VCFHunter
A Step-by-Step Guide
[DRAFT]

Michael Hall

Plant Breeding and Genetics Laboratory
FAO/IAEA Joint Division
Seibersdorf, Austria

Created: July, 2022
Last updated: 14 July 2022

Contents

1 VCFHunter on Sorghum Semi-Dwarfism

1.1 Software Prerequisites

#Download git repository:

```
git clone https://github.com/SouthGreenPlatform/VcfHunter.git
cd VcfHunter
cd bin
ls

#See all available python scripts

VCFtools (0.1.15)
© Adam Auton and Anthony Marcketta 2009

Process Variant Call Format files

For a list of options, please go to:
https://vcftools.github.io/man_latest.html

Alternatively, a man page is available, type:
man vcftools

Questions, comments, and suggestions should be emailed to:
vcftools-help@lists.sourceforge.net
```

1.2 Extract Sample names from VCF

```
head -n 1000 DNAseq_prefiltered.vcf | grep "#CHROM" | sed 's/\t/\n/g' | tail -n +10 >␣
↪all_names.tab
```

1.3 Filter VCF with python script

```
#Filter VCF
python vcfFilter.1.0.py --vcf ../data/Sorghumvcf/freebayes_D2.filtered.vcf --names␣
↪sorghum_all_names.tab
--MinCov 10 --MaxCov 300 --MinAl 3 --nMiss 1 --RmAlAlt 1:3:4:5:6 --prefix DNAseq_
↪Filtered -g y
```

1.4 Separate VCF by chromomsome using vcftools

```
#Separate VCF by Chromosome
vcftools --gzvcf DNAseq_Filtered_filt.vcf.gz --chr Chr01 --recode --out ../data/
↪Sorghumvcf/Chr01_DNAseq_Filtered_filt.vcf.gz
vcftools --gzvcf DNAseq_Filtered_filt.vcf.gz --chr Chr02 --recode --out ../data/
↪Sorghumvcf/Chr02_DNAseq_Filtered_filt.vcf.gz
vcftools --gzvcf DNAseq_Filtered_filt.vcf.gz --chr Chr03 --recode --out ../data/
↪Sorghumvcf/Chr03_DNAseq_Filtered_filt.vcf.gz
vcftools --gzvcf DNAseq_Filtered_filt.vcf.gz --chr Chr04 --recode --out ../data/
↪Sorghumvcf/Chr04_DNAseq_Filtered_filt.vcf.gz
vcftools --gzvcf DNAseq_Filtered_filt.vcf.gz --chr Chr05 --recode --out ../data/
↪Sorghumvcf/Chr05_DNAseq_Filtered_filt.vcf.gz
vcftools --gzvcf DNAseq_Filtered_filt.vcf.gz --chr Chr06 --recode --out ../data/
↪Sorghumvcf/Chr06_DNAseq_Filtered_filt.vcf.gz
vcftools --gzvcf DNAseq_Filtered_filt.vcf.gz --chr Chr07 --recode --out ../data/
↪Sorghumvcf/Chr07_DNAseq_Filtered_filt.vcf.gz
vcftools --gzvcf DNAseq_Filtered_filt.vcf.gz --chr Chr08 --recode --out ../data/
↪Sorghumvcf/Chr08_DNAseq_Filtered_filt.vcf.gz
vcftools --gzvcf DNAseq_Filtered_filt.vcf.gz --chr Chr09 --recode --out ../data/
↪Sorghumvcf/Chr09_DNAseq_Filtered_filt.vcf.gz
vcftools --gzvcf DNAseq_Filtered_filt.vcf.gz --chr Chr10 --recode --out ../data/
↪Sorghumvcf/Chr10_DNAseq_Filtered_filt.vcf.gz
```

1.5 VCF Configuration file

```
#Create a new vcf configuration file and fill it with this information(SorghumVcf.conf)

../data/Sorghumvcf/Chr01_DNAseq_Filtered_filt.vcf.gz.recode.vcf
../data/Sorghumvcf/Chr02_DNAseq_Filtered_filt.vcf.gz.recode.vcf
../data/Sorghumvcf/Chr03_DNAseq_Filtered_filt.vcf.gz.recode.vcf
../data/Sorghumvcf/Chr04_DNAseq_Filtered_filt.vcf.gz.recode.vcf
../data/Sorghumvcf/Chr05_DNAseq_Filtered_filt.vcf.gz.recode.vcf
../data/Sorghumvcf/Chr06_DNAseq_Filtered_filt.vcf.gz.recode.vcf
../data/Sorghumvcf/Chr07_DNAseq_Filtered_filt.vcf.gz.recode.vcf
../data/Sorghumvcf/Chr08_DNAseq_Filtered_filt.vcf.gz.recode.vcf
../data/Sorghumvcf/Chr09_DNAseq_Filtered_filt.vcf.gz.recode.vcf
../data/Sorghumvcf/Chr10_DNAseq_Filtered_filt.vcf.gz.recode.vcf
```

