Bovine viral diarrhoea virus loses quasispecies diversity rapidly in culture
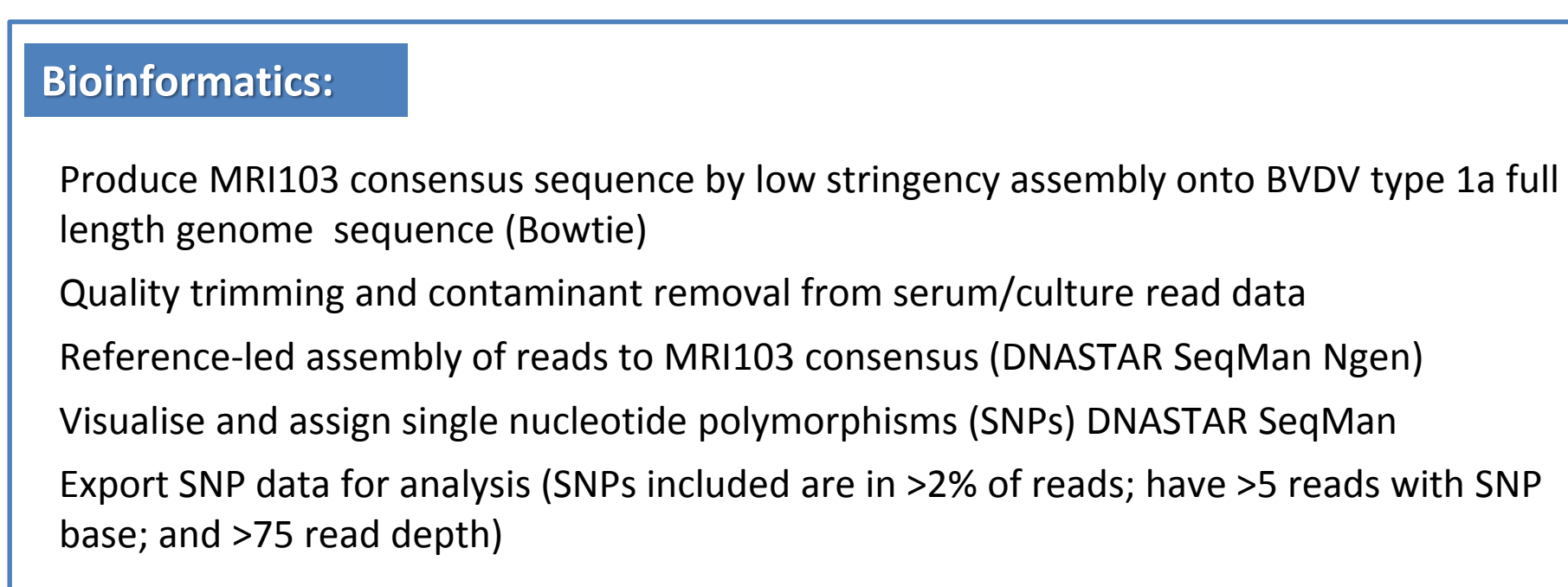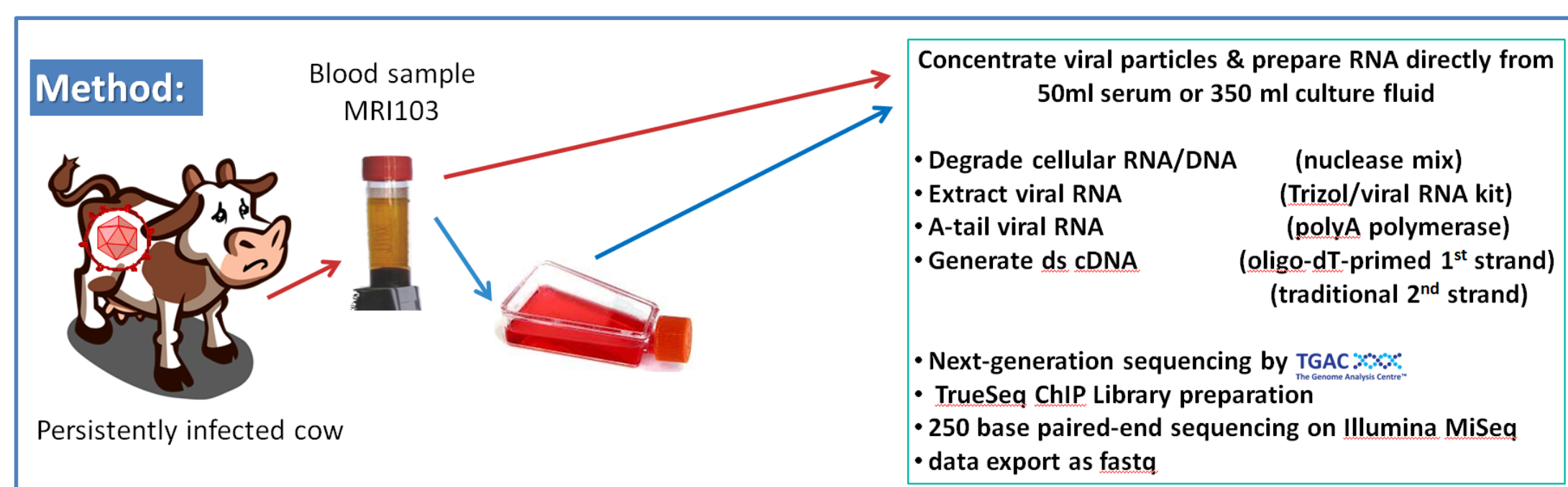
Claudia Bachofen[1,*], Kim Willoughby[1], Ruth Zadoks[2], George Russell[1]
[1] Vaccines & Diagnostics and [2]Disease Control, Moredun Research Institute, Edinburgh, UK
* Current address: Institute of Virology, University of Zürich, Switzerland
E-mail: George.Russell@moredun.ac.uk

