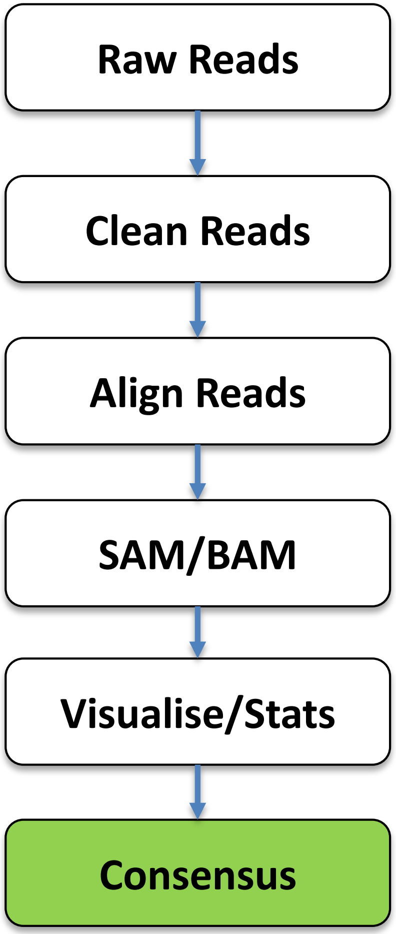
Practical session



- **Plan of Action**
- 15:00 – 16:00 – Consensus/Variant talk + same tutorial on GitHub

- 16:00 – 16:30 Break

- 16:30 – 18:00 Group practical
- Dengue-S
- Dengue-B

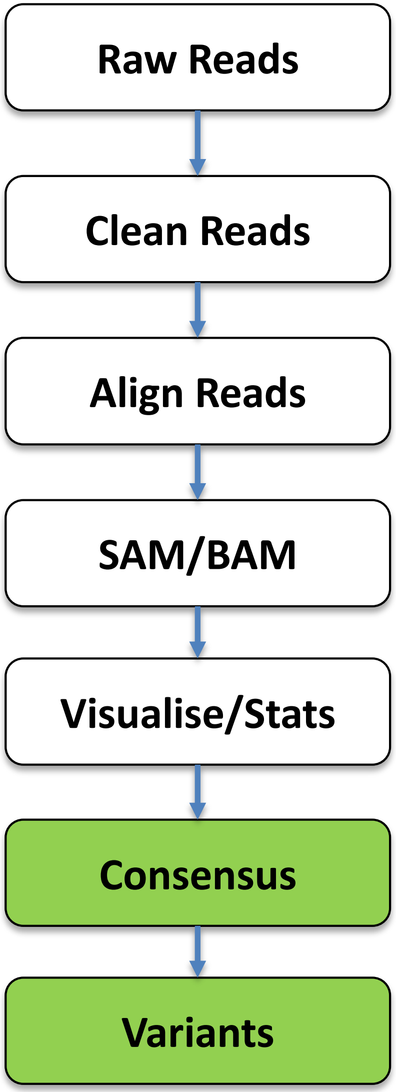
Consensus and Variant Calling

Richard Orton

MRC-University of Glasgow Centre for Virus Research

May 2024

Summary so far …


Raw Reads

Clean Reads

Align Reads

SAM/BAM

Visualise/Stats

Consensus

Variants

```
                1         2         3
Pos:   1234567890123456789012345678901234
Ref:   ACGGTGACACGTAGCAGTACGCGGGTTACACAGA
       ACGGCGA       CAGTTCG      AC-CAGA
             AGACGTA       GCGGGTT
                 GTAGCAGT       TTACACAG
          GCGACAC       TCGCGGG
        CGGCGAC        AGTTCGC     TACACAT
               ACG-AGC       GGGGTAC
Cov:   1223334333333233333343333344433331
```

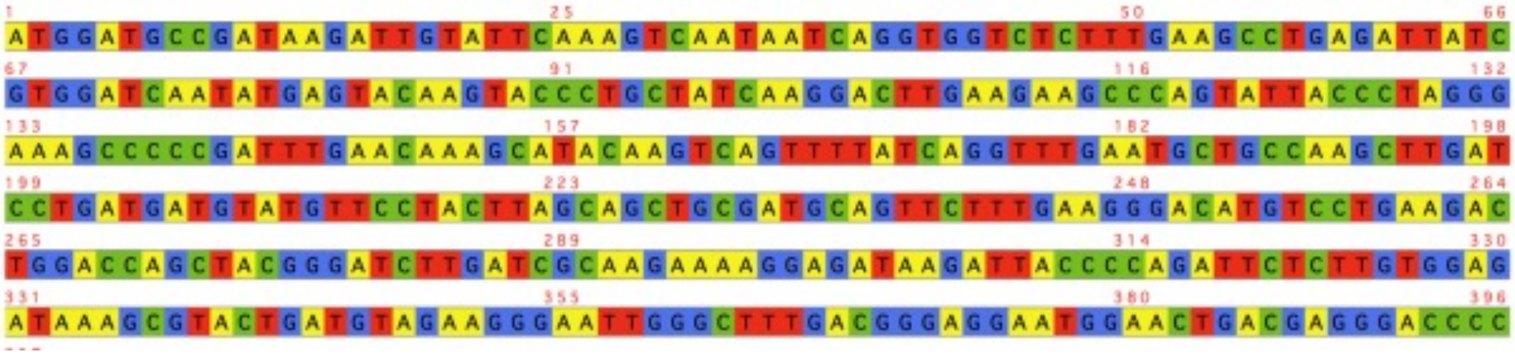
Consensus Sequence

- What is a consensus sequence?
  - At each genome position call the most frequent nucleotide observed

```
                 1         2         3
Pos:    1234567890123456789012345678901234
Ref:    ACGGTGACACGTAGCAGTACGCGGGTTACACAGA
        ACGGCGA       CAGTTCG      AC-CAGA
              AGACGTA       GCGGGTT
                  GTAGCAGT       TTACACAG
           GCGACAC        TCGCGGG
         CGGCGAC       AGTTCGC    TACACAT
                ACG-AGC       GGGGTAC
Cov:    1223334333333233333343333344433331
```

Consensus Sequence

- What is a consensus sequence?
  - At each genome position call the most frequent nucleotide observed



```
1                       25                       50             66
ATGGATGCCGATAAGATTGTATTCAAAGTCAATAATCAGGTGGTCTCTTTGAAGCCTGAGATTATC
67                      91                       116           132
GTGGATCAATATGAGTACAAGTACCCTGCTATCAAGGACTTGAAGAAGCCCAGTATTACCCTAGGG
133                     157                      182           198
AAAGCCCCCGATTTGAACAAAGCATACAAGTCAGTTTTATCAGGTTTGAATGCTGCCAAGCTTGAT
199                     223                      248           264
CCTGATGATGTATGTTCCTACTTAGCAGCTGCGATGCAGTTCTTTGAAGGGACATGTCCTGAAGAC
265                     289                      314           330
TGGACCAGCTACGGGATCTTGATCGCAAGAAAAGGAGATAAGATTACCCCAGATTCTCTTGTGGAG
331                     355                      380           396
ATAAAGCGTACTGATGTAGAAGGGAATTGGGCTTTGACGGGAGGAATGGAACTGACGAGGGACCCC
```

- Genome Position 253, Reference = C

A: 1000	A: 1	A: 750	A: 501	A: 700	A: 1
C: 0	C: 0	C: 0	C: 0	C: 300	C: 1
G: 0	G: 0	G: 250	G: 0	G: 0	G: 0
T: 0	T: 0	T: 0	T: 499	T: 0	T: 0
				Del: 750	

Sequence characters – IUPAC Codes

- The nucleic acid notation currently in use was first formalized by the International Union of Pure and Applied Chemistry (IUPAC) in 1970.

