



Identification of Defective Viral Genomes using NGS Data on HIVE

Konstantinos Karagiannis

Konstantinos.Karagiannis@fda.hhs.gov

FDA/CBER/OBE/HIVE

02/22/2021





Disclaimer

This presentation reflects the views of the author and should not be construed to represent FDA's views or policies



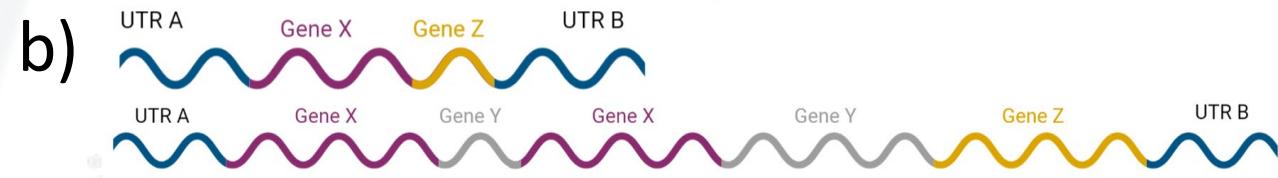
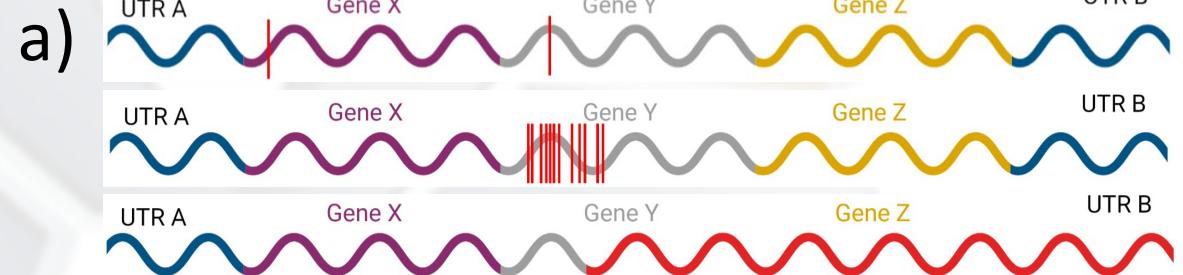
What is a DVG?

DVG

Defective
Viral
Genomes

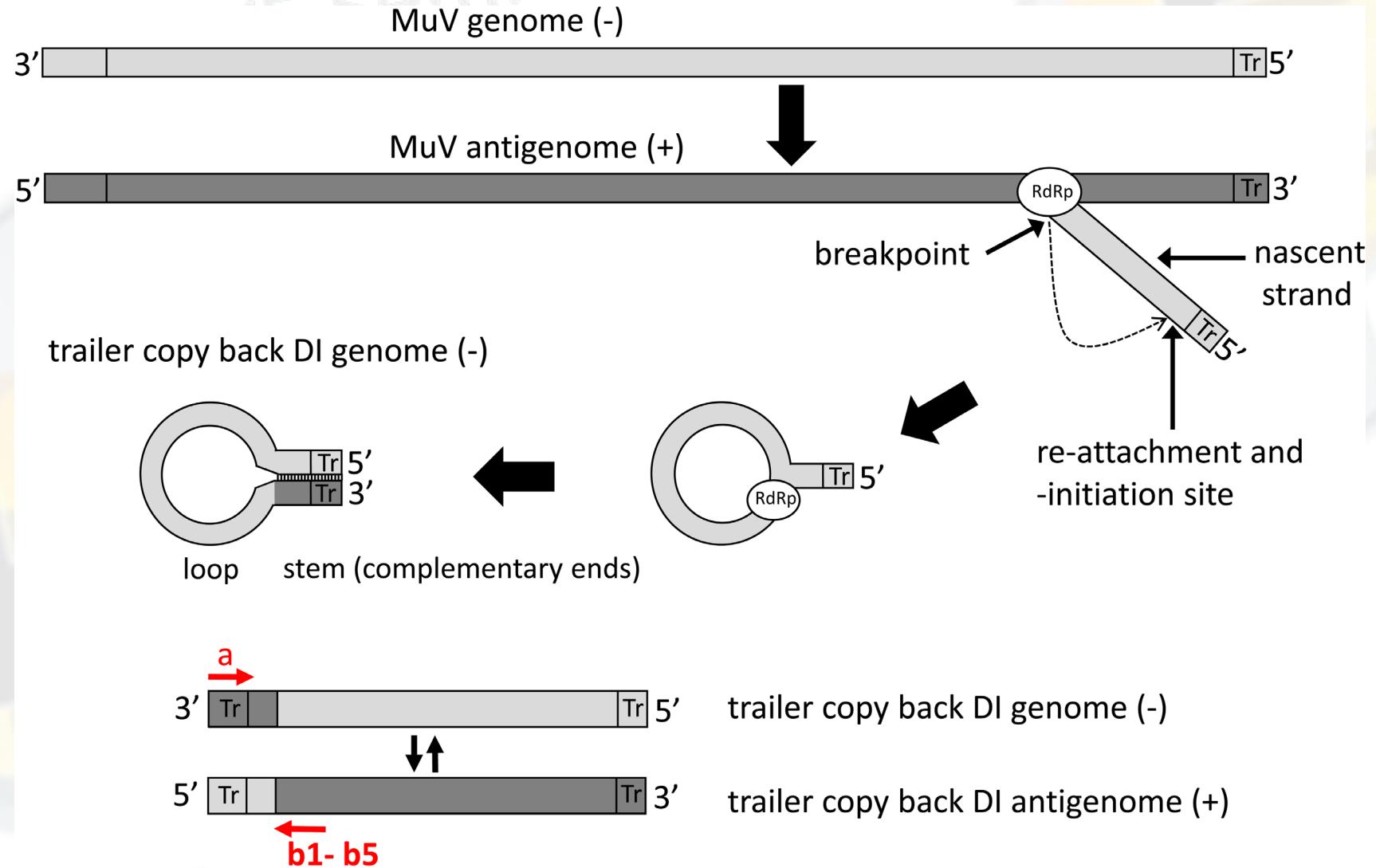
Genomes
unable to
complete a full
replication
cycle

- a) Mutation
- b) Deletions
- c) Copy-back/snap-back





Copy-back DVG



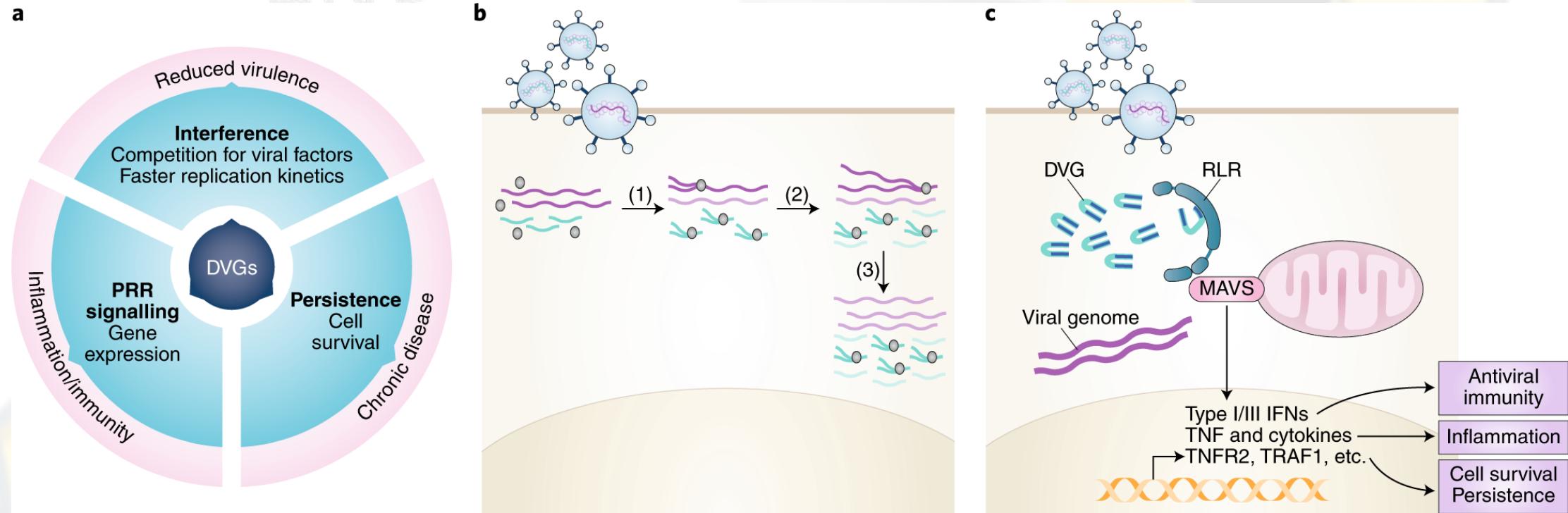


Generation mechanisms

- Stochastic mistakes
Error prone viral RNA polymerase
- Genomic variations (single AA mutation in SeV's nucleoprotein)
- Variants of structural proteins (PPXY domain of LCMV)
- Template switching (intra or inter-genomic)