Introduction

Bovine viral diarrhoea virus (BVDV), a pestivirus of the flaviviridae family, is an important cattle pathogen causing ill-thrift and a lethal form of diarrhoea called mucosal disease. The amazingly broad host range of the virus (all even-toed ungulate species may be infected) may hinder the long-term success of BVDV eradication. **The quasispecies nature of this RNA virus may facilitate interspecies transmission by providing genetic plasticity.** Therefore, analysis of the spectrum of virus variants present in BVDV persistently infected animals may reveal important information on the quasispecies dynamics involved in interspecies transmissions.

AIM: To generate viral cDNA for next generation sequencing without using specific primers or amplification, providing unbiased sequence data for analysis of BVDV quasispecies diversity in serum and after limited culture

Method:
Blood sample MRI103
Persistently infected cow

Concentrate viral particles & prepare RNA directly from 50ml serum or 350 ml culture fluid
• Degrade cellular RNA/DNA (nuclease mix)
• Extract viral RNA (Trizol/viral RNA kit)
• A-tail viral RNA (polyA polymerase)
• Generate ds cDNA (oligo-dT-primed 1st strand)
(traditional 2nd strand)

• Next-generation sequencing by TGAC
• TrueSeq ChIP Library preparation
• 250 base paired-end sequencing on Illumina MiSeq
• data export as fastq

Bioinformatics:
Produce MRI103 consensus sequence by low stringency assembly onto BVDV type 1a full length genome sequence (Bowtie)
Quality trimming and contaminant removal from serum/culture read data
Reference-led assembly of reads to MRI103 consensus (DNASTAR SeqMan Ngen)
Visualise and assign single nucleotide polymorphisms (SNPs) DNASTAR SeqMan
Export SNP data for analysis (SNPs included are in >2% of reads; have >5 reads with SNP base; and >75 read depth)

