Introduction to primer design concepts and tools

"Arboviral Genomic Surveillance in a Climate evolving world" Next Generation Sequencing (NGS) workshop

Duke-NUS Medical School

Singapore

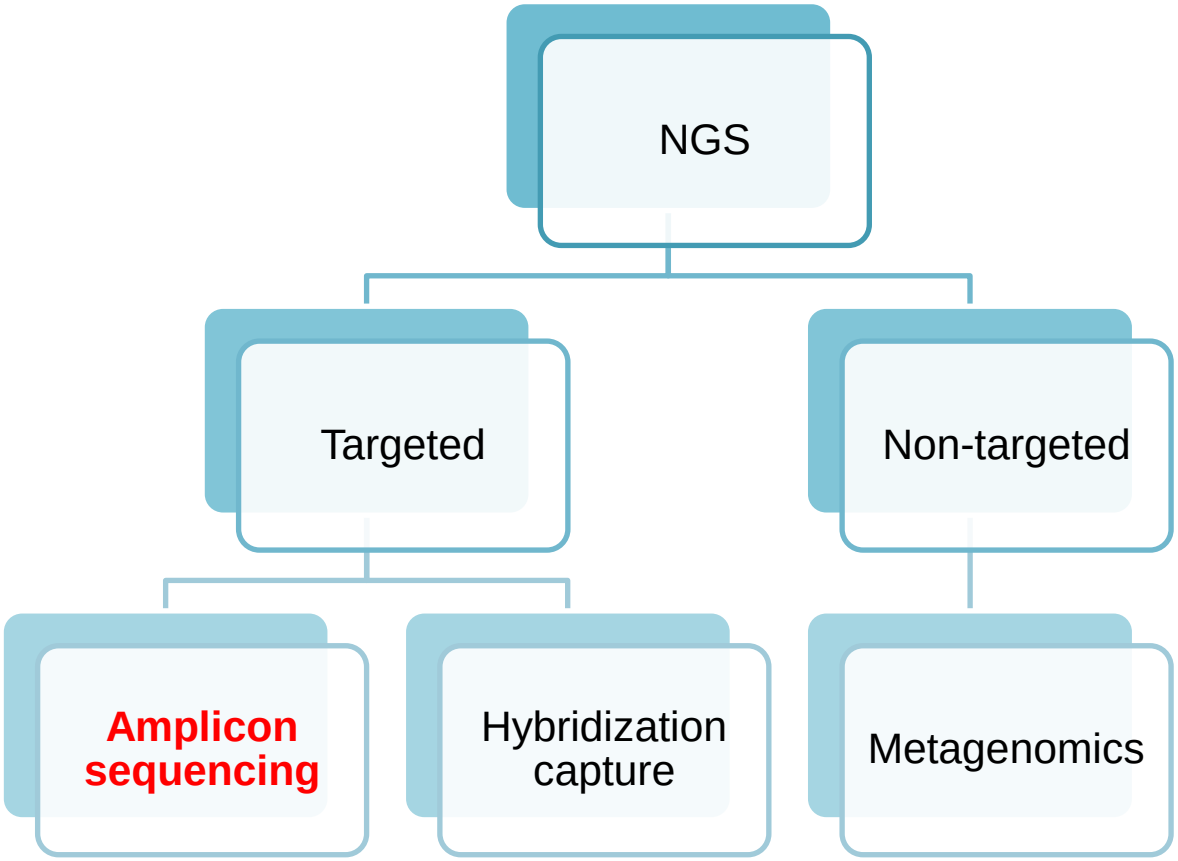
Presented by: Dr. Chanditha Hapuarachchi