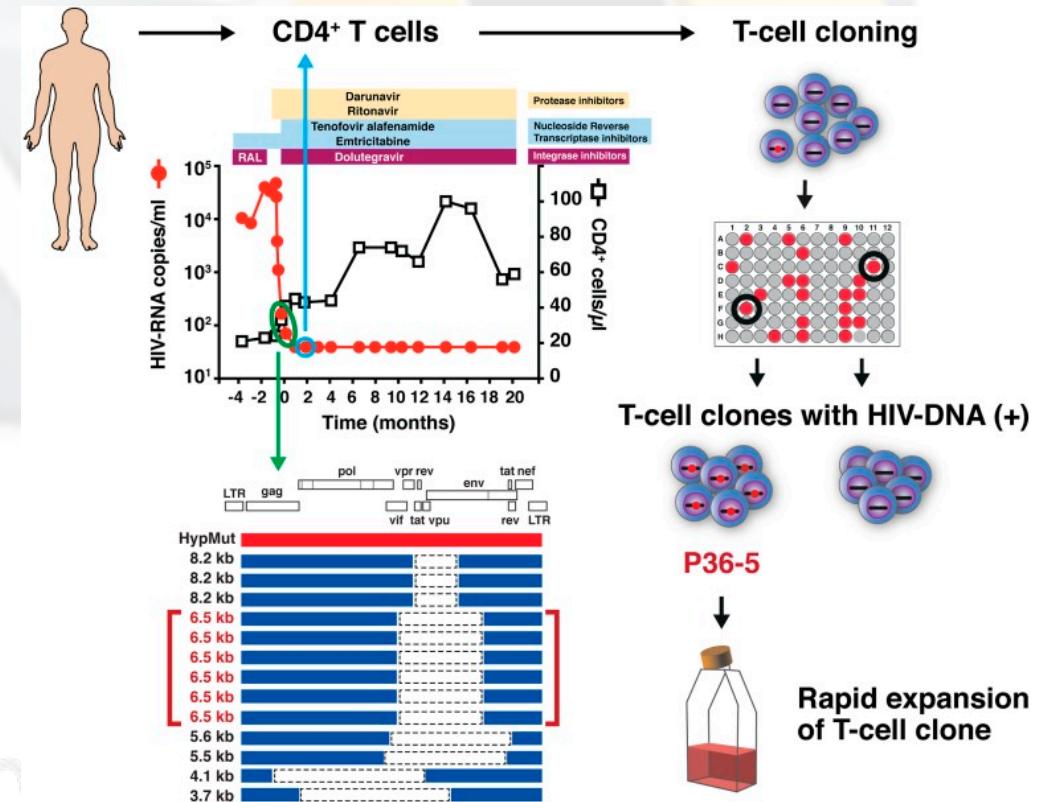
Effects of DVGs





DVG and HIV

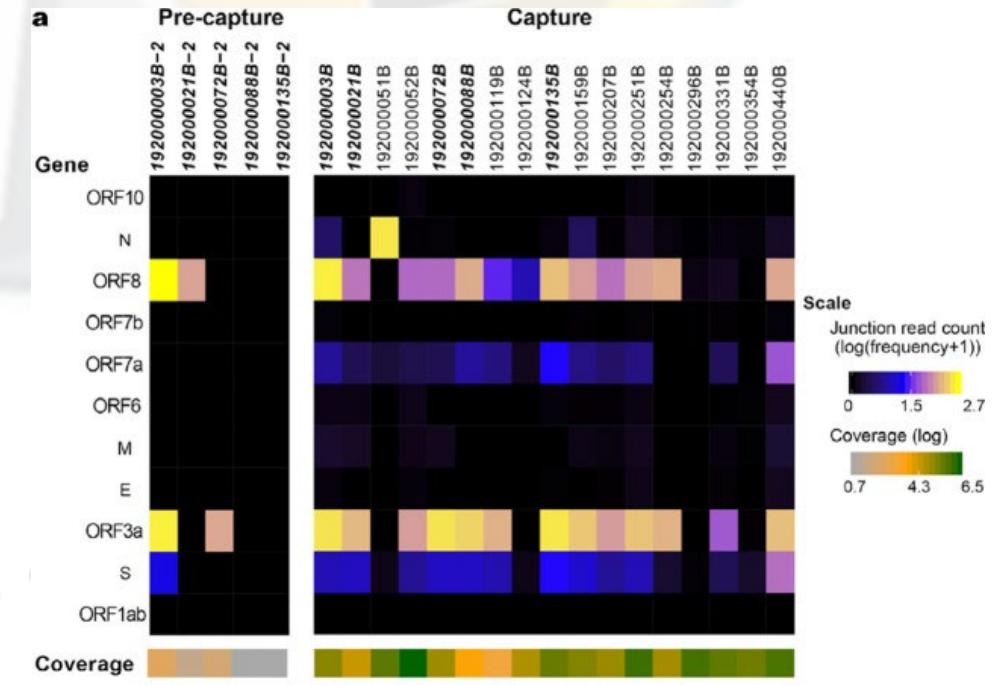
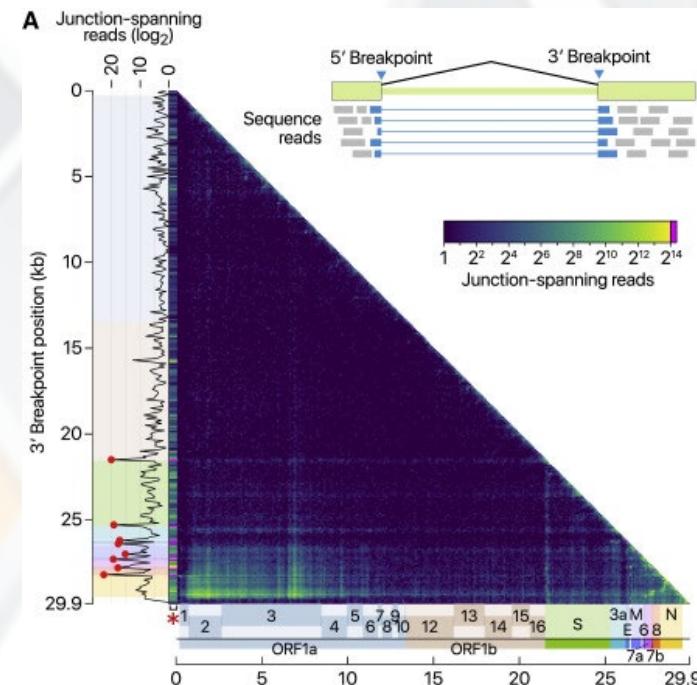
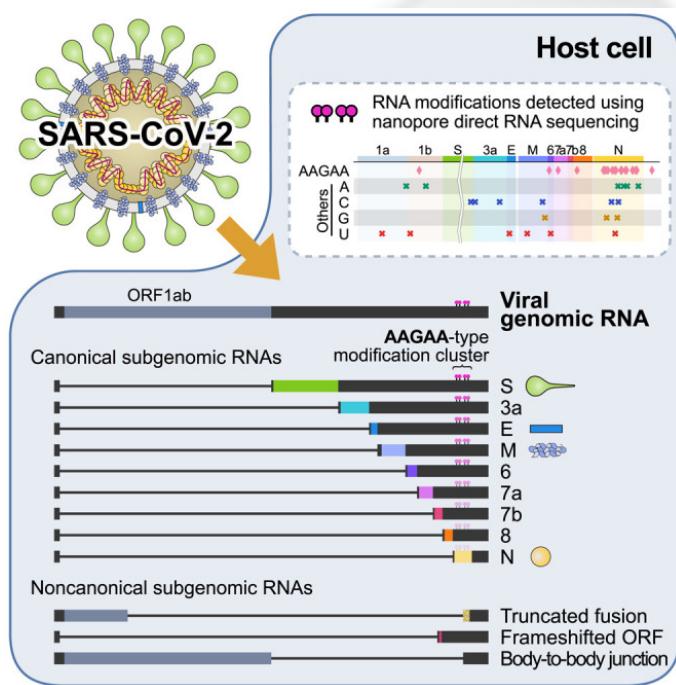
- Types of DVG
 - Mutations
 - Deletions
- Present at all stages of HIV-1 infection
- Replication-incompetent or competent
- Silent pool
 - Persistent seropositivity
 - In CD4⁺ memory cells 17y





DVG in SARS-CoV-2

- Defective RNAs / subgenomic RNA



D. Kim *et al.* The Architecture of SARS-CoV-2 Transcriptome. 2020 May 14, Cell

H. Doddapaneni *et al.* Oligonucleotide Capture Sequencing of the SARS-CoV-2 Genome and Subgenomic Fragments from COVID-19 Individuals. 2020 Dec 11, bioRxiv



