SSCS vs Trimmed Read Variant calls GVA2020

Overview:

Throughout the course we've gone over how errors can pop up in your data and how they can effect confidence in variant calls and knowledge of what variants are actually real. Here we provide you with real data to show the difference between a non-corrected library and a error corrected library.

This tutorial makes use of data generated in the quick Error Correction tutorial. If you have not done that tutorial already you should do it first.

Learning Objectives:

1. Run breseq on fastq files corresponding to error corrected, and non-error corrected data
2. Gain an understanding of just how powerful error correction can be

Running Breseq on SSCS and Trimmed Reads:

The reference files **1400flanking.gff3**, and **REL606.maksed.gff3** should be copied from the $BI/gva_course/mixed_population directory into a new folder GVA_breseq_Error_Correction while **DED110_SSCS.fastq**, and **DED110_all.trimmed.fastq** should be copied from your GVA_Error_Correction folder.

```
mkdir $SCRATCH/GVA_breseq_Error_Correction
cp $BI/gva_course/mixed_population/*.gff3 $SCRATCH/GVA_breseq_Error_Correction
cd $SCRATCH/GVA_breseq_Error_Correction
```

```
cp $SCRATCH/GVA_Error_Correction/DED110_SSCS.fastq .
cp $SCRATCH/GVA_Error_Correction/DED110_all.trimmed.fastq .
```

Just like our previous tutorials these commands should be run on an idev shell and they should be run in the background so both comparisons can be run at the same time.

In this case we are going to be running breseq in both polymorphism mode (**-p**), and targeted sequencing mode (**-t**). By default breseq creates a lot of html files as output which ordinarily are very useful to visualize what the characteristics of any mutation are (in a better way than can be done with IGV). For normal runs this is extremely useful. For targeted deep sequencing runs (especially ones where we have not done error corrections), this can be incredibly time consuming. Since we are ultimately only interested in comparing the difference in number of mutations detected and their frequencies rather tahn the specific characteristics of those mutations, we also include the **--brief-html-output** option. We will also be using 2 different reference files, **1400flanking.gff3** should be used as a standard reference while **REL606.masked.gff3** should be used as a junction only reference.

```
mkdir Logs
module unload samtools
breseq -p -t -o SSCS_output/trimmed -r 1400flanking.gff3 -s REL606.masked.gff3 --brief-html-output DED110_all.
trimmed.fastq >& Logs/trimmed.log.txt &
breseq -p -t -o SSCS_output/SSCS -r 1400flanking.gff3 -s REL606.masked.gff3 --brief-html-output DED110_SSCS.
fastq >& Logs/SSCS.log.txt &
# wait a few moments and check that breseq is running using the tail command
tail Logs/*.log.txt
```

Generate comparison table for the 2 types of samples

Just like the previous tutorial we now want to create a comparison file using "gdtools compare" so we can see the difference of having the trimmed reads vs the SSCS reads. This time you want to create a file "trimmed_vs_SSCS.html".

This was extensively covered in the Advanced Breseq tutorial. You may want to review this section of the tutorial for more detailed explanation while you wait for the 2 runs to complete. Check which commands are still running using the jobs command and the tail command listed above.

```
gdtools compare -h
```

the above will display instructions stating that gdtools compare needs a genebank (gbk) reference file, and .gd files to compare. The -o command is used to direct the output to a specific file name and/or location.

```
gdtools compare -o trimmed_vs_SSCS.html -r reference/REL606.masked.gff3 -r reference/1400flanking.gff3
SSCS_output/SSCS/output/output.gd SSCS_output/trimmed/output/output.gd
```

The comparison table as well as the output for the trimmed and SSCS reads can now be exported to your local computer for viewing.

Evaluating the effect of error correction:

Once you have transferred the **trimmed_vs_SSCS.html** file back to your computer, open it and scroll around to see what the effects of our SSCS error correction were.

Return to GVA2020