Symbol	Description	Bases Represented				Num
A	Adenine	A				1
C	Cytosine		C			1
G	Guanine			G		1
T	Thymine				T	1
U	Uracil				U	1
W	Weak	A			T	2
S	Strong		C	G		2
M	a**M**ino	A	C			2
K	**K**eto			G	T	2
R	pu**R**ine	A		G		2
Y	p**Y**rimidine		C		T	2
B	not A (**B** comes after A)		C	G	T	3
D	not C (**D** comes after C)	A		G	T	4
H	not G (**H** comes after G)	A	C		T	4
V	not T (**V** comes after T & U)	A	C	G		4
N	Any **N**ucleotide	A	C	G	T	4
-	Gap					0

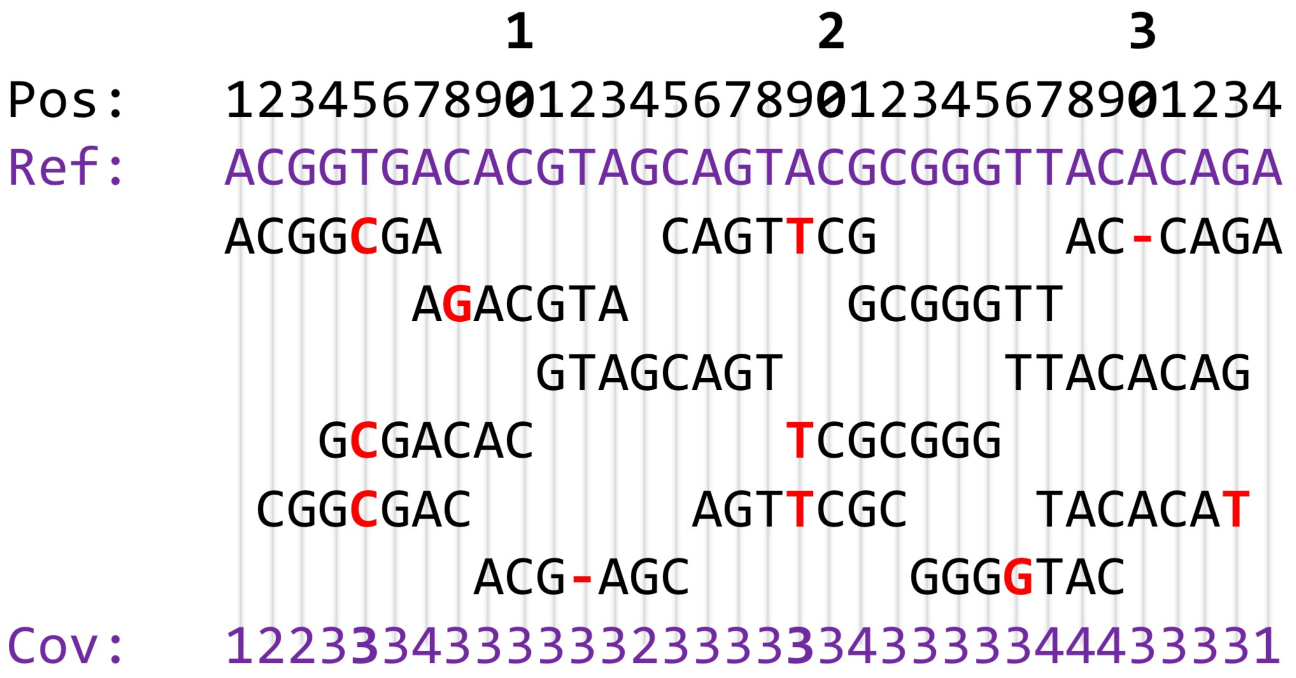
```
A: 750
C: 0
G: 250
T: 0

A: 501
C: 0
G: 0
T: 499

A: 700
C: 300
G: 0
T: 0
Del: 750
```

What do you need

- **BAM file**
  - Reads aligned to a reference
- **Reference file**
  - The reference file used in the BAM

```
                1         2         3
Pos:   1234567890123456789012345678901234
Ref:   ACGGTGACACGTAGCAGTACGCGGGTTACACAGA
       ACGGCGA       CAGTTCG      AC-CAGA
             AGACGTA       GCGGGTT
                 GTAGCAGT       TTACACAG
          GCGACAC       TCGCGGG
        CGGCGAC       AGTTCGC    TACACAT
               ACG-AGC       GGGGTAC
Cov:   1223334333333233333343333344433331
```

Pileup the data – samtools mpileup



```
                1         2         3
Pos:   1234567890123456789012345678901234
Ref:   ACGGTGACACGTAGCAGTACGCGGGTTACACAGA
       ACGGCGA       CAGTTCG      AC-CAGA
             AGACGTA       GCGGGTT
                 GTAGCAGT       TTACACAG
          GCGACAC        TCGCGGG
        CGGCGAC       AGTTCGC    TACACAT
               ACG-AGC       GGGGTAC
Cov:   1223334333333233333343333344433331
```

Ref	Pos	Cov	RefBase	Bases	Qualities
HQ156345.1	1	1	A	A	I
HQ156345.1	2	2	C	CC	IH
HQ156345.1	3	2	G	GG	IB
HQ156345.1	4	3	G	GGG	CCB
HQ156345.1	5	3	T	**CCC**	CID
HQ156345.1	6	3	G	GGG	FFF
HQ156345.1	7	4	A	AAAA	IIEE
HQ156345.1	8	3	C	**G**CC	III

Pileup the data – samtools mpileup

Ref	Pos	Reads	RefBase	Bases	Qualities
HQ156345.1	1	1	A	A	I
HQ156345.1	2	2	C	CC	IH
HQ156345.1	3	2	G	GG	IB
HQ156345.1	4	3	G	GGG	CCB
HQ156345.1	5	3	T	**CCC**	CID
HQ156345.1	6	3	G	GGG	FFF
HQ156345.1	7	4	A	AAAA	IIEE
HQ156345.1	8	3	C	**G**CC	III

. = match to the reference base in forward direction

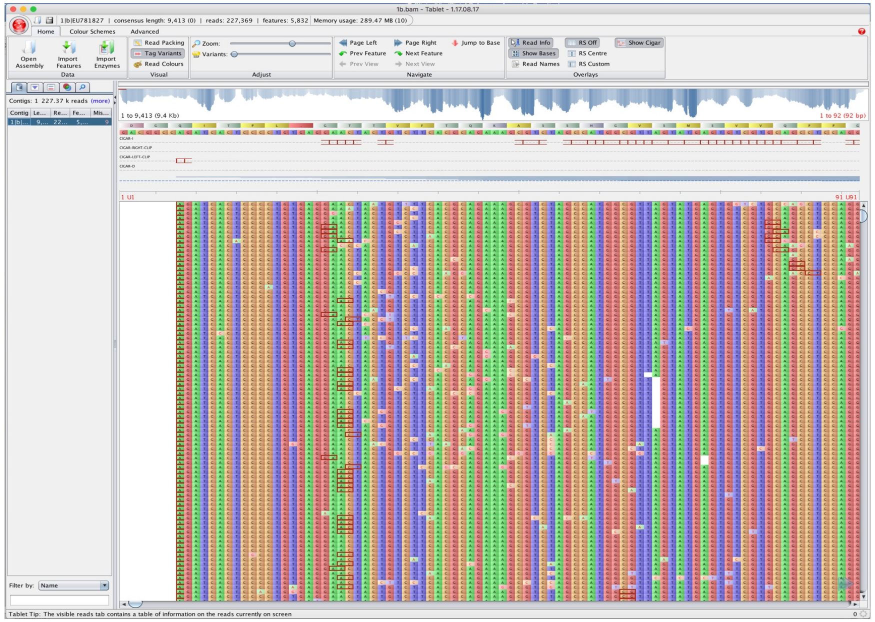
, = match to the reference base in reverse direction

ACGTN = mismatch to ref in forward direction

acgtn = mismatch to ref in reverse direction

Ref	Pos	Reads	RefBase	Bases	Qualities
HQ156345.1	1	1	A	.	I
HQ156345.1	2	2	C	.,	IH
HQ156345.1	3	2	G	.,	IB
HQ156345.1	4	3	G	.,.	CCB
HQ156345.1	5	3	T	CcC	CID
HQ156345.1	6	3	G	.,.	FFF
HQ156345.1	7	4	A	.,..	IIEE
HQ156345.1	8	3	C	G.,	III

samtools mpileup command



```
samtools mpileup -aa -d 0 -Q 0 -B –A my.bam > my_mpileup.txt
```

Ref	Pos	Reads	RefBase	Bases	Qualities
HQ156345.1	1	1	A	.	I
HQ156345.1	2	2	C	.,	IH
HQ156345.1	3	2	G	.,	IB
HQ156345.1	4	3	G	.,.	CCB
HQ156345.1	5	3	T	CcC	CID
HQ156345.1	6	3	G	.,.	FFF
HQ156345.1	7	4	A	.,..	IIEE

iVar: https://github.com/andersen-lab/ivar

- We will be using the iVar consensus caller in this practical
  - Used alot for SARS-CoV-2 data
  - Can also be used for trimming amplicon primers based on BAM alignment co-ordinates

- iVar uses samtools mpileup to feed data in
  - `samtools mpileup -aa –A -d 0 -Q 0 my.bam | ivar consensus -p myseq`

- **-aa:** output data for **a**ll positions (even positions with zero coverage)
- **-A:** don’t discount orphan reads (not in a pair)
- **-d 0:** disable the maximum depth to report [default is 8000]
- **-Q 0:** minimum base quality 0