1.6 Create a Origin tab delimited file

```
#Create a new Origin tab delimited file and fill it with this information (SorghumOrigin.
↪tab)

con-all AA
D2      BB
```

1.7 Create a Color configuration file

```
#Create a new color configuration file and fill it with this information (SorghumColor.
↪conf)

AA 0 255 0
BB 255 0 0
```
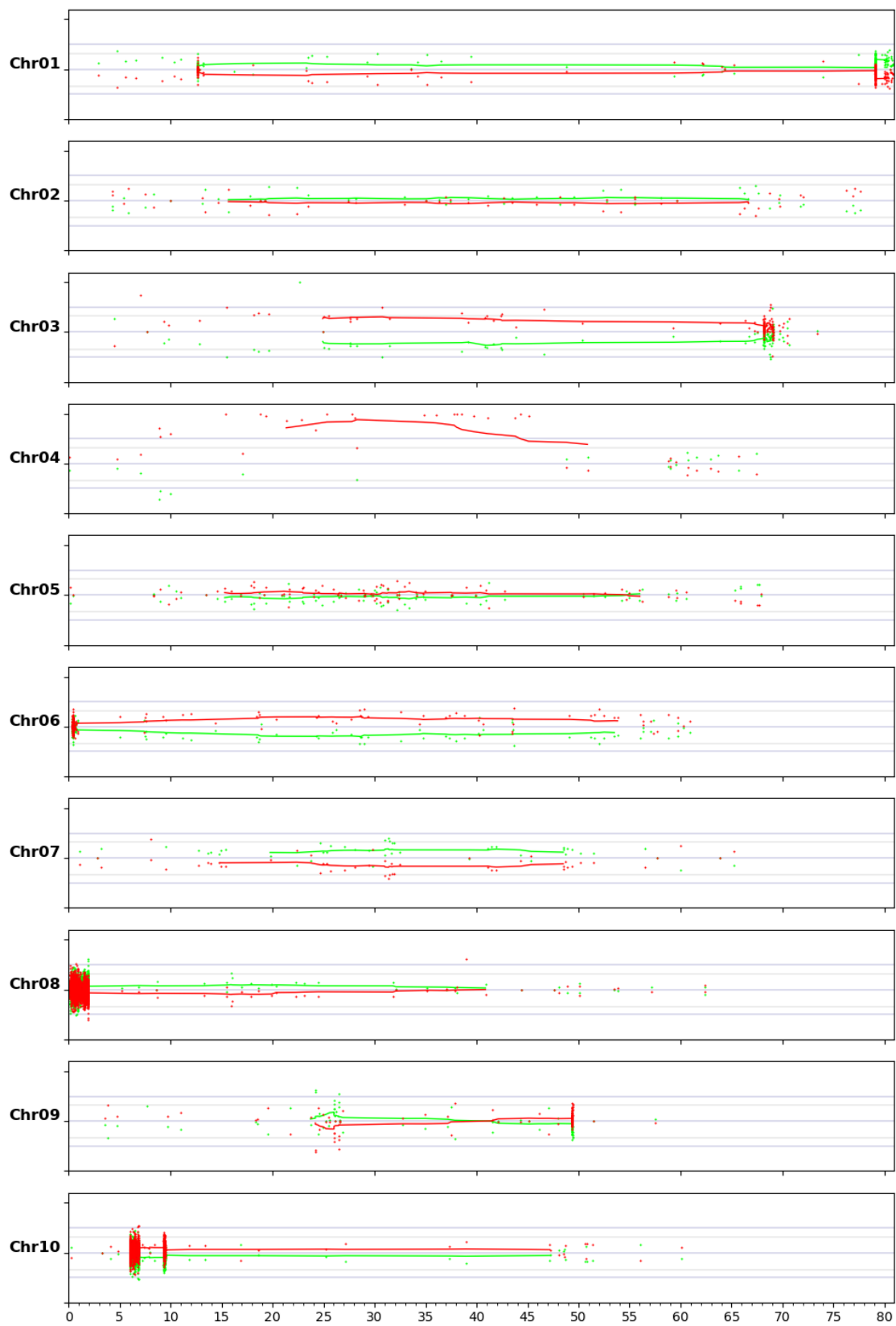
1.8 Run vcf2allPropandCov python script

```
#Run python script
python vcf2allPropAndCov.py --conf ../data/config/SorghumVcf.conf --origin ../data/
↪config/SorghumOrigin.tab --acc D2_F2_tt --ploidy 2 --dcurve y --col /data/config/
↪SorghumColor.conf
```