03 Jul 2024

National Environment Agency

Safeguard · Nurture · Cherish

Next Generation Sequencing for genomic surveillance



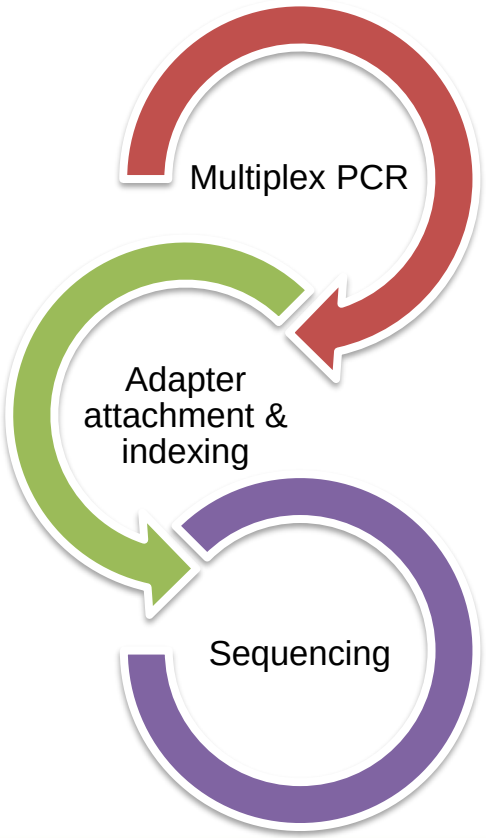
Target enrichment	▪ PCR ▪ Isolation	Long oligo probes	Background depletion

Amplicon sequencing

Amplicon sequencing is a type of targeted next generation sequencing that uses multiplex PCR to create DNA sequences called amplicons

Advantages

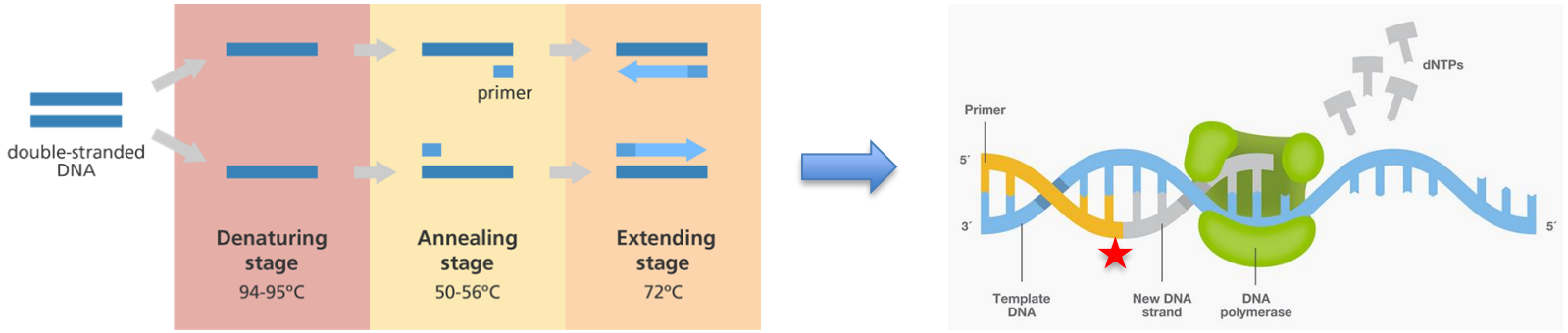
- Fast turnaround time
- Low sequencing costs
- Low DNA input requirements – up to 50-100 genome copies per reaction
- Supports multiplexing of multiple samples in a single sequencing reaction
- Amenable to a wide range of experimental designs, such as complete genome sequencing, allelic variant identification, marker gene sequencing for microbiome studies etc.



A successful PCR relies on well-designed primers ……..

What makes a good primer?

High priming efficiency	High specificity
No homo/hetero dimers, **esp. at 3' end** (prevent self-extension)	Long enough and reasonably high melting temperature (T_m)
No hairpin stems **esp. at 3' end** (prevent self-priming)	Uniqueness of the sequence
No repetitive sequences & mismatches **esp. at 3' end**	Moderately stable **3' end**



Characteristics of a “good” primer

Requirement	Feature
High priming efficiency	Approximate T_m values (±5^{0}C between F & R)
	Equal proportions of A/Ts & G/Cs (50% ±5-10%)
	Includes G/Cs at the 3’ end
	Lack of long repeats & runs– no more than 4 repeats & 3 bp runs
	Lack of duplexes and structures
High specificity	Length 18-22 bp
	Unique sequence esp. at 3’ end but with moderate stability (ideally no more than 2 Gs or Cs in last 5 bases)