- **my.bam:** the name of the bam file

- **|:** pipe/pass the data/results/output into the next command

- **ivar:** the name of the program we are using
- **consensus:** the name of the function within ivar we are using
- **-p myseq:** the prefix of the output file that ivar will create -> **myseq.fasta**

Consensus sequences - alternatives

- bcftools (with bedtools to mask low coverage regions)

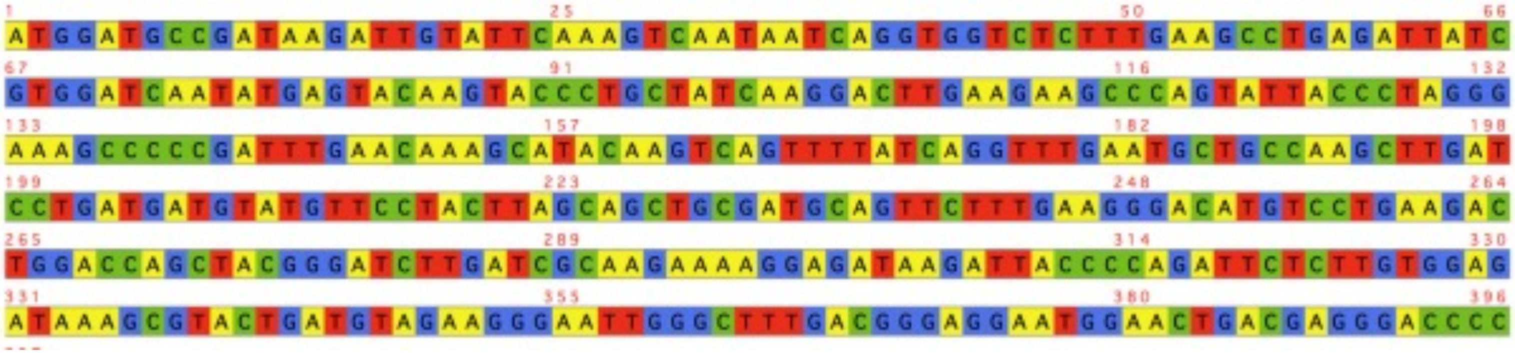
- **samtools** consensus

- VirusConsensus
- https://github.com/niemasd/ViralConsensus

- ConsensusFixer
- https://github.com/cbg-ethz/ConsensusFixer

- Kindel
- https://github.com/bede/kindel

- VarScan2
- http://varscan.sourceforge.net


1 25 50 66
ATGGATGCCGATAAGATTGTATTCAAAGTCAATAATCAGGTGGTCTCTTTGAAGCCTGAGATTATC
67 91 116 132
GTGGATCAATATGAGTACAAGTACCCTGCTATCAAGGACTTGAAGAAGCCCAGTATTACCCTAGGG
133 157 182 198
AAAGCCCCCGATTTGAACAAAGCATACAAGTCAGTTTTATCAGGTTTGAATGCTGCCAAGCTTGAT
199 223 248 264
CCTGATGATGTATGTTCCTACTTAGCAGCTGCGATGCAGTTCTTTGAAGGGACATGTCCTGAAGAC
265 289 314 330
TGGACCAGCTACGGGATCTTGATCGCAAGAAAAGGAGATAAGATTACCCCAGATTCTCTTGTGGAG
331 355 380 396
ATAAAGCGTACTGATGTAGAAGGGAATTGGGCTTTGACGGGAGGAATGGAACTGACGAGGGACCCC

Consensus sequences – minimum coverage?

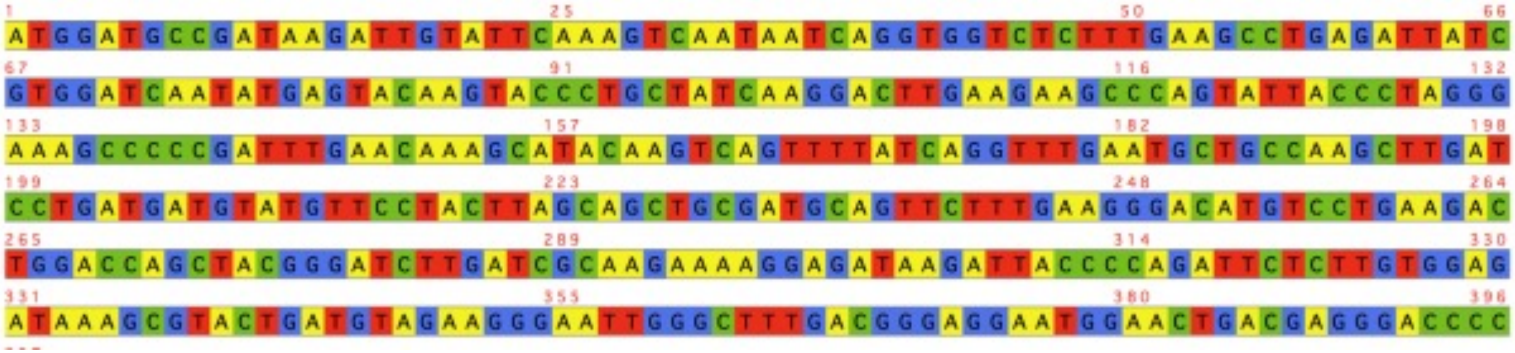
- iVar default is 10 – illumina - has been used a lot for ARTIC SARS-CoV-2 samples

- How much data do you have?
- How desperate are you to get (any) sequence?

- 100 is a strong threshhold
- 20 is a good threshold
- 10 is a decent threshold
- 5 is a weak threshold [high quality, low ambiguity]
- 2 & 1 are desperate thresholds

- Low coverage = potential for many ambiguities: coverage 5, 3As 2 Ts -> consensus = A, but very high probability the consensus could have been T

- The 5’/3’ ends of genomes/segments are typically poorly covered – RACE
  - RACE: Rapid amplification of cDNA ends

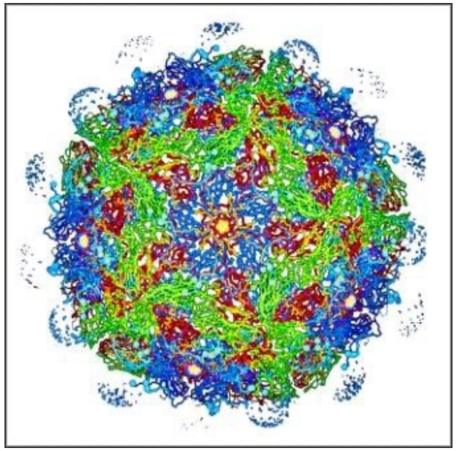

1 25 50 66
ATGGATGCCGATAAGATTGTATTCAAAGTCAATAATCAGGTGGTCTCTTTGAAGCCTGAGATTATC
67 91 116 132
GTGGATCAATATGAGTACAAGTACCCTGCTATCAAGGACTTGAAGAAGCCCAGTATTACCCTAGGG
133 157 182 198
AAAGCCCCCGATTTGAACAAAGCATACAAGTCAGTTTTATCAGGTTTGAATGCTGCCAAGCTTGAT
199 223 248 264
CCTGATGATGTATGTTCCTACTTAGCAGCTGCGATGCAGTTCTTTGAAGGGACATGTCCTGAAGAC
265 289 314 330
TGGACCAGCTACGGGATCTTGATCGCAAGAAAAGGAGATAAGATTACCCCAGATTCTCTTGTGGAG
331 355 380 396
ATAAAGCGTACTGATGTAGAAGGGAATTGGGCTTTGACGGGAGGAATGGAACTGACGAGGGACCCC

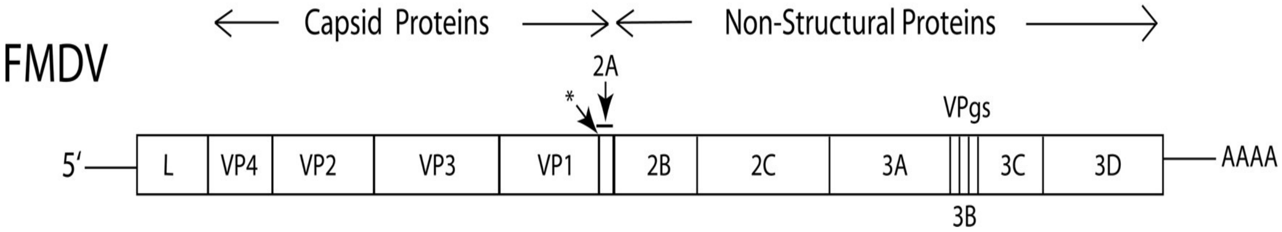
Variant calling

- Consensus level (e.g. >50%)

- Low frequency variants (25%, 10%, 1%, 0.1%)