1.9 Plots
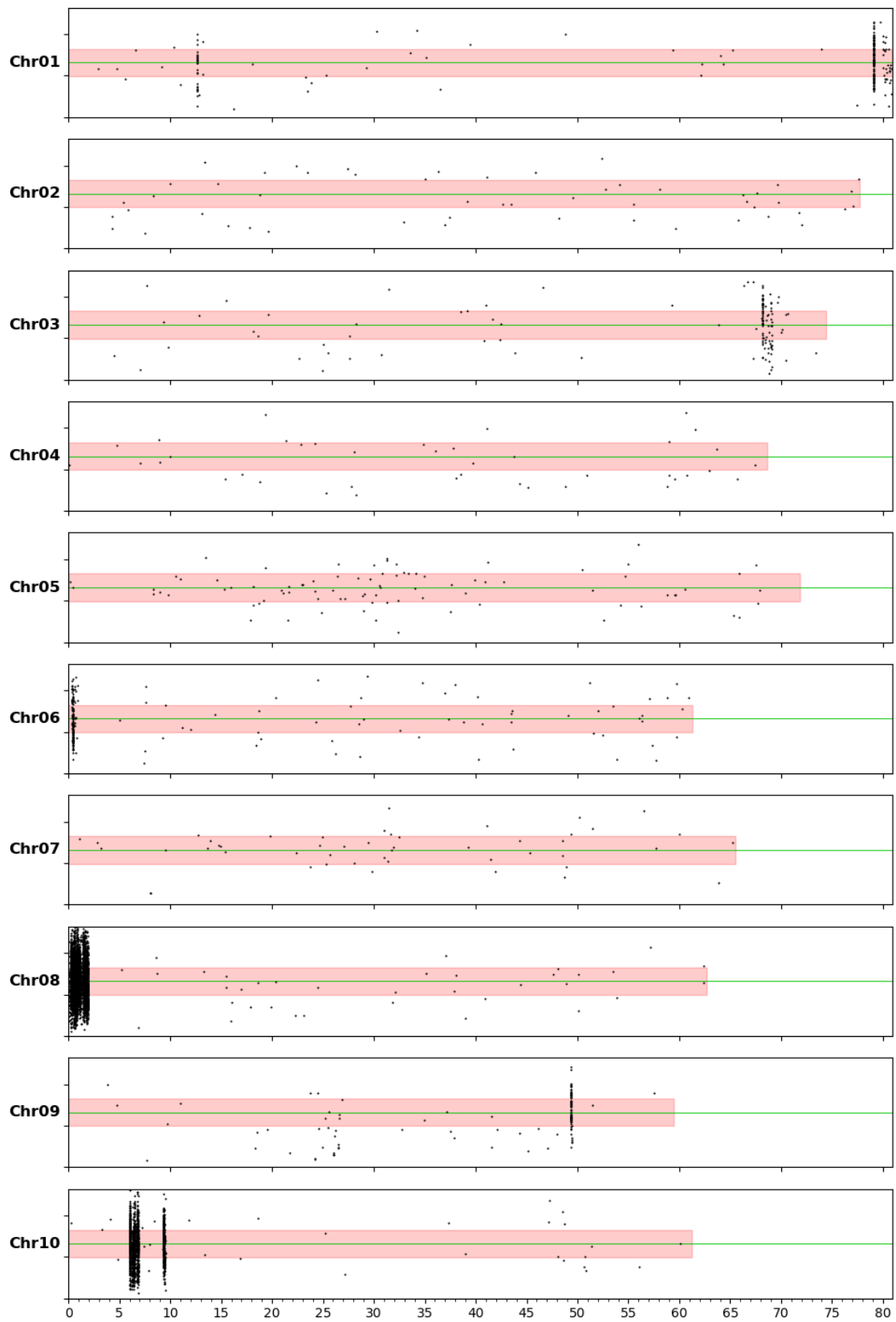
	V1	V2	V3	V4	V5
1193	Chr01	2943267	A	BB	0.4358974
1194	Chr01	2943267	T	AA	0.5641026
1195	Chr01	4720049	A	BB	0.3162393
1196	Chr01	4720049	G	AA	0.6837607
1197	Chr01	5567202	A	AA	0.5760870
1198	Chr01	5567202	G	BB	0.4239130
1199	Chr01	6582529	CG	BB	0.4135802
1200	Chr01	6582529	CCACTG	AA	0.5864198
1201	Chr01	9151187	GTTTTTTTTC	BB	0.3801653
1202	Chr01	9151187	GTTTTTTTTTTC	AA	0.6198347
1203	Chr01	10361552	G	BB	0.4285714
1204	Chr01	10361552	A	AA	0.5714286
1205	Chr01	10972432	A	AA	0.5949367
1206	Chr01	10972432	G	BB	0.4050633
1207	Chr01	12616225	A	BB	0.4537037
1208	Chr01	12616225	G	AA	0.5462963
1209	Chr01	12616497	G	BB	0.4869565
1210	Chr01	12616497	T	AA	0.5130435
1211	Chr01	12616805	T	BB	0.4520548
1212	Chr01	12616805	A	AA	0.5479452
1213	Chr01	12616990	GAAAAAAAAC	BB	0.4900662
1214	Chr01	12616990	GAAAAAAAAAC	AA	0.5099338
1215	Chr01	12617272	CGT	BB	0.4250000
1216	Chr01	12617272	CT	AA	0.5750000
1217	Chr01	12617582	G	BB	0.5874126
1218	Chr01	12617582	A	AA	0.4125874
1219	Chr01	12617785	T	BB	0.4973262
1220	Chr01	12617785	C	AA	0.5026738