Other considerations

- Avoid regions with high secondary structures (hairpins) & GC islands
- Try to have the lowest level of degeneracy, if degenerate bases are needed
- **For difficult to amplify regions**
  - Select primers with high T_m and GC content (than usual)
  - Add solvents/additives such as DMSO
  - Increase primer concentration, but not more than 1 µmol (to avoid nonspecific primer-primer amplifications)
- **Designing probes**
  - Do not put Gs at the 5´ end
  - Keep the size of PCR product or amplicon to 50 - 150 bp. Shorter amplicons work most efficiently
  - Probe should ideally be close to one of the primers without overlapping

Post-design analysis

OligoAnalyzer

Instructions | Definitions | Feedback

Sequence: 5' MOD | INTERNAL | 3' MOD | MIXED BASES

TACAACATGATGGGGAARAGAGARAA

Bases 26 — CLEAR SEQUENCE

Try the new batch mode here

Parameter sets: SpecSheet (Default)

Target type	DNA	
Oligo Conc	0.25	µM
Na^+ Conc	50	mM
Mg^{++} Conc	1.50	mM
dNTPs Conc	200	mM

ANALYZE | HAIRPIN | SELF-DIMER | HETERO-DIMER | NCBI BLAST | TM MISMATCH | ADD TO ORDER

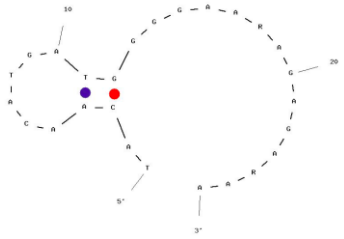
Results

RESUSPENSION | DILUTION

SEQUENCE	5'- TAC AAC ATG ATG GGG AAR AGA GAR AA -3'
COMPLEMENT	5'- TTY TCT CTY TTC CCC ATC ATG TTG TA -3'
LENGTH	26
GC CONTENT	38.5 %
MELT TEMP RANGE	MIN 63 °C; MEAN 63.6 °C; MAX 64.7 °C

Structures

structure	Image	ΔG (kcal.mole^{-1})	T_m (°C)	ΔH (kcal.mole^{-1})	ΔS (cal.K^{-1}mole^{-1})	Output
1		-0.04	25.5	-21.9	-73.33	Ct Det
2		0.02	24.6	-14.7	-49.38	Ct Det
3		0.41	17.7	-16.5	-56.73	Ct Det



Homo-Dimer Analysis

The delta G is calculated by taking into account the longest stretch of complementary bases. These pairs of complementary bases are represented by a solid line. Dotted lines represent additional complementary bases for that dimer structure, but their presence does not impact calculated delta G values. Actual delta G values may vary based on presence of additional complementary bases. The Maximum Delta G value refers to the free energy of the oligo sequence binding to its perfect complement.

Dimer Sequence:
5'- TACAACATGATGGGGAARAGAGARAA -3'
Maximum Delta G: -46.01 kcal/mole

Delta G: -5.38 kcal/mole **Base Pairs:** 4

```
5'              TACAACATGATGGGGAARAGAGARAA
                  :: |||| ::
3' AARAGAGARAAGGGGTAGTACAACAT
```

Hetero-Dimer Analysis

The delta G is calculated by taking into account the longest stretch of complementary bases. These pairs of complementary bases are represented by a solid line. Dotted lines represent additional complementary bases for that dimer structure, but their presence does not impact calculated delta G values. Actual delta G values may vary based on presence of additional complementary bases. The Maximum Delta G value refers to the free energy of the oligo sequence binding to its perfect complement.