Viral Populations





- Small, compact genomes – gives high depth with HTS

- Mutation rate ~10^{-4} mutations per nucleotide per transcription cycle – every genome replication introduces new mutations

- **Evolve rapidly**: Large population size, high replication rate, error prone RNA polymerase

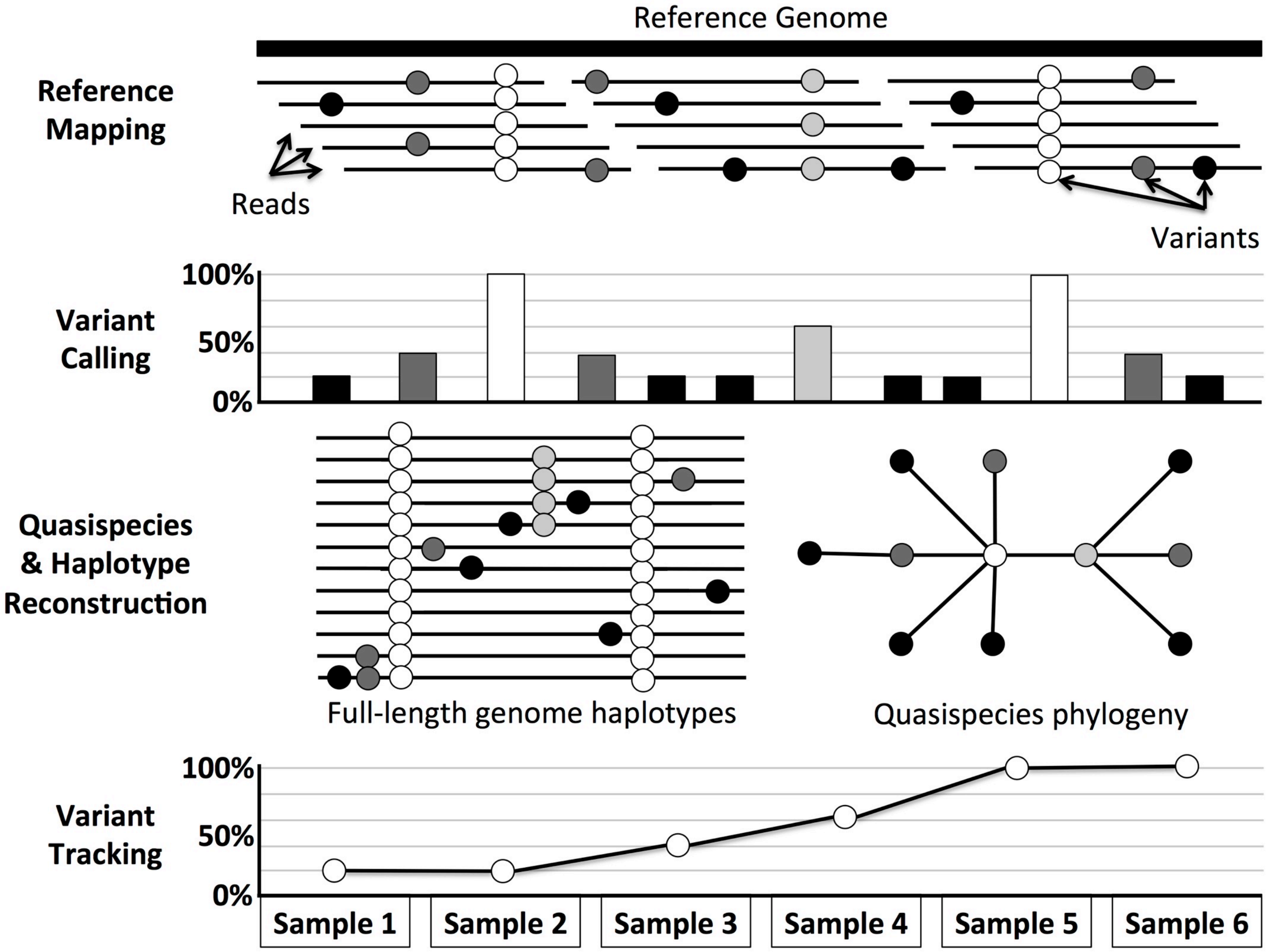
- Enables them to rapidly adapt to new (host) environments and selective pressures such as drug treatments

- Exist within their hosts as large, complex and heterogeneous populations

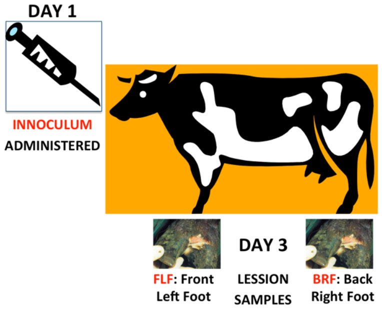
- Comprising a spectrum of related but non-identical genome sequences termed the **quasispecies**.

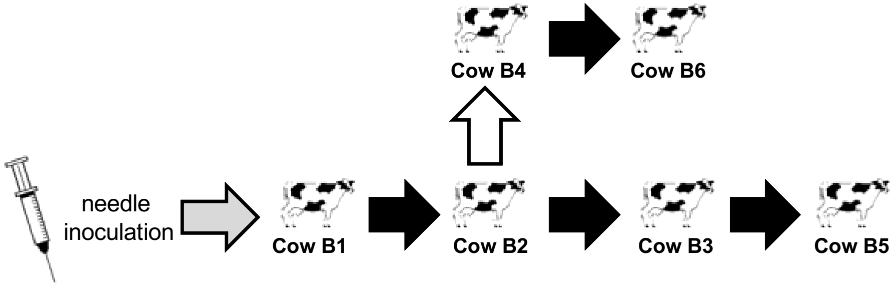
http://www7.nationalgeographic.com/ngm/0707/feature5/index.html
https://i.pinimg.com/originals/d4/ae/cf/d4aecfdda061a49313043e0397e780de.jpg
https://www.pirbright.ac.uk/news/2017/12/protein-shown-vital-foot-and-mouth-disease-virus-replication