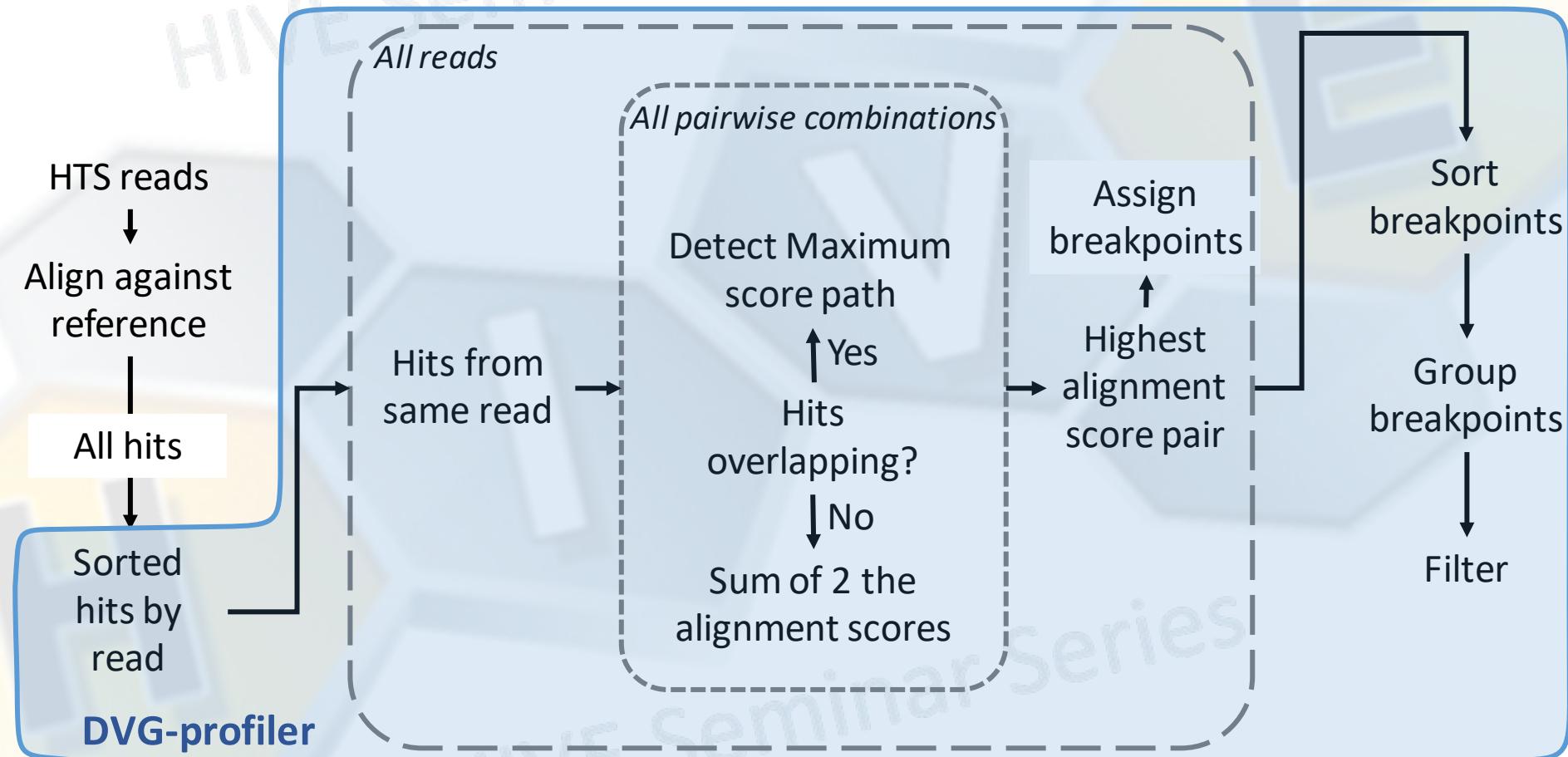
RNA viruses with described DVGs

Virus	Type of DVG	Known DVG functions
nsRNA viruses		
Arenaviridae		
Lymphocytic choriomeningitis virus	Deletion	Interference/modulate virulence/persistence
mammarenavirus		
Filoviridae		
Ebola virus	Deletion /copy-back	Persistence
Orthomyxoviridae		
Influenza virus	Deletion	Interference/IFN-induction/persistence/modulate virulence
Paramyxoviridae		
Human parainfluenza virus 3	Deletion /copy-back	Interference
Parainfluenza virus 5	Deletion/copy-back	IFN-induction
Measles virus	Deletion/copy-back	Interference/IFN-induction/persistence/modulate virulence
Mumps virus		Persistence/modulate virulence
Sendai virus	Deletion/copy-back	Interference/IFN-induction/immune stimulation/persistence/modulate virulence
Peribunyaviridae		
Bunyamwera virus	Deletion	Interference
Phenuiviridae		
Rift Valley Fever virus (RVFV)	Deletion	Interference/modulate virulence
Toscana virus	Deletion	Interference
Pneumoviridae		
Human metapneumovirus	Copy-back	IFN-induction
Human respiratory syncytial virus	Deletion/copy-back	Interference/IFN-induction/persistence
Rhabdoviridae		
Vesicular stomatitis virus	Deletion/copy-back	Interference/IFN-induction/persistence/modulate virulence
Rabies virus	Deletion	Interference/persistence
Tospoviridae		
Tomato spotted wilt virus	Deletion	Modulate virulence
dsRNA viruses		
Birnaviridae		
Infectious necrotic pancreatic virus	UN	Persistence
Partitiviridae		
Rosellinia necatrix virus	Deletion	Interference
Reoviridae		
Type 3 reovirus	Deletion	Interference
Wound tumor virus	Deletion	UN

Virus	Type of DVG	Known DVG functions
psRNA viruses		
Arteriviridae		
Porcine reproductive and respiratory syndrome virus	Deletion	UN
Equine arteritis virus	Deletion	Interference
Closteroviridae		
Citrus tristeza virus	Deletion	UN
Coronaviridae		
Berne virus	Deletion	Interference
Bovine coronavirus	Deletion	UN
Infectious bronchitis virus	Deletion	UN
Mouse hepatitis virus	Deletion	Interference/persistence
Transmissible gastroenteritis virus	Deletion	Interference
Flaviviridae		
Dengue virus	Deletion	Persistence
Japanese encephalitis virus		Persistence
Hepatitis C virus	Deletion	Persistence
Murray Valley encephalitis virus	Deletion	Persistence
Tick-borne encephalitis virus	Deletion	Modulate virulence
West Nile virus	Deletion	Interference/persistence
Nepoviridae		
Tomato black ring virus	Deletion	Interference
Nodaviridae		
Flock House virus	Deletion	Modulate virulence
Picornaviridae		
Encephalomyocarditis virus	Deletion	Interference
Foot-and-mouth disease virus	Deletion	Interference
Mengo virus	Deletion	Interference
Polio virus	Deletion	Modulate virulence
Togaviridae		
Rubella virus		Persistence
Semliki Forest virus	Deletion	Interference/modulate virulence
Sindbis virus	Deletion	Interference/IFN-induction
Tombusviridae		
Cucumber necrosis virus	Deletion	Modulate virulence
Tomato bushy stunt virus	Deletion	Interference/modulate virulence
Turnip crinkle virus	Deletion	Interference/modulate virulence
Retroviruses		
Retroviridae		
Human immunodeficiency virus 1	Deletion/ hypermutation/frame shift	Persistence