Primary Sequence: 5'- TACAACATGATGGGGAARAGAGARAA -3'
Secondary Sequence: 5'- GTGTCCCAGCCGCGTGTCATCGC -3'

Maximum Delta G: -50.44 kcal/mole

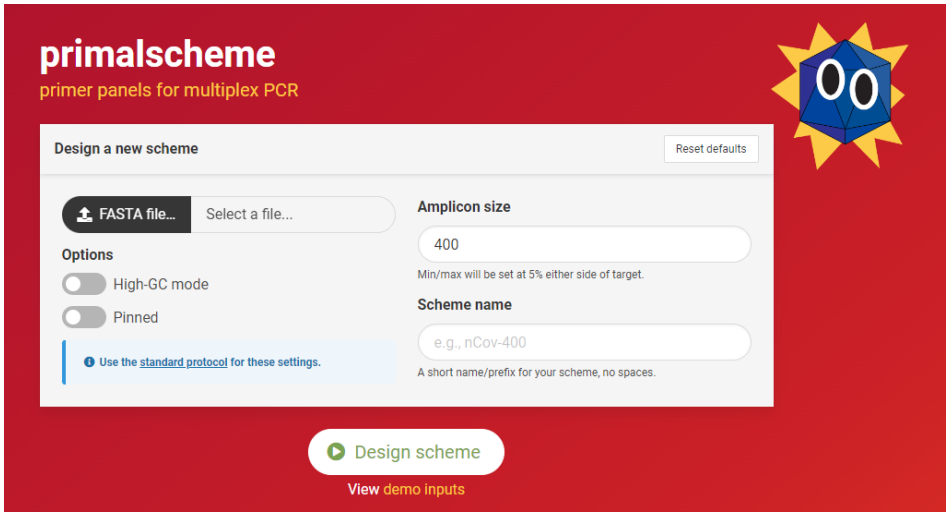
Delta G: -8.09 kcal/mole **Base Pairs:** 4

```
5'       TACAACATGATGGGGAARAGAGARAA
           :::    : ||||  :
3' CGCTACTGTGCGCCGACCCTGTG
```

[OligoAnalyzer Tool - Primer analysis and Tm Calculator / IDT (idtdna.com)]

Primer designing for amplicon sequencing

- 'Primal Scheme' primer design tool (https://primalscheme.com/)
- Multiplex primer sets that are divided into two pools (pool 1 & 2)
- User-defined fragment length (default 400 bp; 200-2000 bp)
- Overlapping fragments – tiling along the genome with overlapping regions (default 75 bp)
- High annealing temperature that is high enough to perform 2-step PCR (65-68°C)



Quick J et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017 Jun;12(6):1261-1276.