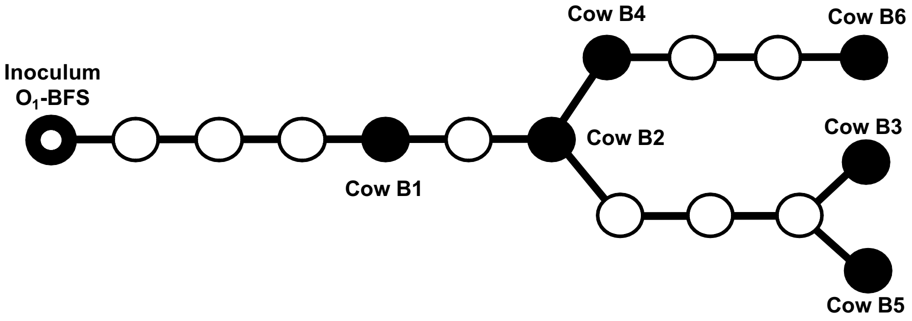
HTS Applications



Viral mutation tracking



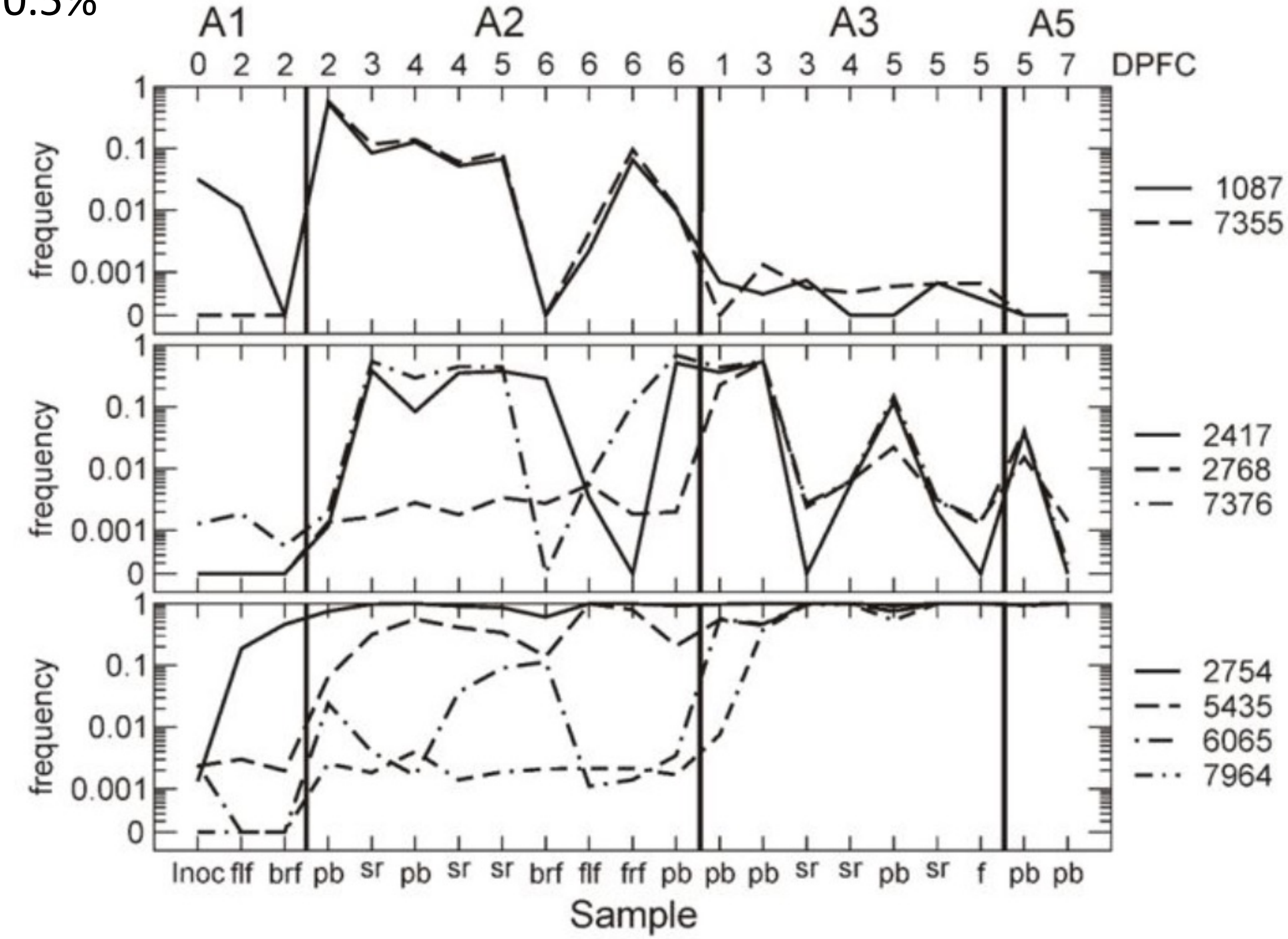




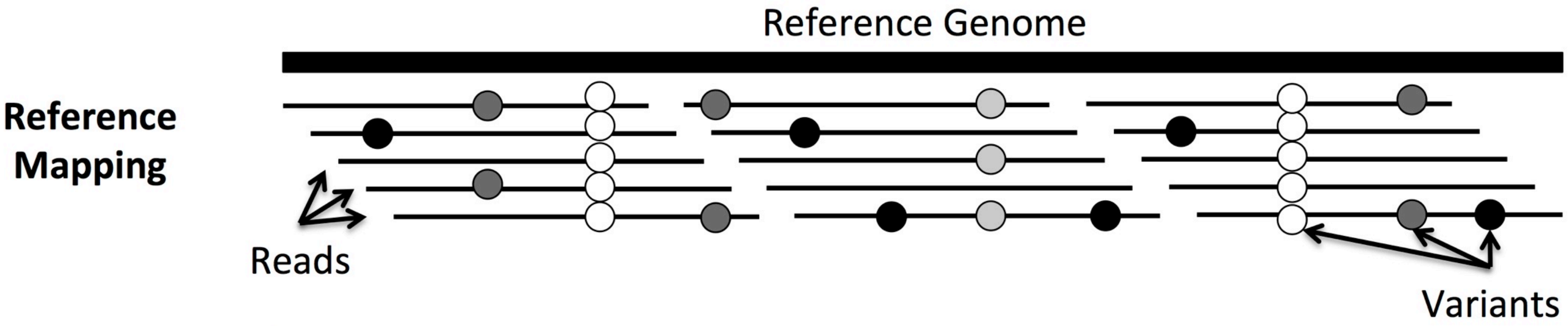
Wright et al. (2011): **Beyond the consensus:** dissecting within-host viral population diversity of FMDV by using NGS

Viral mutation tracking

- Morelli et al 2013. BMC Veterinary Medicine; 44: 12
- Min coverage of 1000, quality 30 filtering
- Sequenced in duplicate
- PCR control data 0.5%



Sequence Errors



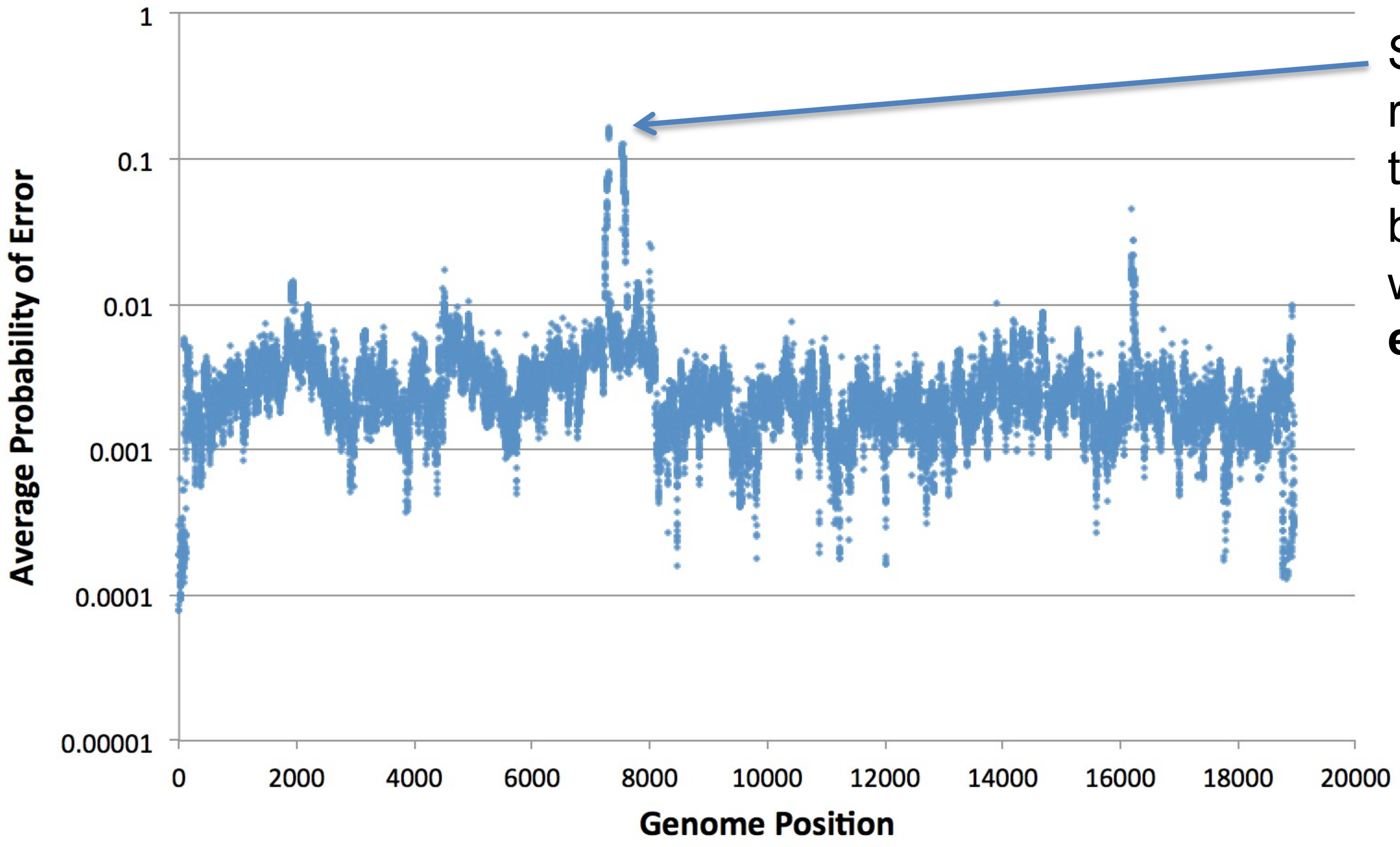
- Sequencing errors make it hard to identify low frequency variants in the population

- Coverage of 20,000
- All at Q40 (P=0.0001, 0.01%)
- Expect 2 errors (variant frequency 0.01%)

- Coverage of 20,000
- All at Q30 (P=0.001, 0.1%)
- Expect 20 errors (variant frequency 0.1%)

