Oyster Genome Resequencing Data

Dear all,

I am happy to write that the initial bioinformatic analysis of our 90 resequencing samples is complete. I have created a GitHub repository https://github.com/jpuritz/OysterGenomeProject that contains QC information, sample metadata (that I have), and step-by-step documentation of the various steps used for data processing.

Overall, we had about 37.8 million read pairs per sample after trimming and adapter removal with a minimum of 21.4 million read pairs and maximum of 81.7 read pairs. Reads were mapped with `BWA`, duplicates marked with `Picard`, and variants were called using `Freebayes`. We ended up with 56,358,904 total raw variant calls. After some filtering (detailed in repo), we have between 10 and 30 million SNPs and 5,667,164 INDEL variants.

I have made nearly all the data available to you via SFTP from my server. The login information is below:

Criteria	Credential
Server address	KITT.uri.edu
Port	2292
User	user1
Password:	...OyGen.2.0.1.8

Everything will be within the directory `/home/user1/files/`, and I **DO NOT** recommend trying to download the whole thing. It's 2.9 TB. You will likely be interested in the contents of `./VCF_files/` which contain the relevant variant files:

File Name	Variants	Number of Loci	Size
INDELs.TRSdp5g75FnDNA.vcf.gz	Filtered INDel variants with missing data of less than 25% allowed	5,667,164	14 GB
SNP.TRSdp5g75FnDNAmaf01.vcf.gz	Filtered SNPs with a minor allele frequency > 0.01	23,579,809	36 GB
SNP.TRSdp5g75FnDNAmaf05.vcf.gz	Filtered SNPs with a minor allele frequency > 0.05	10,983,744	20 GB
SNP.TRSdp5g75FnDNA.vcf.gz	Filtered SNPs	29,175,863	41 GB

SNP.TRSdp5g95FnDNAmaf05.vcf.gz	Filtered SNPs with a minor allele frequency > 0.01 and missing data < 5%	7,030,071	14 GB
TotalRawSNPs.vcf.gz	The original raw variant VCF	56,358,904	74 GB
SNP.TRSdp5g75mtDNA.recode.vcf	Filtered SNPs from mitochondrial genome	1337	9.5 MB

Please let me know if you have any questions,

Jon