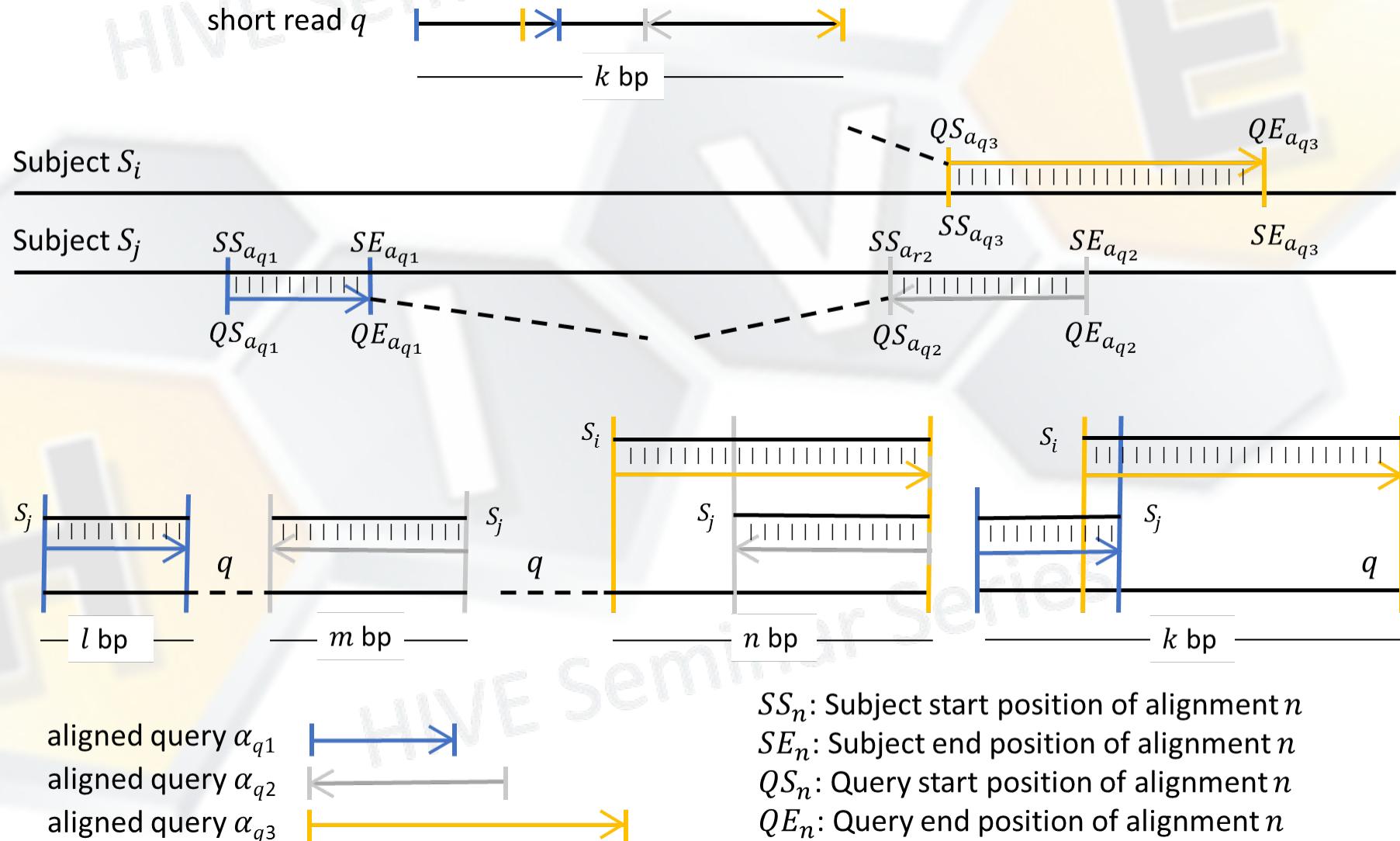


DVG-profiler



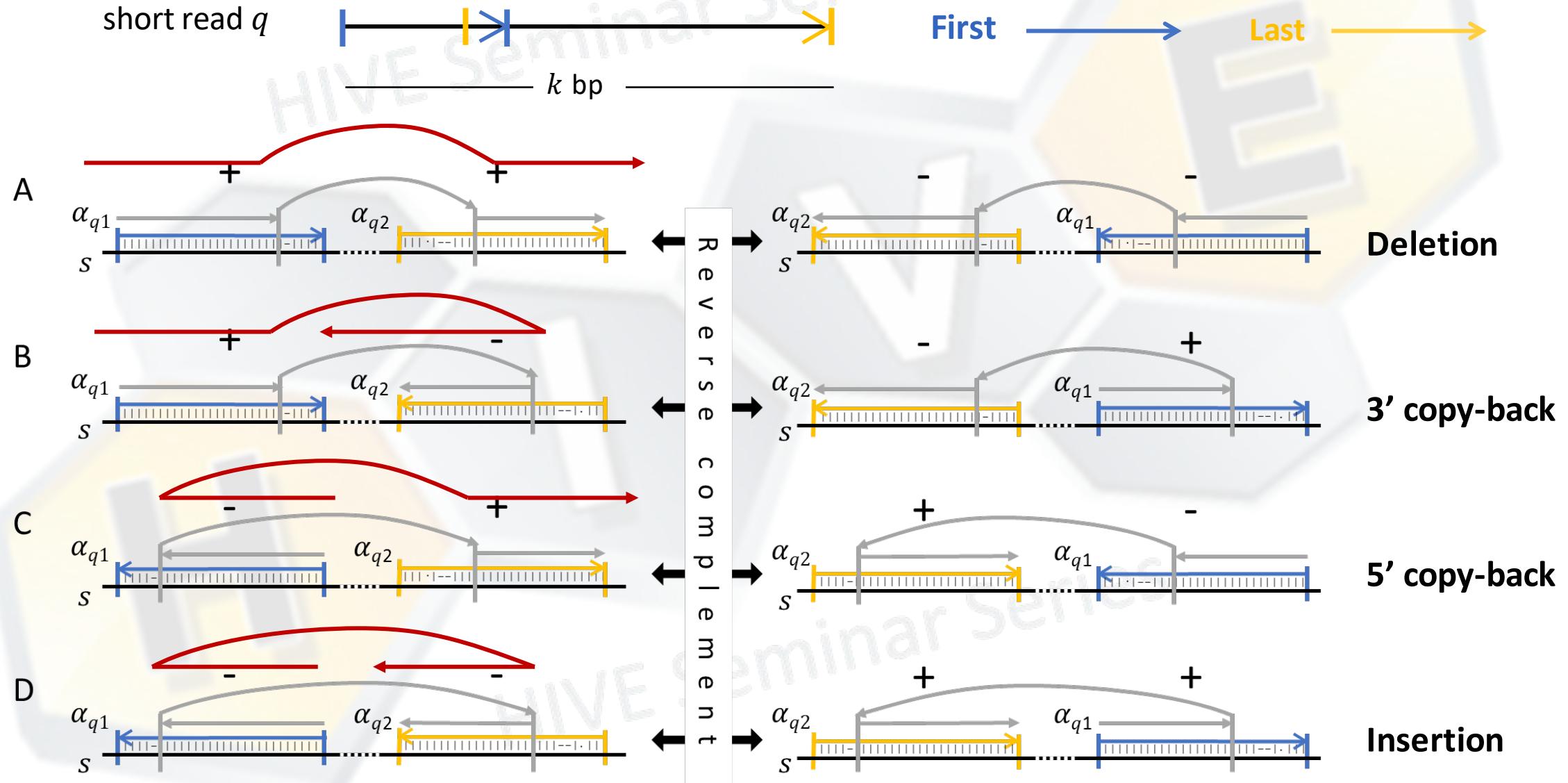


DVG-profiler: Annotations



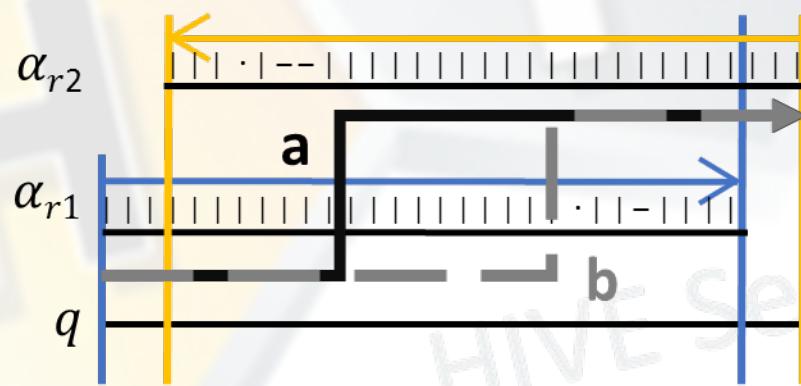
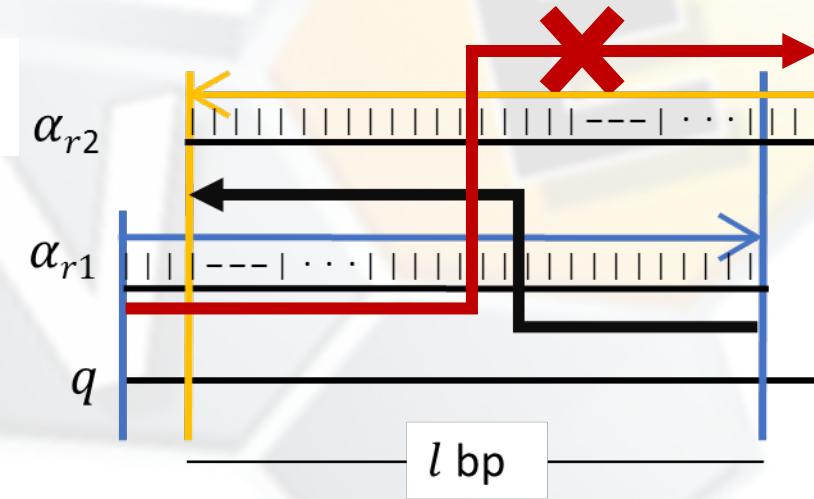
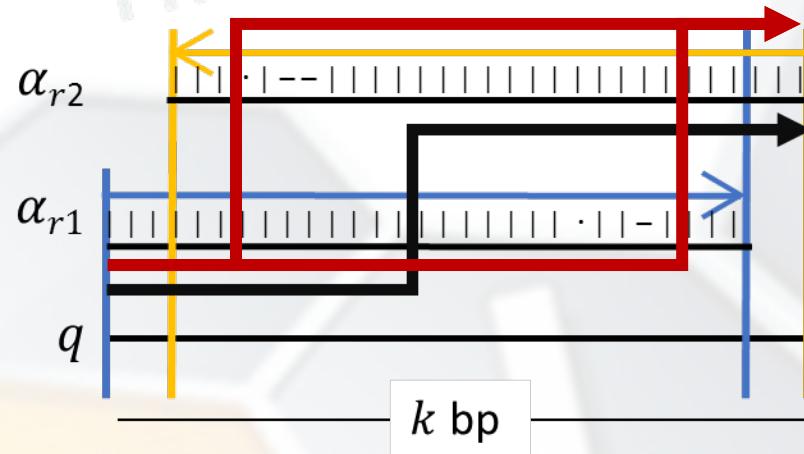


DVG profiler: Junction types





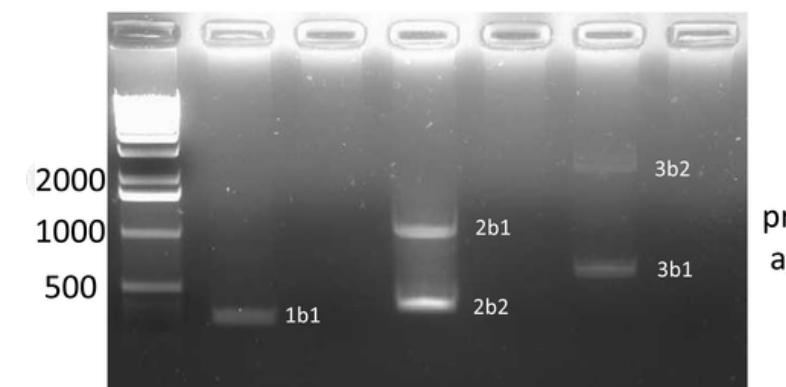
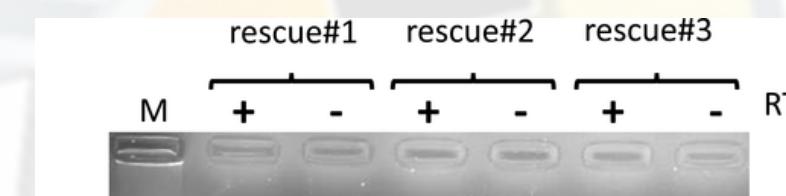
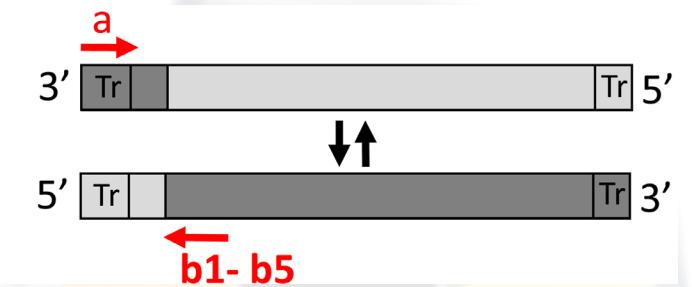
Stable junction calling