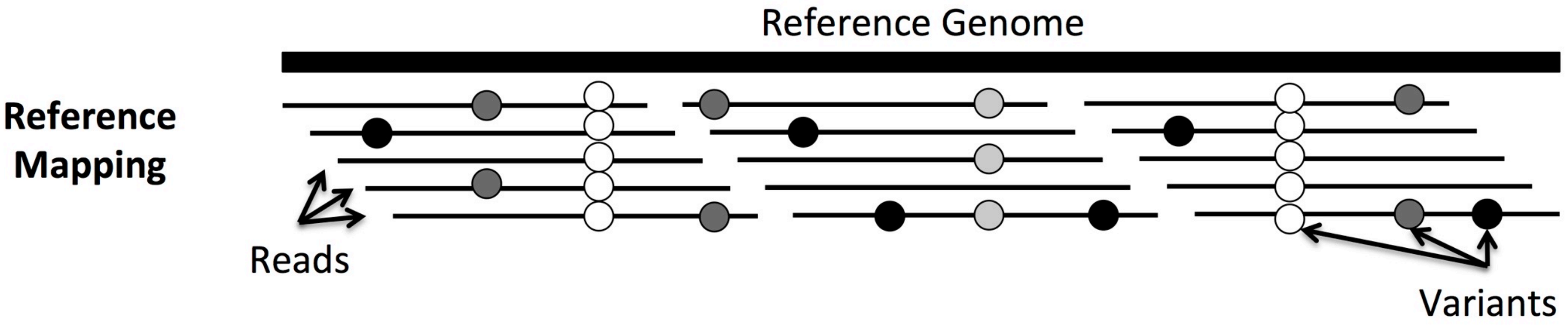
Genome Position Specific

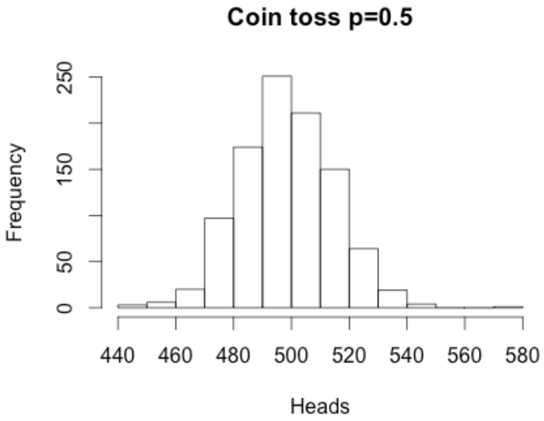
- At each genome position, sum all the Q score probabilities of every base aligned there, calculate average probability of error

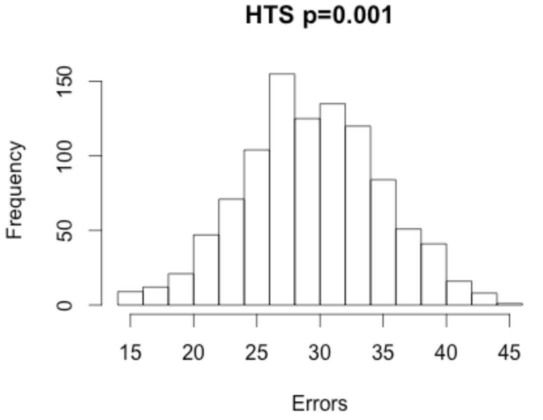


Some positions- regions more prone to **poor quality** bases and therefore will have more **errors**

Basic Modelling of Sequence Errors







- Binomial probability distribution- number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p.

- Coin toss
  - n = number of throws (1000)
  - p = probability of getting a head (0.5, 50%)
  - 1000 replicates
  - On **average** you will observe 500 heads (50%) - **variation**

- HTS
  - n = coverage at genome position (30,000)
  - p = average error probability at genome position (0.001 = 0.1%)
  - 1000 replicates
  - On **average** you will observe 30 errors (0.1%) – **variation**

Systematic Errors

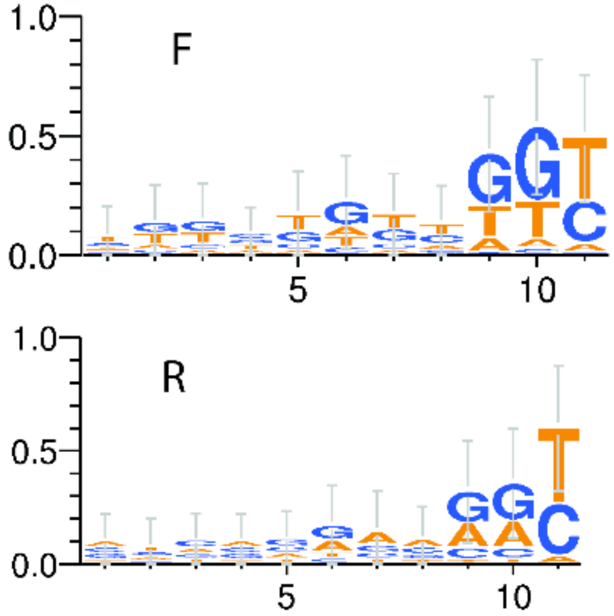
- Defined as an error associated with a flaw in the equipment or experiment

- 454 has well known problems with homopolymers – MinION deletions as well

- Illumina errors more likely at some sequence motifs than others – typically independent of Q scores
  - Meacham et al 2011 – BMC Bioinformatics
  - Nakamura et al 2011 – Nucl Acids Res
  - Li et al 2012 – Genome Biology
  - MIRA manual

- Inverted Repeats
- At or downstream of GGC, GGT, GGX – **not fully understood**

- Compounds variant identification – especially at lower frequencies



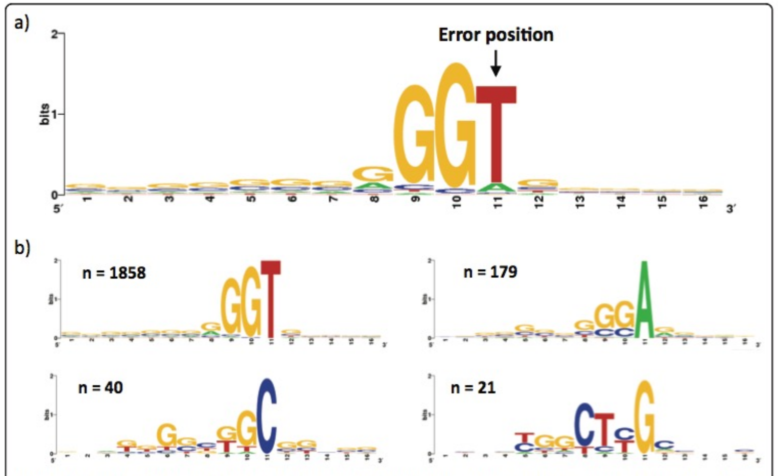


Figure 3 Sequence motifs at systematic error sites. (a) The motif around systematic errors reveals a strong enrichment for instances preceded by an occurrence of *GG* and for the error to occur at locations where the reference genome is *T*. (b) Categorized by the nucleotide at the error location. The number of systematic errors in each subset is denoted by *n*.

Strand Bias Examples

- A true variant should appear equally on reads going in both directions - can be used to identify systematic errors

- Ref Seq
- ACCGTAGCCTGGTATGTACGTAG

- Fwd Reads →
- ACCGTAGCCT**GGg**ATGTACGTAG
- ACCGTAGCCT**GGg**ATGTACGTAG
- ACCGTAGCCT**GGg**ATGTACGTAG
- ACCGTAGCCT**GGg**ATGTACGTAG
- ACCGTAGCCT**GGg**ATGTACGTAG

- Rev Reads ←
- TGGCATCGGACC**ATA**CATGCATC
- TGGCATCGGACC**ATA**CATGCATC
- TGGCATCGGACC**ATA**CATGCATC
- TGGCATCGGACC**ATA**CATGCATC

Ref Seq
ACGTACGTACGTTTTTTTACGTACGT

Fwd Reads →
ACGTACGTACG**TTTTTTTt**CGTACGT
ACGTACGTACG**TTTTTTTt**CGTACGT
ACGTACGTACG**TTTTTTTt**CGTACGT
ACGTACGTACG**TTTTTTTt**CGTACGT
ACGTACGTACG**TTTTTTTt**CGTACGT

Rev Reads ←
TGCATGCATGCAAAAAAA**TGCATGCA**
TGCATGCATGCAAAAAAA**TGCATGCA**
TGCATGCATGCAAAAAAA**TGCATGCA**
TGCATGCATGCAAAAAAA**TGCATGCA**

- Statistical tests for strand bias such as Fisher's exact test
- Does the variant show the same/similar bias to the reference

Variant Calling - LoFreq

- LoFreq
  - Fast and sensitive variant calling from next-gen sequencing data
  - https://csb5.github.io/lofreq/
  - Wilm et al. 2012