Example panels

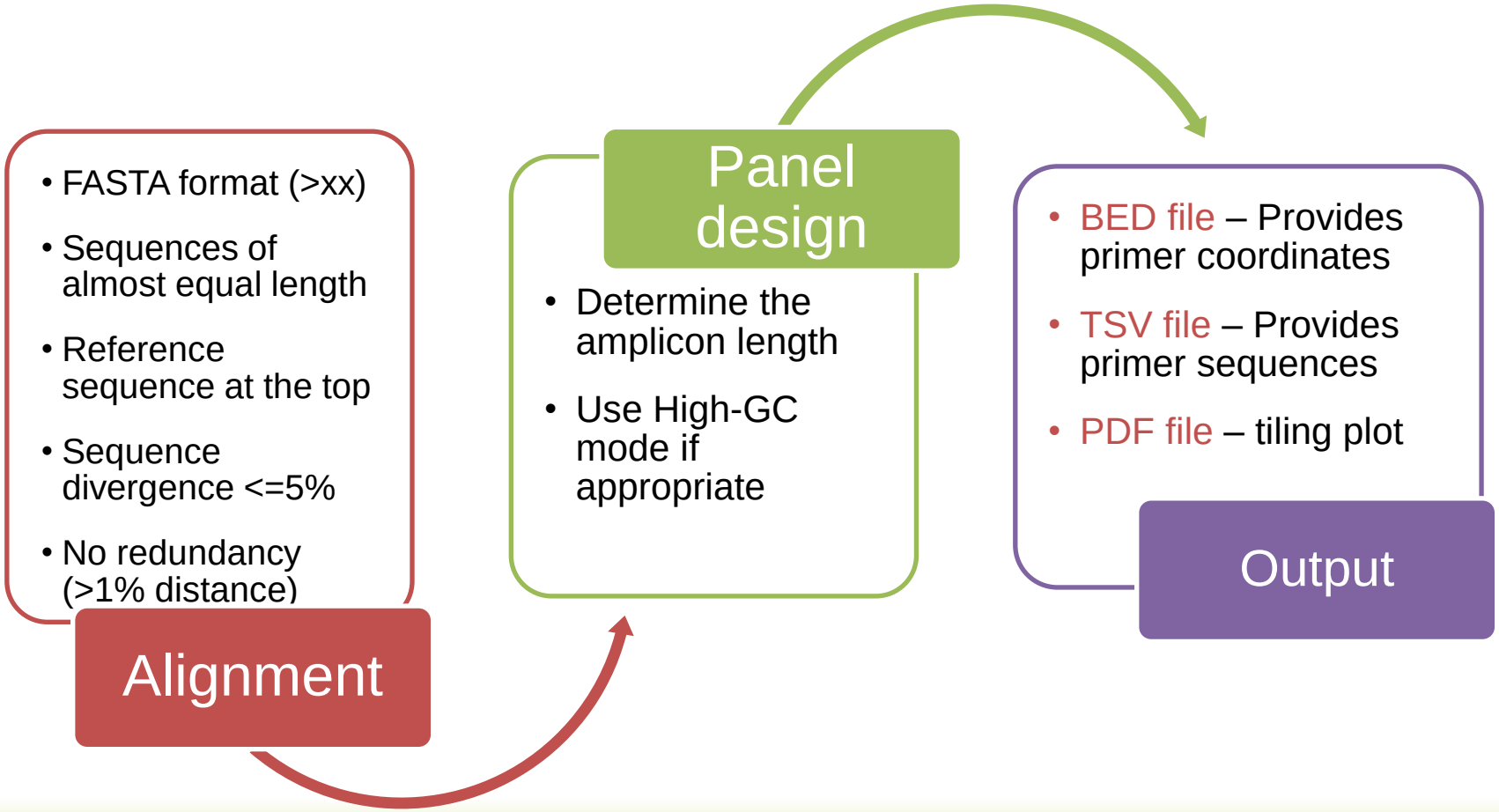
1. DENV (Nathan Grubaugh)
   https://www.protocols.io/view/dengueseq-a-pan-serotype-whole-genome-amplicon-seq-kqdg39xxeg25/v2?version_warning=no

2. ZIKV & CHIKV

   Quick J et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017 Jun;12(6):1261-1276.

Primalscheme: Primer designing workflow

Design requirement - research question, target region etc



Primalscheme: Primer designing output files

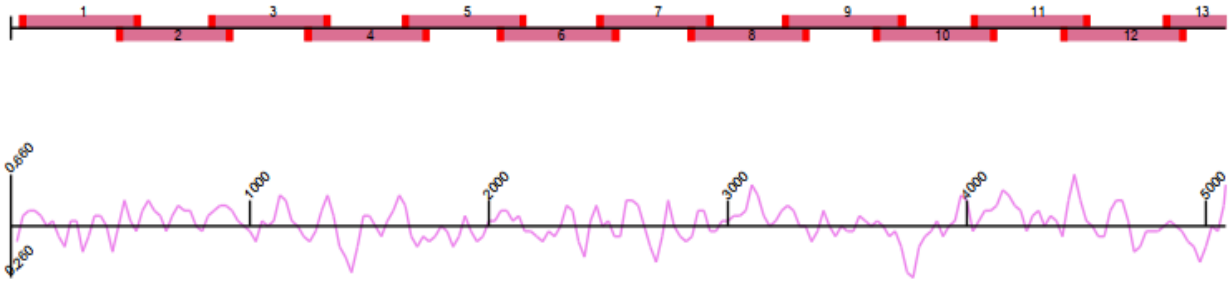
BED file – Provides primer coordinates

```
AY618991(country        39      61      DENV-4_Revised_1_LEFT   1       +
AY618991(country        518     541     DENV-4_Revised_1_RIGHT  1       -
AY618991(country        444     466     DENV-4_Revised_2_LEFT   2       +
AY618991(country        904     926     DENV-4_Revised_2_RIGHT  2       -
AY618991(country        831     852     DENV-4_Revised_3_LEFT   1       +
AY618991(country        1312    1334    DENV-4_Revised_3_RIGHT  1       -
```