BVDV sequence assembly summary:

| Source	serum	culture
Assembled reads	11492	115739
Mean read length	234	214
Mean read quality	36	36
Read pairs	5281	39242
Consistent pairs	5263	39190
Contig length	12310	12384
Mean coverage	220	2024

RESULTS:
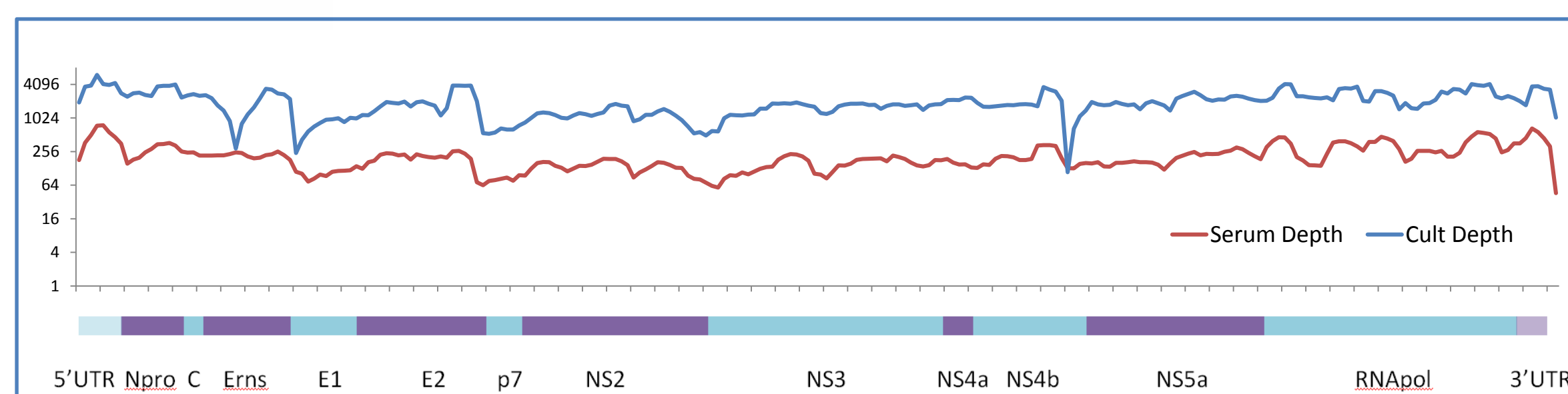
Fig 1. Read depth plotted against sequence position for serum- and culture-derived virus sequences. Depth (number of reads) is plotted on a logarithmic scale to allow both curves to be clearly viewed.
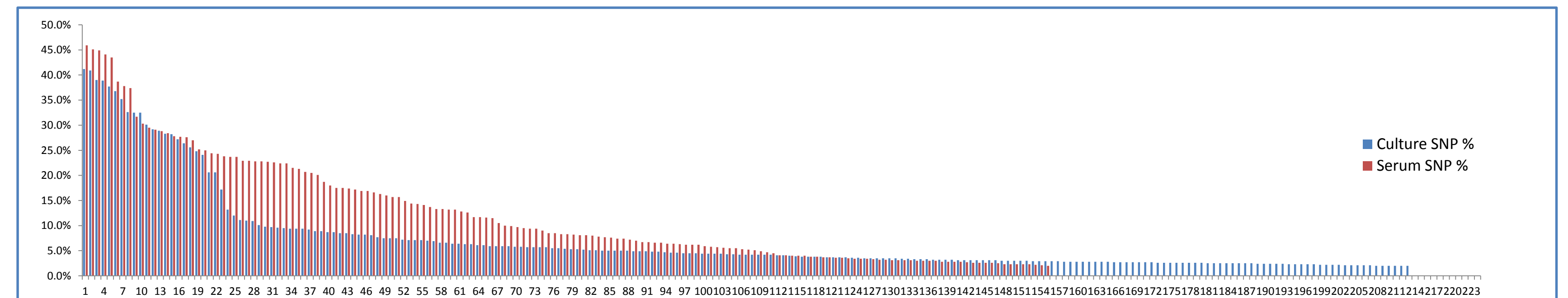
Fig 2. Comparison of SNP frequency in serum versus cultured virus. SNP frequency, measured as % of total read number with variant base call, was plotted in order of percentage for all SNP that met the analysis criteria. A larger number of valid SNPs were found in the cultured virus dataset (blue; probably a consequence of greater read depth), while comparable SNPs generally had higher % polymorphism in the serum dataset (red).
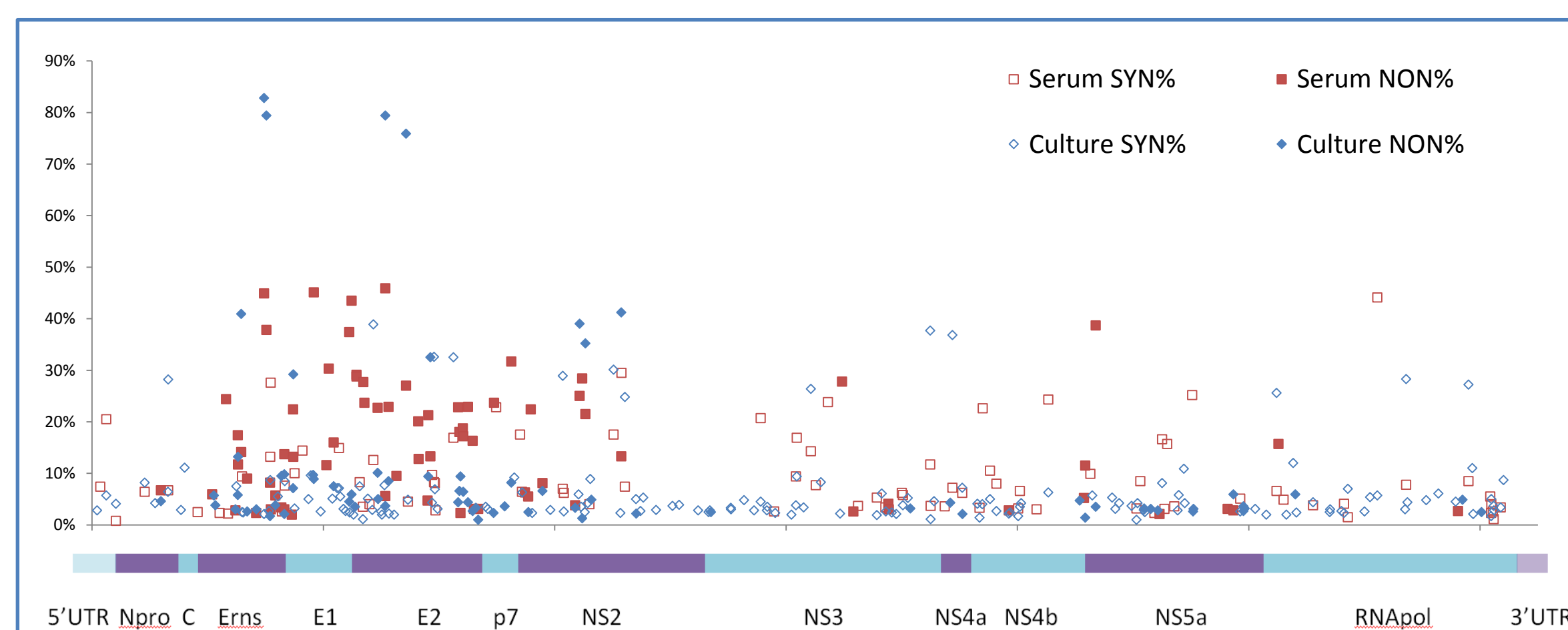