- LoFreq is an advanced variant caller that use a range of data to call variants
  - Quality Scores
  - Transition/Transversion mutation matrix
  - Strand Bias test

- Python
  - BAM file
  - Reference file
  - Output VCF file

- `lofreq call -f ref.fasta -o var.vcf my.bam`

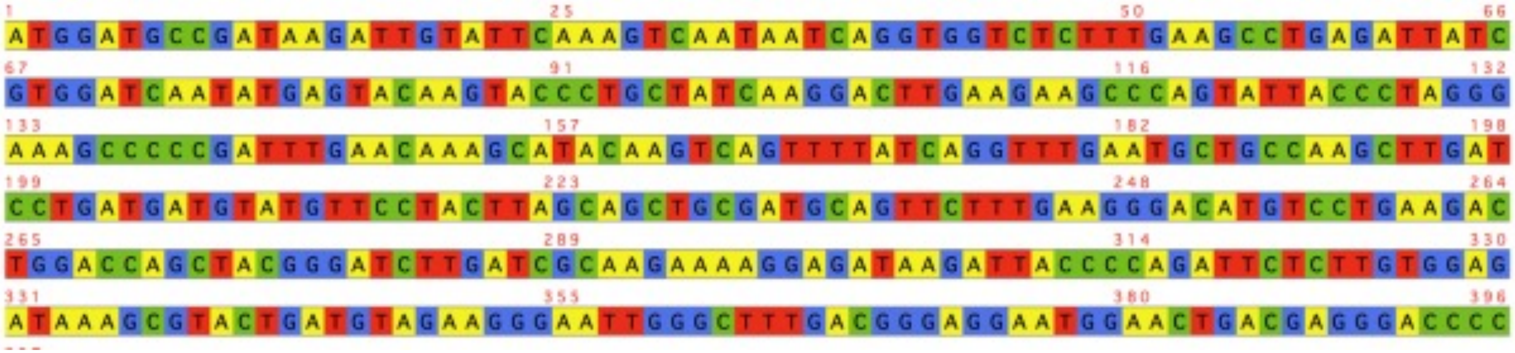
- more/less/cat var.vcf

Variant calling– minimum coverage & frequency

- A good first step is to use LoFreq or another variant caller and trust the results

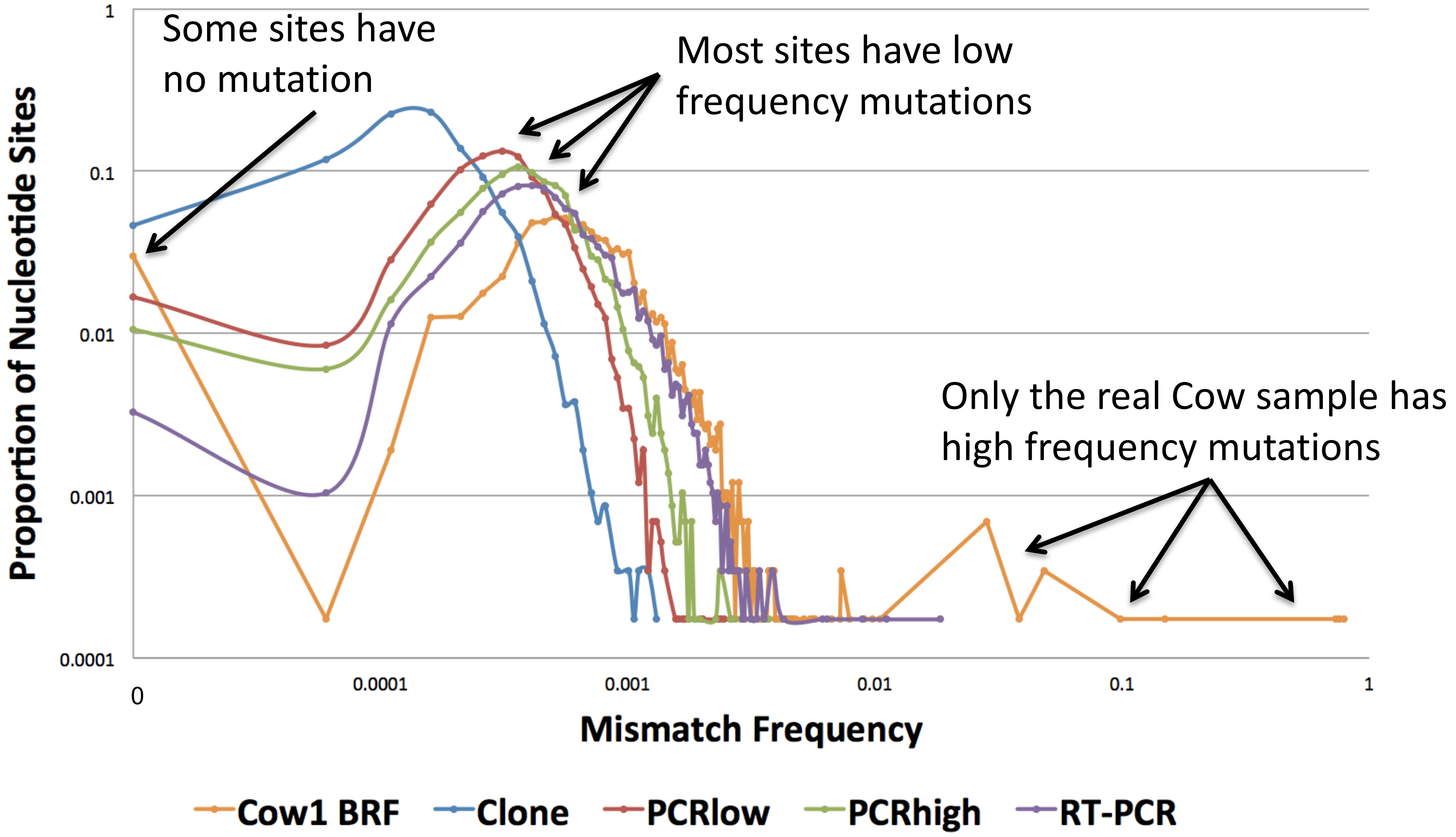
- You can further filter the results
  - Depth (DP) of the site > 1000
  - Only trust variants above 0.5% or 1%
  - Some tools allow base quality filtering e.g. Q30

- LoFreq does not account for RT-PCR errors



```
1                       25                       50             66
ATGGATGCCGATAAGATTGTATTCAAAGTCAATAATCAGGTGGTCTCTTTGAAGCCTGAGATTATC
67                      91                       116            132
GTGGATCAATATGAGTACAAGTACCCTGCTATCAAGGACTTGAAGAAGCCCAGTATTACCCTAGGG
133                     157                      182            198
AAAGCCCCCGATTTGAACAAAGCATACAAGTCAGTTTTATCAGGTTTGAATGCTGCCAAGCTTGAT
199                     223                      248            264
CCTGATGATGTATGTTCCTACTTAGCAGCTGCGATGCAGTTCTTTGAAGGGACATGTCCTGAAGAC
265                     289                      314            330
TGGACCAGCTACGGGATCTTGATCGCAAGAAAAGGAGATAAGATTACCCCAGATTCTCTTGTGGAG
331                     355                      380            396
ATAAAGCGTACTGATGTAGAAGGGAATTGGGCTTTGACGGGAGGAATGGAACTGACGAGGGACCCC
```