TSV file – Provides primer sequences

name	pool	seq	size	%gc	tm (use 65)
DENV-4_Revised_1_LEFT	1	ATGCTGAAACGCGAGAGAAACC	22	50	61.93
DENV-4_Revised_1_RIGHT	1	ACCAGCAATCAATGTCTTCAGGT	23	43.48	60.76
DENV-4_Revised_2_LEFT	2	CTTATTGCCATGGACCTGGGTG	22	54.55	61.46
DENV-4_Revised_2_RIGHT	2	CATCCTCCATGTTCTAGCACCA	22	50	60.01
DENV-4_Revised_3_LEFT	1	TACGGAATGCGATGCGTAGGA	21	52.38	61.79
DENV-4_Revised_3_RIGHT	1	GGGGTTATCGTAGCTGTCACTC	22	54.55	60.4

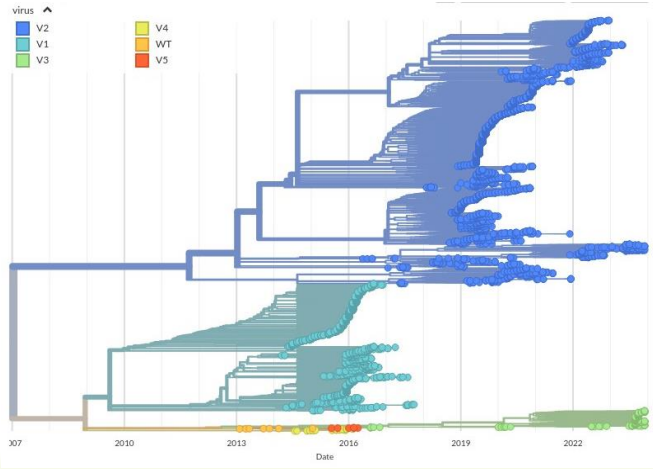
PDF file – tiling plot



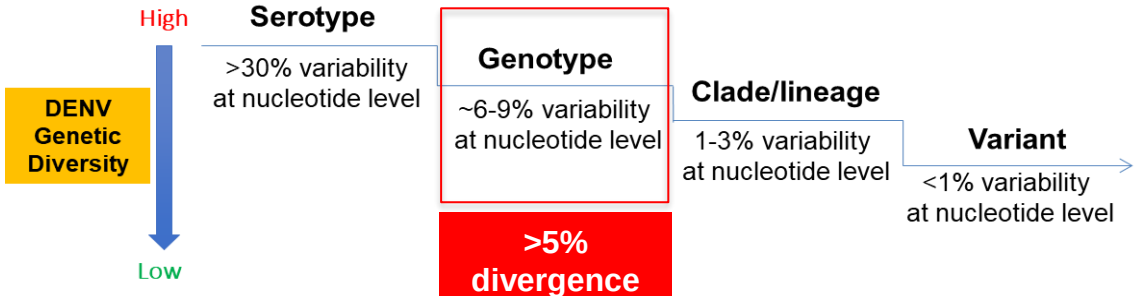
Amplicon sequencing: Primer designing considerations

- Selection of sequences: to represent the true diversity of the pathogen/genomic region of interest
- No. of panels: based on the level of genetic divergence
- Quality of nucleic acids: Short amplicons for degraded samples