Stable junction detection



Experimental confirmation

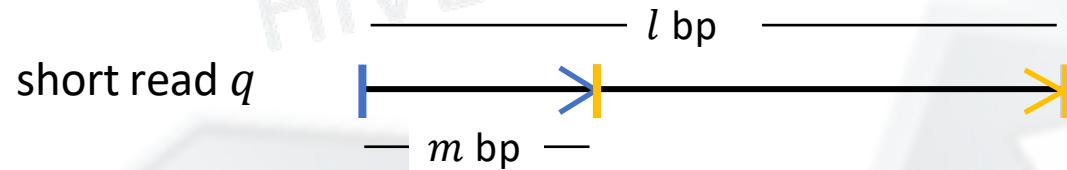


primer pair
a/b1

primer pair
a/b2

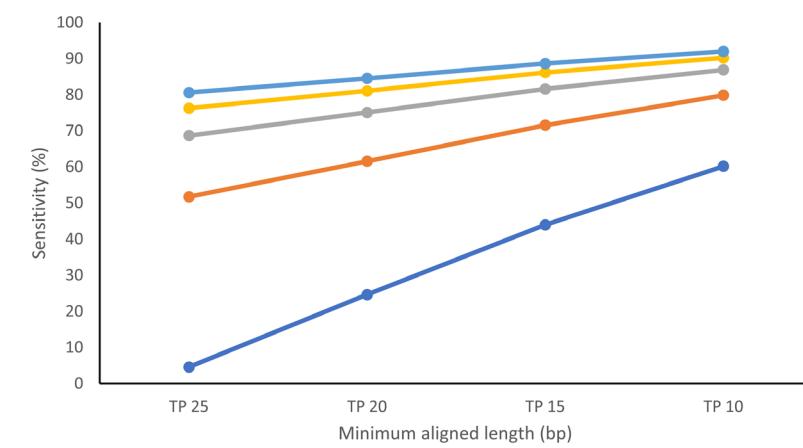


Performance assessment

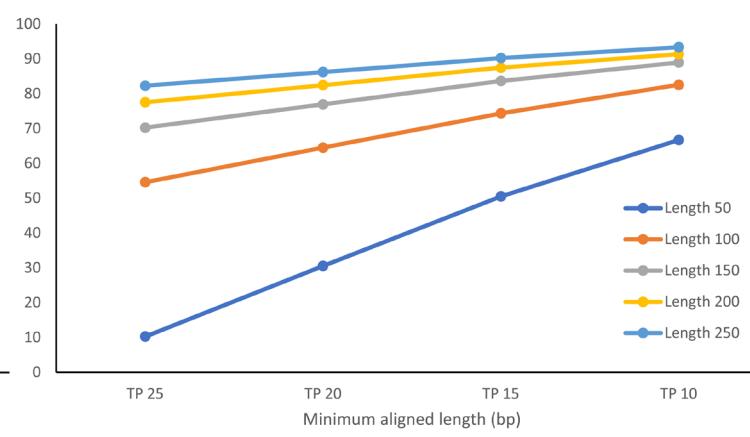


m = Minimum aligned length
 l = Read length

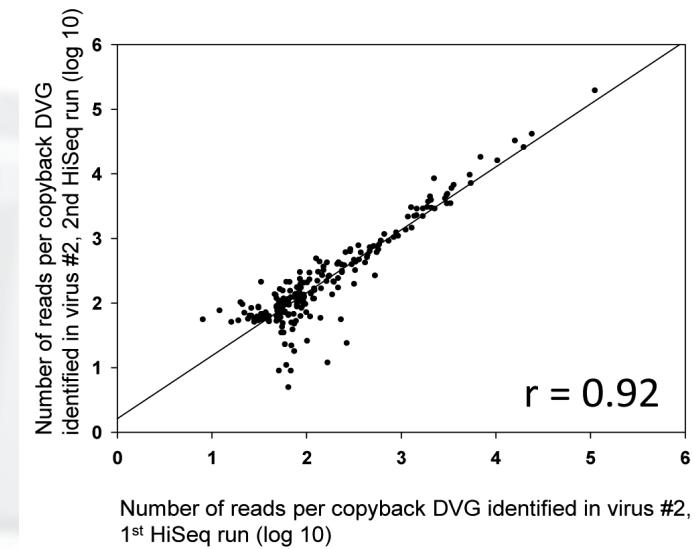
DVG S1



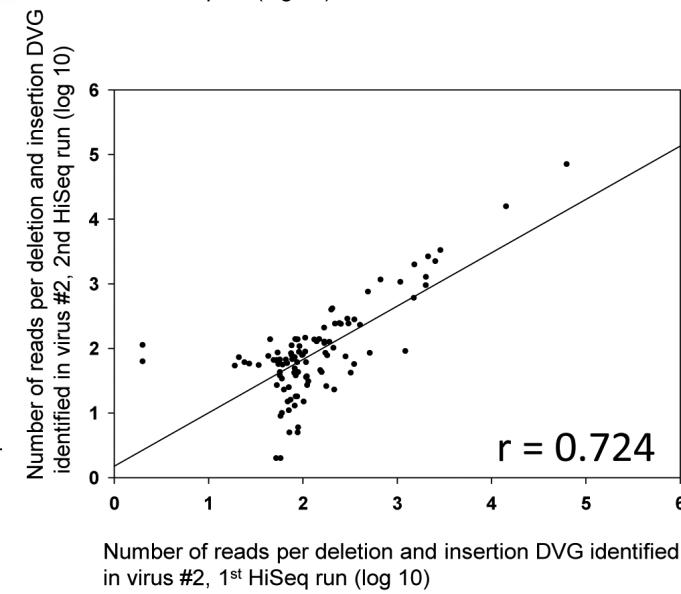
DVG S2



Reproducibility

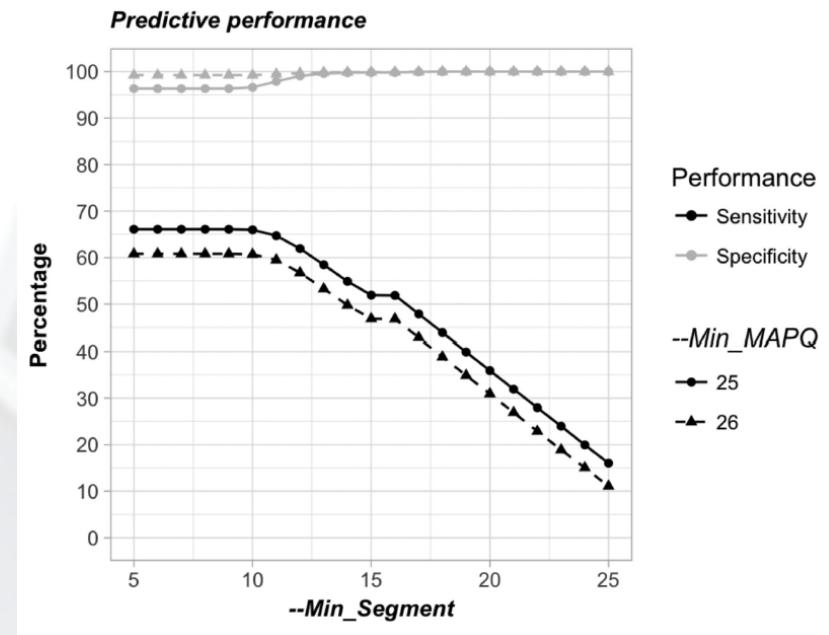


Copy-back
in/dels





DI-tector



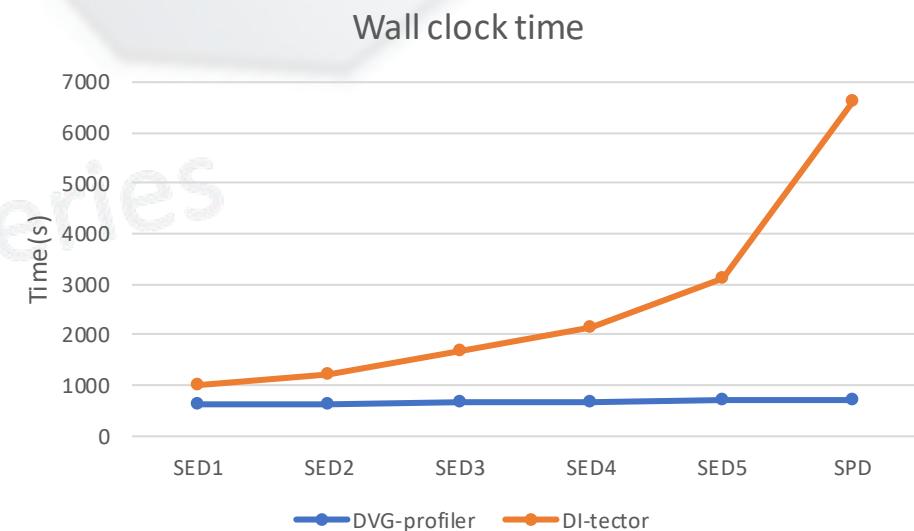
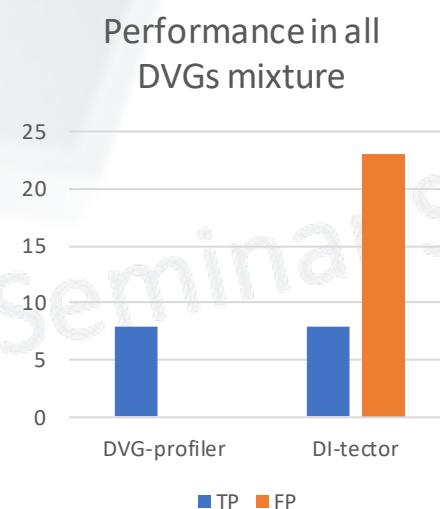
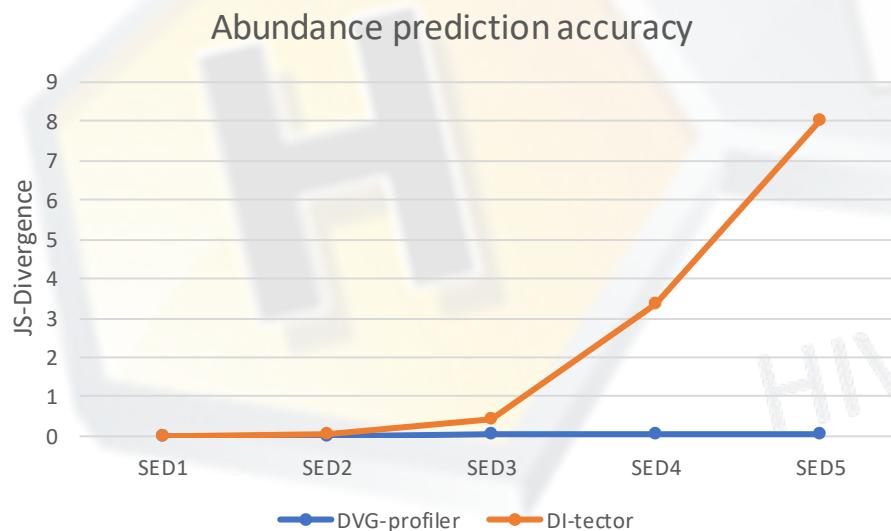
- Uses artificial split reads and BWA
- Complexity increases with the read length
- Available only as end-to-end



Comparison DI-tector and DVG profiler

In silico samples:

Name	Type	Breakpoint	Reinitiation point	Delta	Length (b)
REF	Reference	n/a	n/a	n/a	15384
DVG1	5' cb	13363	14924	561	2483
DVG2	5' cb	13257	14191	934	3322
DVG3	3' cb	520	2312	1792	2832
DVG4	3' cb	5585	5686	101	11271
DVG5	Deletion	522	5587	5065	10320
DVG6	Deletion	521	542	21	15364
DVG7	Insertion	2311	2292	19	15404
DVG8	Insertion	5686	5586	100	15485



Sensitivity concentrations:

Dataset	Composition	Av. Depth of Coverage	Prevalence (%)
SED1	REF/DVG3	6500.26/3.53	99.95 / 0.05
SED2	REF/DVG3	6500.26/3.53	99.46 / 0.054
SED3	REF/DVG3	6500.26/353.31	94.85 / 5.15
SED4	REF/DVG3	6500.26/3531.07	64.80 / 35.20
SED5	REF/DVG3	6500.26/35310.73	15.55 / 84.45



Tutorial – Reads in HIVE

Seminar Series

The screenshot shows the HIVE web interface. On the left, a sidebar titled "HIVE Space" lists various projects and datasets. In the center, a main panel displays a table of files under the "reads" category. The table has columns for ID, Name, Records, Size, and Created. Two entries are shown:

ID	Name	Records	Size	Created
258327	19C_R2_paired.fastq	7,523,581	1.78 GB	9/7/2016
258325	19C_R1_paired.fastq	7,523,581	1.78 GB	9/7/2016

Below the table, a histogram plot shows "base count" on the y-axis (0 to 4,000,000) against an unlabeled x-axis (0 to 80). The plot features two overlapping bell-shaped curves: a blue curve peaking around 25-30 and a green curve peaking around 15-20.

<https://hive.biochemistry.gwu.edu/dna.cgi?cmd=main>



Tutorial – Align Reads against reference

The screenshot shows the Hexagon Aligner software interface. On the left, the main menu includes 'Hexagon Aligner', '19C_R1_paired.fastq versus 88-mod.fa', 'Parameters' (with tabs for General, System, Batch, Advanced), 'Progress', and 'Results'. The 'Parameters' section is expanded, showing 'General' settings like 'Name: 19C_R1_paired.fastq' and 'Alignment Algorithm: HEXAGON 2.0 Native HIVE-hexagon algorithm op...nce Cloud computing environments'. Under 'Minimum Match', the 'Length' is set to 17 and 'Unit' is 'base(s)'. Under 'Matches to Keep', it says 'All Matches within acceptable thresholds'. Under 'Mismatches', 'Percent Allowed' is set to 15 and 'Computed on' is 'Minimum match length'. On the right, there is a 'Service Help' tab and a detailed 'HIVE-Hexagon' help page. The help page describes the HIVE-hexagon algorithm as finding short read alignments by seeding, extension and optimal alignment. It also mentions that the hexagon algorithm runs faster than other industry favorites. Below the help page is a results table showing a single entry:

ID	Summary	Created
259534	88-mod.fa 1 Genomic sequence(s)	9/14/2020

A sidebar on the right lists various projects and datasets: HIVE Space, All objects, Inbox, Trash, 359185, 362309, Censuscope, COVID-19, CVM, Hexahedron-Hept, hg19_chroms_m, hip arthroplasty, and Masa Project. A message at the bottom of the sidebar states: 'available to you through your user home directory. Select one or multiple reads by clicking on the corresponding row.'

<https://hive.biochemistry.gwu.edu/dna.cgi?cmd=main>



Tutorial – From Alignments to DVG profiler

Hexagon Aligner

19C_R1_paired.fastq versus 88-mod.fa

Parameters

Progress

Results

Pie Chart

Histogram

Saturation

Hit List

Alignments

Stack

Hit Tables

Help

Downloads

What's Next?

Modify and Resubmit

Sequence Profiling

Reference Recombination

DVG Profiling

Population Analysis

Targeted Alignment QC

Hit List Downloads

id	Reference	Hits	Length	RPKM	Density
0	Unaligned	946.61K	-	-	-
1	88-1961-mod	$14.95 \cdot 10^6$	15,384	65,002.6	90650.9
+	total	$14.95 \cdot 10^6$	15,384	-	90650

1-3 50 click here to select anno

Pie Chart Saturation

Download graph as SVG file

Histogram Alignments Stack Hit Tables Help

Start	Alignment	End	Sequence	Repeats
14456	ACAACTGCTCTGAATAGGGCTCGCACCCtGAATGAACAAGGCTTTCACTCATCCCACCTGAATTGGTTaGTGAGTACTGGAGGAGG	14542	88-1961-mod	
14	ACAACTGCTCTGAATAGGGCTCGCACCCtGAATGAACAAGGCTTTCACTCATCCCACCTGAATTGGTTgGTGAGTACTGGAGGAGG	100	SN1029:244:HVKN7...117 2:N:0:GTGAAA 1	
14451	CCTTCACAACTGCTCTGAATAGGCTCGCACCCtGAATGAACAAGGCTTTCACTCATCCCACCTGAATTGGttAGTGAGtACTGGAGGAGGC	14549	88-1961-mod	
2	CCTTCACAACTGCTCTGAATAGGGcCGCACCCGAATGAACAAGGCGTCACCATCCCACCTGAATTGGccAGTGAGcACTGGAGGAGGC	100	SN1029:245:HVHKN...793 1:N:0:GTGAAA 1	
14472	GGGCTCGCACCCtGAATGAACAAggCTTTCACCATCCCACCTGAATTGGTTaGTGAGTaCTGGAGGAGGC	14550	88-1961-mod	
1	GGGCTCGCACCCtGAATGAACAGGGCTTTCACCATCCCACCTGAATTGGTTgGTGAGTgCTGGAGGAGGC	79	SN1029:245:HVHKN...390 2:N:0:GTGAAA 1	
14450	TCCTTCACAACTGCTCTGAATAGGGCTCGCACCCtGAATGAACAAGGCTTTCACCATCCCACCTGAATTGGTTAGTGAGTACTGGAGGAGG	14545	88-1961-mod	
4	TCCTTCACAACTGCTCTGAATAGGGCTCGCACCCtGAATGAACAAGGCTTTCACCATCCCACCTGAATTGGTTAGTGAGTACTGGAGGAGG	99	SN1029:245:HVHKN...607 2:N:0:GTGAAA 1	
14450	TCCTTCACAACTGCTCTGAATAGGGCTCGCACCCtGAATGAACAAggCTTTCACATCC	14510	88-1961-mod	
40	TCCCTCACAACTGCTCTGAATGGGCTCGCACCCtGAATGAACAGGGCTTTCACATCC	100	SN1029:245:HVHKN...146 1:N:0:GTGAAA 1	
14450	TCCTTCACAACTGCTCTGAATAGGGCTCGCACCCtGAATGAACAAGGCTTTCACATCC	14511	88-1961-mod	
		75		
14	TCCTTCACAACTGCTCTGAATAGGGCTCGCACCCtGAATGAACAAGGCTTTCACATCC	75	SN1029:244:HVKN7...963 1:N:0:GTGAAA 1	
14450	TCCTTCACAACTGCTCTGAATAGGGCTCGCACCCtGAATGAACAAGGCTTTCACATCCCACCTGA	14517	88-1961-mod	

Position 14500

Search

<https://hive.biochemistry.gwu.edu/dna.cgi?cmd=main>



Tutorial – DVG profiler parameters

HIVE Defective Viral Genomes Detection

General System Batch

Name DVG profile based on

Alignments ID=425368 | 19C_R1_paired.fastq versus 88-mo...>HIVE-hexagon Alignment 100% !

Distance of aligned parts

- Maximum distance 50 1
- Maximum overlap 5 2
- Minimum depth of coverage 2
- Minimum length of aligned read 70

Analyze All References

Advanced Service Help

Overlap score threshold(%) 0

Minimum length of aligned subject 20

Group width 5

Treat multiple alignment combinations Keep combination with best score

Is genome antisense

Print construct details

Analyze All References

Parameters

General System Batch Advanced Alignment

Progress

Results

1) Maximum distance

— d bp —

2) Maximum overlap

— o bp —



Tutorial – DVG profiler results

HIVE Defective Viral Genomes Detection

DVG profile based on alignment : 19C_R1_paired.fastq versus 88-mod.fa (425368)

Parameters

Progress

Results

All hits

Grouped hits

What's Next?

Home Page

Modify and Resubmit

Back to Alignment

All hits Grouped hits

Download DI reads Download DI alignments Load New Data Data Source New Column Graphs Analysis Glue Tables Download the table Reset Table 2001-3000 10