Clone Results



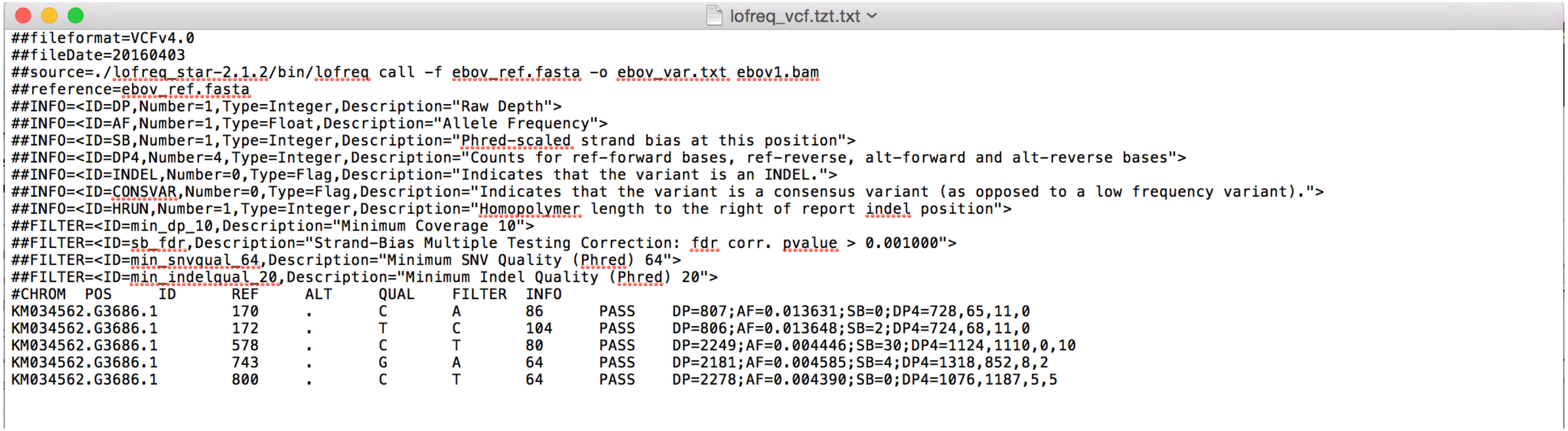
Some sites have no mutation

Most sites have low frequency mutations

Only the real Cow sample has high frequency mutations

Variant Call Format - .vcf files

- Many tools output this data in the VCF (Variant Call Format)
- Text file – tab delimited: more/less/cat/head/tail



```
##fileformat=VCFv4.0
##fileDate=20160403
##source=./lofreq_star-2.1.2/bin/lofreq call -f ebov_ref.fasta -o ebov_var.txt ebov1.bam
##reference=ebov_ref.fasta
##INFO=<ID=DP,Number=1,Type=Integer,Description="Raw Depth">
##INFO=<ID=AF,Number=1,Type=Float,Description="Allele Frequency">
##INFO=<ID=SB,Number=1,Type=Integer,Description="Phred-scaled strand bias at this position">
##INFO=<ID=DP4,Number=4,Type=Integer,Description="Counts for ref-forward bases, ref-reverse, alt-forward and alt-reverse bases">
##INFO=<ID=INDEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
##INFO=<ID=CONSVAR,Number=0,Type=Flag,Description="Indicates that the variant is a consensus variant (as opposed to a low frequency variant).">
##INFO=<ID=HRUN,Number=1,Type=Integer,Description="Homopolymer length to the right of report indel position">
##FILTER=<ID=min_dp_10,Description="Minimum Coverage 10">
##FILTER=<ID=sb_fdr,Description="Strand-Bias Multiple Testing Correction: fdr corr. pvalue > 0.001000">
##FILTER=<ID=min_snvqual_64,Description="Minimum SNV Quality (Phred) 64">
##FILTER=<ID=min_indelqual_20,Description="Minimum Indel Quality (Phred) 20">
#CHROM  POS     ID      REF     ALT     QUAL    FILTER  INFO
KM034562.G3686.1        170     .       C       A       86      PASS    DP=807;AF=0.013631;SB=0;DP4=728,65,11,0
KM034562.G3686.1        172     .       T       C       104     PASS    DP=806;AF=0.013648;SB=2;DP4=724,68,11,0
KM034562.G3686.1        578     .       C       T       80      PASS    DP=2249;AF=0.004446;SB=30;DP4=1124,1110,0,10
KM034562.G3686.1        743     .       G       A       64      PASS    DP=2181;AF=0.004585;SB=4;DP4=1318,852,8,2
KM034562.G3686.1        800     .       C       T       64      PASS    DP=2278;AF=0.004390;SB=0;DP4=1076,1187,5,5
```

Variant Call Format

Ref	Pos	Reads	RefBase	Bases	Qualities
HQ156345.1	1	1	A	.	I
HQ156345.1	2	2	C	.,	IH
HQ156345.1	3	2	G	.,	IB
HQ156345.1	4	3	G	.,.	CCB
HQ156345.1	5	3	T	CcC	CID
HQ156345.1	6	3	G	.,.	FFF
HQ156345.1	7	4	A	.,..	IIEE
HQ156345.1	8	3	C	G.,	III

CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO
HQ156345.1	5	.	T	C	60	PASS	DP=3;AF=1;SB=0;DP4=0,0,2,1
HQ156345.1	8	.	C	G	58	PASS	DP=3,AF=0.333;SB=0;DP4=1,1,1,0

CHROM: Chromosome Name (i.e. viral genome name)
POS: Position the viral genome
ID: Name of the known variant from DB
REF: The reference allele (i.e. the reference base)
ALT: The alternate allele (i.e. the variant/mutation observed)
QUAL: The quality of the variant on a Phred scale
FILTER: Did the variant pass the tools filters
INFO: Information

- **DP** = Depth
- **AF** = Alternate Allele Frequency
- **SB** = Strand Bias P-value
- **DP4**=Extra Depth Information (forward ref; reverse ref; forward non-ref; reverse non-ref)

Variant Calling – alternatives

- LoFreq

- iVar

- VPhaser2 - old
- https://www.broadinstitute.org/viral-genomics/v-phaser-2

- SegminatorII - old
- http://www.bioinf.manchester.ac.uk/segminator/

- VarScan2
- http://varscan.sourceforge.net

- **Freebayes**
- https://github.com/freebayes/freebayes

- SNVer
- http://snver.sourceforge.net

SNPeff

- SNPeff
  - Takes variants in VCF format
  - Uses information of gene start/stop co-ordinates
  - Adds synonymous/missense annotations to the vcf

- `snpEff -ud 0 NC_045512.2 sars2.vcf > sars2_snpeff.vcf`

- **-ud 0:** Set upstream downstream interval length to 0
- **NC_045512.2:** the reference **name** – not filename
- **sars2.vcf:** the input vcf file name
- **sars2_snpeff.vcf:** the annotated output vcf file name

SNPeff

```
CHROM        POS   ID    REF   ALT   QUAK  FILTER    INFO
NC_045512.2 23403.     A     G     44968.0    PASS DP=1286;AF=0.995334;SB=0;DP4=1,1,628,652
```

- DP=1286;AF=0.995334;SB=0;DP4=1,1,628,652;ANN=G|**missense_variant**|MODERATE|**S**|GU280_gp02|transcript|GU280_gp02|protein_coding|1/1|c.1841A>G|**p.Asp614Gly**|1841/3822|1841/3822|614/1273||

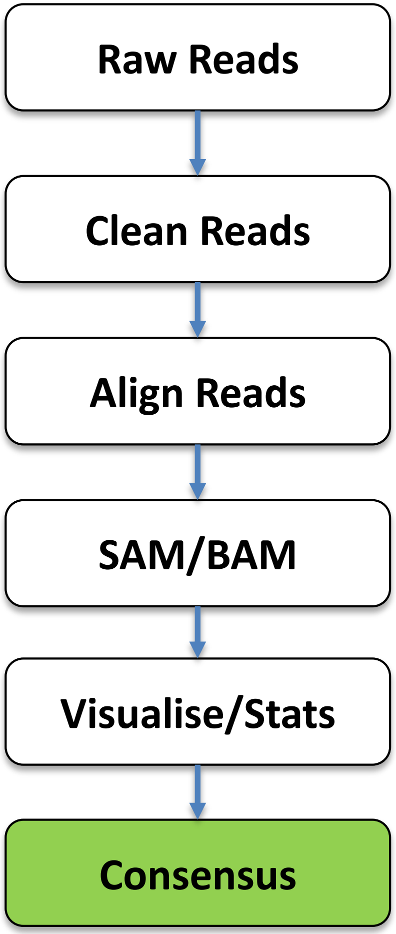
;ANN=G|	
missense_variant|	**Annotation**
MODERATE|	Annotation Impact
S|	**Gene Name**
GU280_gp02|	GeneID
transcript|	Feature Type
GU280_gp02|	FeatureID
protein_coding|	Transcript BioType
1/1|	Rank
c.1841A>G|	HGVS.coding [Human Genome Variation Society]
p.Asp614Gly|	**HGVS.protein**
1841/3822|	cDNA.pos / cDNA.length
1841/3822	CDS.pos / CDS.length
|614/1273|	AA.pos / AA.length
|	

- The A to G mutation at genome position **23403** corresponds to position **1841** (out of **3822**) within the Spike (S) gene which corresponds to codon **614** (out of **1273**) within Spike(S)

SNPeff alternatives

- SNPeff alternatives
  - vcf-annotator: https://github.com/rpetit3/vcf-annotator

Practical session



- DENV3 sample
- denv3.bam

- **ivar** -> consensus

- **lofreq** -> variants

- **conda** activate snpeff

- **snpeff** to add characterisation to SNPs

- Extra things:
  - Check the ivar manual – call variants on the denv3.bam – how do they compare to lofreq

- **Plan of Action**
- 15:00 – 16:00 – Same tutorial on GitHub
- 16:00 – 16:30
- 16:30 – 18:00 Group practical

The end