High variant diversity within a single clade of DENV-2

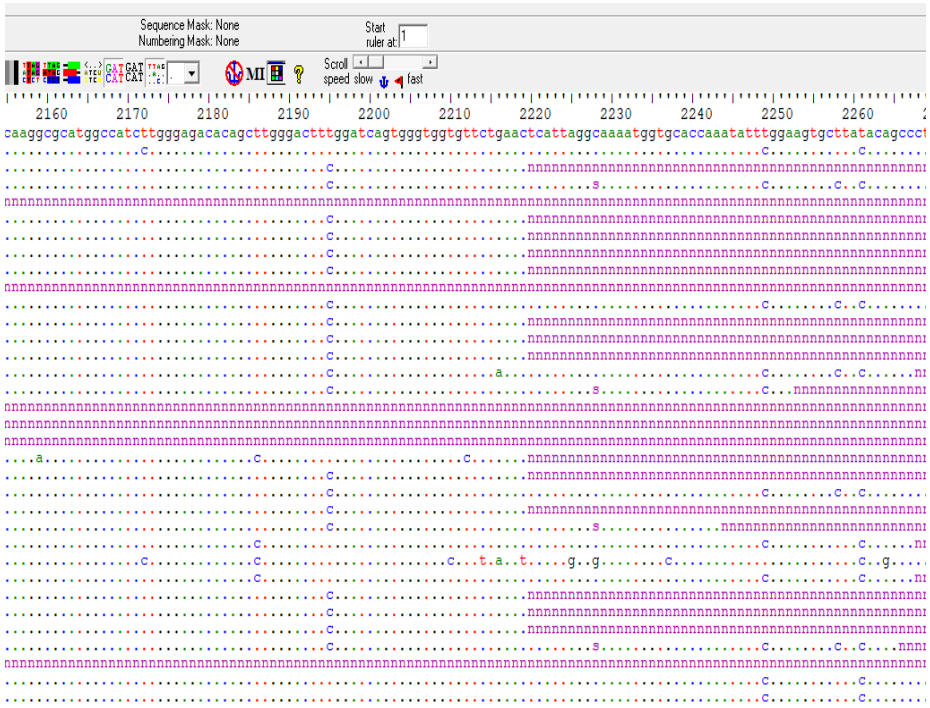


Dengue virus diversity cascade: Ideally needs panels to represent each genotype



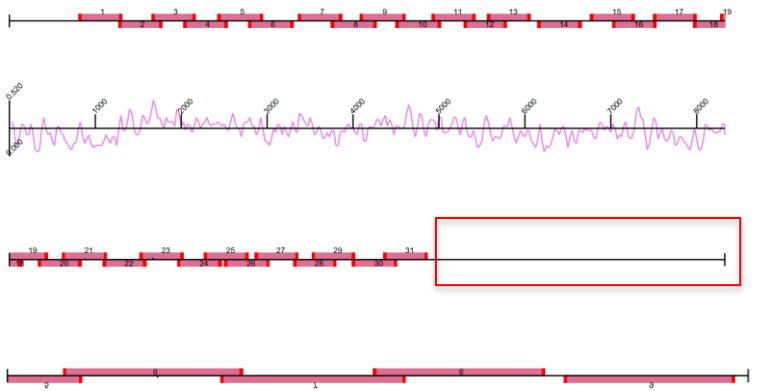
Primer panels do not always generate ideal results!!

Drop-out regions

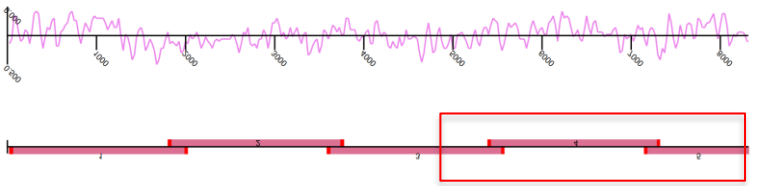


Design limitations

500 bp design



2000 bp design



Amplicon sequencing: Primer designing troubleshooting

Improving the primers in drop-out regions

- Map the primers to determine the likely reason of their failure
  - Modify the primer if possible, while maintaining the same Tm
  - Choose a replacement primer from another panel
  - Redesign the panel using a different reference dataset
- If fails, fill-in gaps using Sanger sequencing

Our Environment

Safeguard • Nurture • Cherish

National Environment Agency

Safeguard · Nurture · Cherish