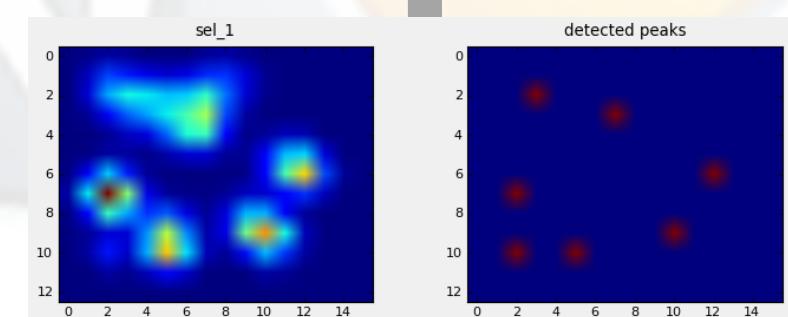
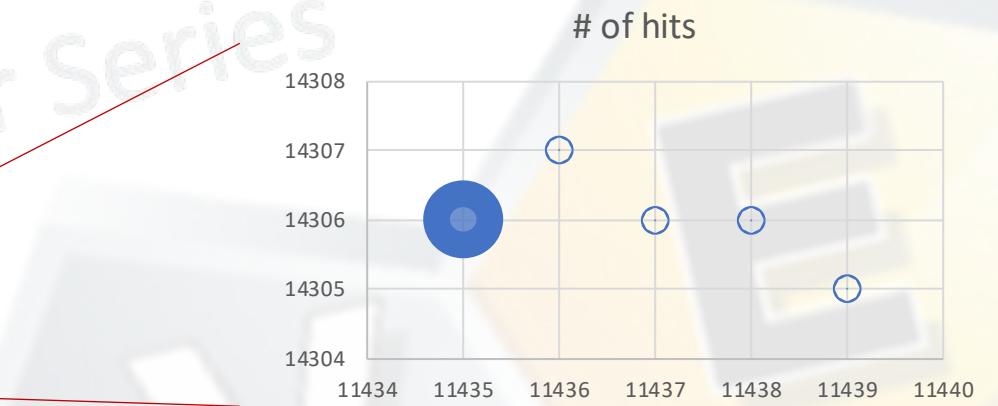
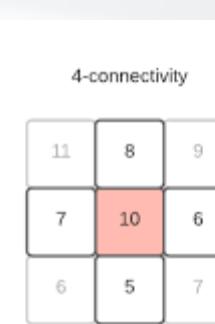
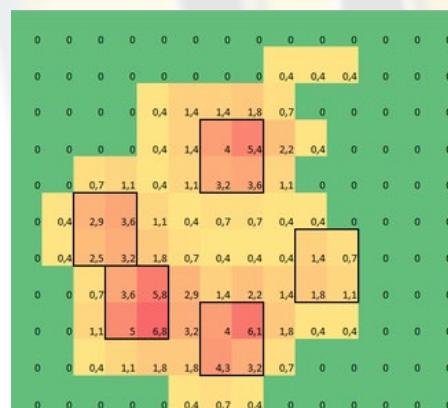
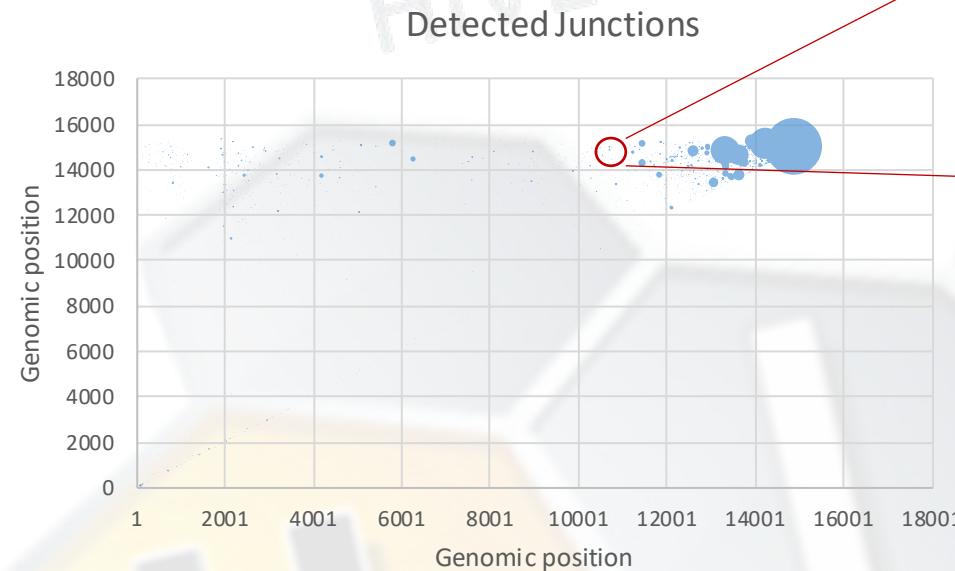
10269 total rows 9 total columns

Position (left)	Strandness (left)	Position (right)	Strandness (right)	DVG	Length	Delta	Forward hits	Reverse hits
6212	-	11117	+	5' cb	13443	4905	0	1
6225	+	13309	+	deletion	8301	7084	1	1
6239	+	6264	-	3' cb	12503	25	1	0
6242	-	6459	-	insertion	15602	217	1	0
6249	+	14475	+	deletion	7159	8226	1	0
6250	+	14476	+	deletion	7159	8226	491	168
6250	+	14478	+	deletion	7157	8228	4	0
6251	+	14477	+	deletion	7159	8226	1	0
6256	+	14465	-	3' cb	20721	8209	1	0
6257	+	14455	-	3' cb	20712	8198	1	0
6257	+	14465	-	3' cb	20722	8208	5	0
6257	+	14485	+	deletion	7157	8228	3	0
6257	+	14731	+	deletion	6911	8474	1	1
6261	+	14427	+	deletion	7219	8166	1	0
6269	+	11987	+	deletion	9667	5718	2	0
6272	-	14532	+	5' cb	9968	8260	2	2
6274	-	6582	-	insertion	15693	308	2	0
6275	-	14116	+	5' cb	10381	7841	1	0
6275	+	8186	+	deletion	13474	1911	1	1
6276	-	13093	+	5' cb	11403	6817	0	1
6289	+	6355	+	deletion	15319	66	1	0
6291	+	6328	+	deletion	15348	37	1	1
6292	+	7920	+	deletion	13757	1628	2	0
6293	-	6461	-	insertion	15553	168	0	1
6294	-	6554	+	5' cb	17924	260	0	3
6302	+	6350	+	deletion	15337	48	1	1
6303	-	13911	+	5' cb	10558	7608	2	0
6303	+	6363	+	deletion	15325	60	2	0
6306	+	15277	+	deletion	6414	8971	0	1
6307	+	6564	+	deletion	15128	257	1	0
6307	+	13326	+	deletion	8366	7019	1	1
6310	-	13558	+	5' cb	10904	7248	1	0
6324	+	6351	+	deletion	15358	27	0	1
6326	+	15295	+	deletion	6416	8969	0	1

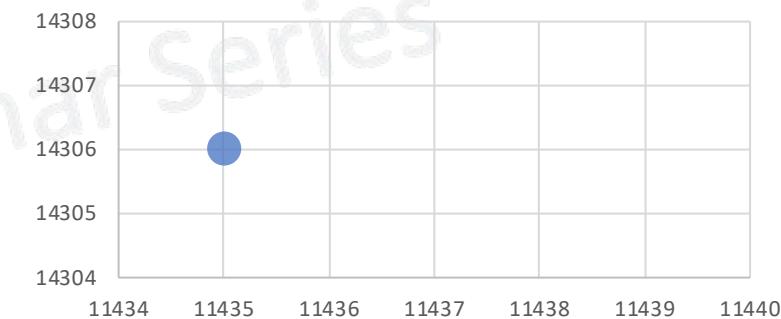
<https://hive.biochemistry.gwu.edu/dna.cgi?cmd=main>



2D peak detection



of hits (grouped)





Tutorial – DVG profiler results refined

HIVE Defective Viral Genomes Detection

DVG profile based on alignment : 19C_R1_paired.fastq versus 88-mod.fa (425368)

Parameters

Progress

Results

All hits

Grouped hits

What's Next?

Home Page

Modify and Resubmit

Back to Alignment

All hits Grouped hits

Download DI sam Download DI alignments Download DI reads Load New Data Data Source New Column Graphs Analysis Glue Tables Download the table Reset Table

4489 total rows 13 total columns

Position (left)	Group start (left)	Group end (left)	Strandness (left)	Position (right)	Group start (right)	Group end (right)	Strandness (right)	DVG	Length	Delta	Forward hits	Reverse hits
544	-	-	+	14476	-	-	+	deletion	1453	13932	5	0
6250	6249	6251	+	14476	14475	14478	+	deletion	7159	8226	497	168
6386	-	-	-	14476	-	-	+	5' cb	9910	8090	2	0
12903	12900	12905	-	14476	-	-	+	5' cb	3393	1573	1	5
14138	14133	14141	-	14476	14476	14480	-	insertion	15723	338	1	3



Identifying junctions is the first step

- Nested events
 - RT-PCR detected deletions in cpRNAs
- DVGs mutations and quasispecies

Predicted size of DVG (nt)	Break-point position	Reinitiation position	Loop size (nt)	Stem size (nt)	Detected by RT-PCR using primer pair ^b	Size of PCR fragment in bp ^c
1584 1125 ^e	13907–13909 ^a	15279–15277 ^a	1368– 1372 909–913	108–106 108–106	a/b1 a/b2 a/b3 a/b3	993 (2a1) 1231(2b1) 1331 (2c1) 870 (2c2) ^e
1429	14456	14885	429	500	a/b1	838 (2a2) ^d
1381 922 ^f	14223–14224 ^a	15166–15165 ^a	941–943 482–484	219–220 219–220	a/b1 a/b3	790 (2a2) ^d 668 (2c3) ^f
1320	14342	15108	766	277	a/b1	728 (2a3) ^d
1278	14596	14896	300	489	a/b1	685 (2a3) ^d
1092	14761	14917	156	468	a/b1	500 (2a4)
1014	14730–14733 ^a	15026–15023 ^a	290–296	362–359	a/b1	423 (2a5)
870	14868	15032	164	353	a/b2	407 (2b2)