Fig 3. SNP frequency plotted against BVDV genome position in serum (red) and cultured (blue) viral genomic RNA. SNP frequencies at each position were calculated with respect to the serum consensus sequence. Open symbols indicate synonymous substitutions (SYN), while closed symbols indicate non-synonymous (NON) substitutions. Symbols at >50% frequency indicate a change of consensus sequence in the cultured virus
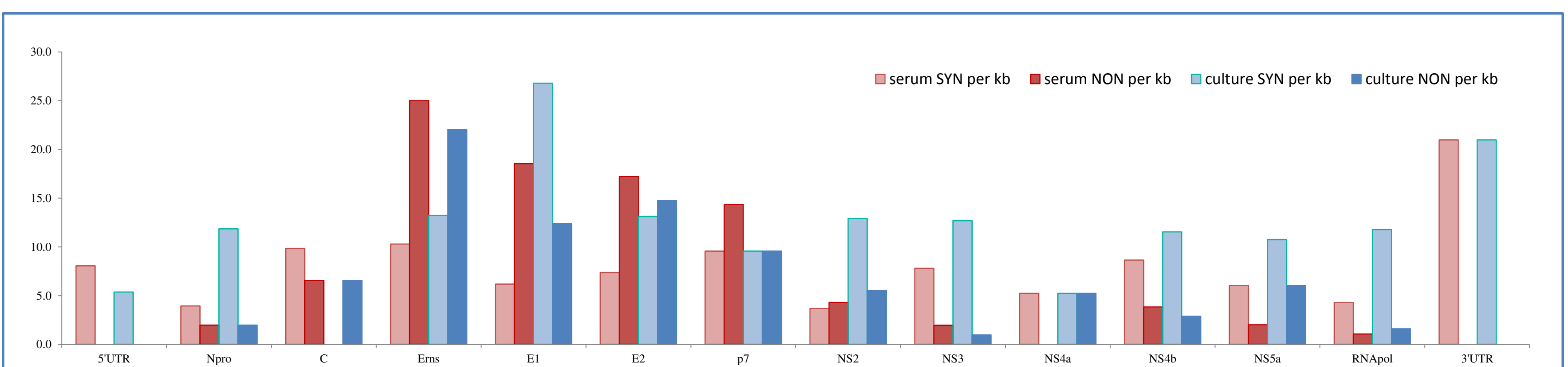
Fig 4. SNP frequency in each BVD gene. For each source of virus, the number of synonymous (SYN) or non-synonymous (NON) SNPs were normalised to the gene length in kb. This demonstrates that the virus envelope protein (E) coding genes have a generally higher SNP frequency than the non-structural (NS) genes

N.B. SNPs in the 5'UTR and 3'UTR were all classified as synonymous because these regions do not encode viral proteins. However, sequence variation may be constrained in these regions for functional reasons such as maintaining essential secondary structure in the viral genome.

CONCLUSIONS:

• Quasispecies diversity and consensus sequence were both influenced by virus propagation in culture
• Average SNP frequency in the serum virus sequence (12.6%) was almost double that found after 3 passages in culture (7.4%)
• The highest SNP density was observed in the envelope protein coding regions. After passage, SNP-switches in several positions lead to changes in the consensus sequence.
• Non-synonymous SNP's (causing amino acid changes) were mainly found in the envelope and capsid protein coding region but rarely in regions coding for non-structural proteins.

SUMMARY:

• **We developed an amplification-free method for unbiased quasispecies analysis**
• **BVDV quasispecies diversity was reduced rapidly by cell culture passage**
• **Highest diversity was found in envelope protein regions**
• **Lower diversity/synonymous SNPs in non-structural protein coding regions**
• **Possible selection for preservation of essential (non-structural) protein functions**
• **Studies of quasispecies diversity should avoid virus propagation in culture**