data to generate the snps, tags, and alleles files for each sample.

pstacks script:

```
#!/bin/bash

STACKS=/share/bioinformatics/stacks-1.23/bin

BASE=/scratch/ggoetz/omykiss-rad-mac

IN=${BASE}/bowtie
OUT=${BASE}/stacks_results
LOGS=${BASE}/logs

cd bowtie

for file in `ls -1 *.bowtie`
do
    NAME=`echo ${file} | awk -F "." '{print $1}'`
    echo ${NAME}
    ID=`echo ${NAME} | awk -F "_" '{print $1}'`

    ${STACKS}/pstacks \
        -t bowtie \
        -f ${IN}/${file} \
        -p 6 \
        -o ${OUT} \
        --model_type 'bounded' \
        --bound_low 0.001 \
        --bound_high 0.01 \
        -m 3 \
        -i ${ID} \
        > ${LOGS}/${NAME}.pstacks.log 2>&1
done
```

Samples Selected For Catalog. Randomly selected ten samples from each group (natural and hatchery) that had between 2.5million and 3.5million reads.

```
35_H    151_N
82_H    86_N
80_H    143_N
93_H    84_N
118_H    83_N
108_H    139_N
```

```
120_H    78_N
34_H    133_N
85_H    4_N
45_H    134_N
```

Running cstacks to build catalog from the selected samples.

cstacks script:

```
#!/bin/bash

STACKS=/share/bioinformatics/stacks-1.23/bin

BASE=/scratch/ggoetz/omykiss-rad-mac

IN=${BASE}/stacks_results
OUT=${IN}
LOGS=${BASE}/logs

LIST=`cat list.*.txt`

LIST_CMD=""
for item in ${LIST}
do
    LIST_CMD="${LIST_CMD} -s ${IN}/${item}"
done

${STACKS}/cstacks \
    -b 1 \
    -p 4 \
    -o ${OUT} \
    ${LIST_CMD} \
    > ${LOGS}/cstacks.log 2>&1
```

Running sstacks to compare samples to catalog

sstacks script:

```
#!/bin/bash

STACKS=/share/bioinformatics/stacks-1.23/bin

BASE=/scratch/ggoetz/omykiss-rad-mac

SAMPLES=${BASE}/process_radtags

IN=${BASE}/stacks_results
OUT=${IN}
LOGS=${BASE}/logs
```

```
cd ${IN}

for file in `ls -1 ${SAMPLES}/*.fq`
do
    NAME=`echo ${file} | awk -F "/" '{print $NF}' | sed -e "s/.fq//"`
    echo ${NAME}
    ${STACKS}/sstacks \
        -b 1 \
        -c ${IN}/batch_1 \
        -s ${IN}/${NAME} \
        -p 4 \
        > ${LOGS}/${NAME}.sstacks.log 2>&1
done
```

Script to make popmap file for populations, uses awk.

popmap script:

```
{
    NAME="";
    if ($2 == "N")
        NAME="natural";
    else
        NAME="hatchery";

    print $1"_"$2"\t"NAME;
}
```

Populations, initial script

populations script:

```
#!/bin/bash

STACKS=/share/bioinformatics/stacks-1.23/bin

BASE=/scratch/ggoetz/omykiss-rad-mac

IN=${BASE}/stacks_results
LOGS=${BASE}/logs

cd ${IN}

${STACKS}/populations \
    -b 1 \
    -P . \
    -M popmap.txt \
    -m 10 \
```

```
    --vcf \
    --genepop \
    --write_single_snp \
    --structure \
    > ${LOGS}/populations_1.log 2>&1
```

2016/06/02

Comparing the 2015 data to the 2014 catalog. First need to fix the ids in the 2015 data. Tried this a couple times, had to make some changes to the 'fix script'.

Fix script:

```
#!/bin/bash

for file in $(ls *_H.tags.tsv *_N.tags.tsv)
do
    echo ${file}
    awk -F "\t" 'BEGIN {OFS="\t"} {$2=$2+200; print}' ${file} \
        > ../combined_2014_2015/${file}
done

for file in $(ls *_H.snps.tsv *_N.snps.tsv)
do
    echo ${file}
    awk -F "\t" 'BEGIN {OFS="\t"} {$2=$2+200; print}' ${file} \
        > ../combined_2014_2015/${file}
done

for file in $(ls *_H.alleles.tsv *_N.alleles.tsv)
do
    echo ${file}
    awk -F "\t" 'BEGIN {OFS="\t"} {$2=$2+200; print}' ${file} \
        > ../combined_2014_2015/${file}
done
```

Running sstacks again using the 'fixed' 2015 files. Copied the 2014 catalog in to the combined folder.

sstacks script:

```
#!/bin/bash

# Using older version of stacks to match 2014 data
STACKS=/share/bioinformatics/stacks-1.23/bin

# Using the 2014 catalog
```

```
for file in $(ls -1 *_[HN].snps.tsv)
do
    NAME=$(echo ${file} | sed -e "s/.snps.tsv//")
    echo ${NAME}
    ${STACKS}/sstacks \
        -b 1 \
        -c batch_1 \
        -s ${NAME} \
        -p 6 \
        > ${NAME}.sstacks.log 2>&1
done
```

Populations script with whitelist:

```
#!/bin/bash

STACKS=/share/bioinformatics/stacks-1.23/bin
#STACKS=/share/bioinformatics/stacks-1.39/bin

${STACKS}/populations \
    -b 1 \
    -P . \
    -M popmap.2014_2015.txt \
    -m 10 \
    --vcf \
    --genepop \
    --write_single_snp \
    --structure \
    -W ../whitelist_936SNPs.txt \
    > populations.run2.log 2>&1
```

Looks like populations finished without any problems.