Quasispecies

4×10^{-4} errors / nucleotide / round of replication



Highly diverse populations, mutant clouds, swarms
Quasispecies

Quasispecies is a mathematical model by Manfred Eigen¹

$$w_{ij} = A_i q_{ij} - D_i \delta_{ij}$$

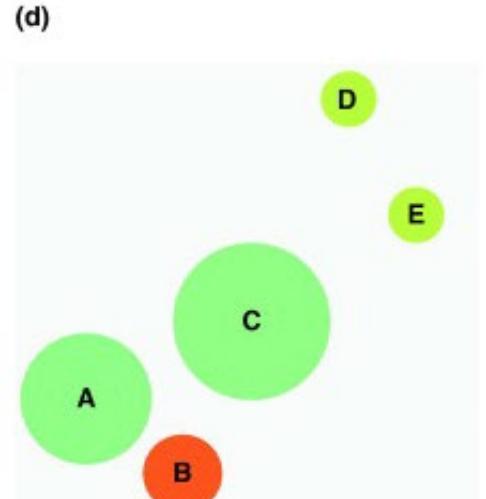
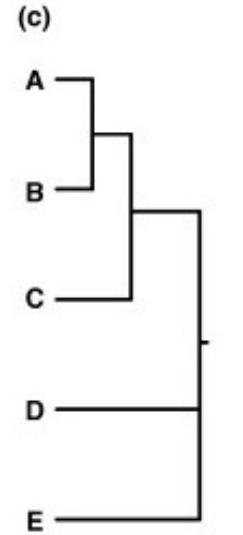
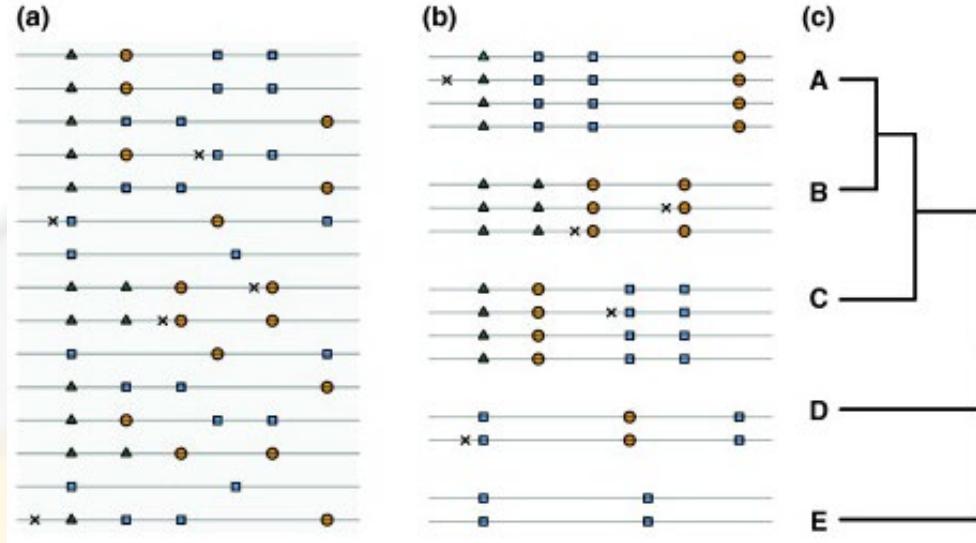
w_{ij} : type i sequence from parent sequence j

$A_i q_{ij}$: offspring A_i with mutation rate q_{ij}

$D_i \delta_{ij}$: death rate D_i of sequence i ($\delta_{ij} = 0 \forall i = j$ and $\delta_{ij} = 1 \forall i \neq j$)



Problem definition

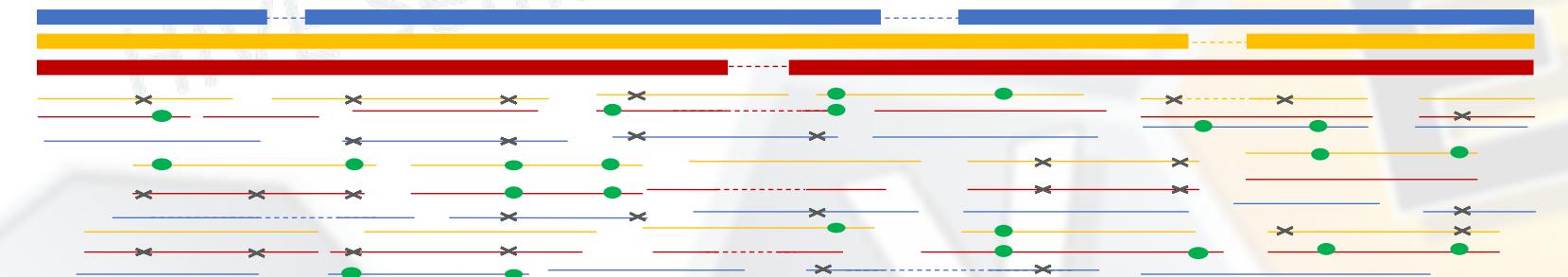


Quasispecies Spectrum Reconstruction (QSR) means:

- the set of sequence consensuses and
- the relative frequency of each sequence in the sample



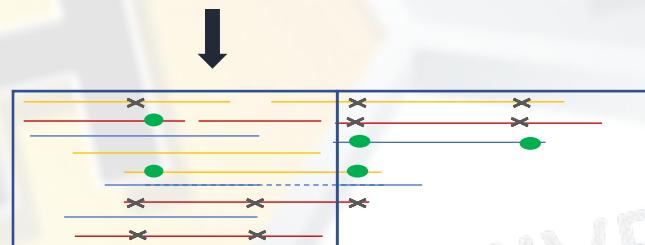
HexaHedron algorithm



Alignment length to window half length



start end
Sliding window



Window profile

Pos	A	C	G	T	Ins	Del
1	50	1	0	0	0	0
2	0	0	65	0	0	1
3	0	82	1	0	0	0
4	0	67	0	64	0	0
5	0	44	0	86	0	0
6	1	0	99	0	0	0
7	76	0	41	0	0	0
8	0	32	41	47	0	1
9	104	0	4	0	0	1
10	0	11	0	80	1	0
11	0	56	0	17	0	0
12	64	0	0	0	0	0
13	0	1	48	0	0	0
14	0	0	0	39	0	0

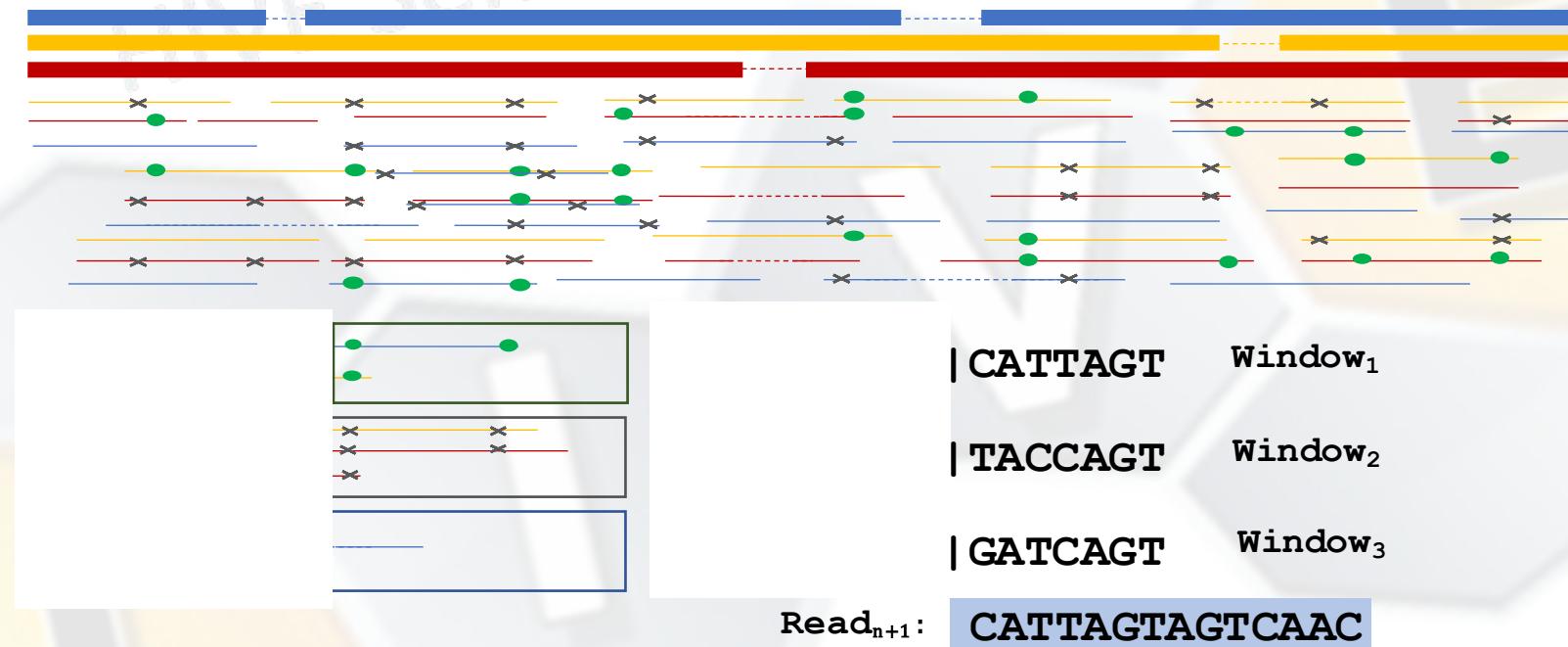
← start

end

overlap



HexaHedron algorithm



Hamming similarity normalized by Sørensen–Dice index

$$H = \frac{HSim(R_{R \cap W}, W_{R \cap W})}{QS(R, W)}$$

$$QS(R, W) = \frac{2|R \cap W|}{|R| + |W|}$$

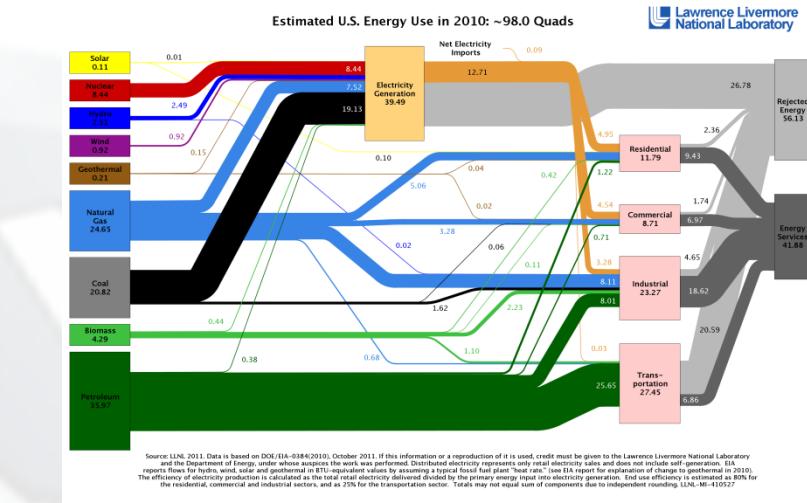
Exceptions:

- During bifurcation
- Paired end read with assigned mate

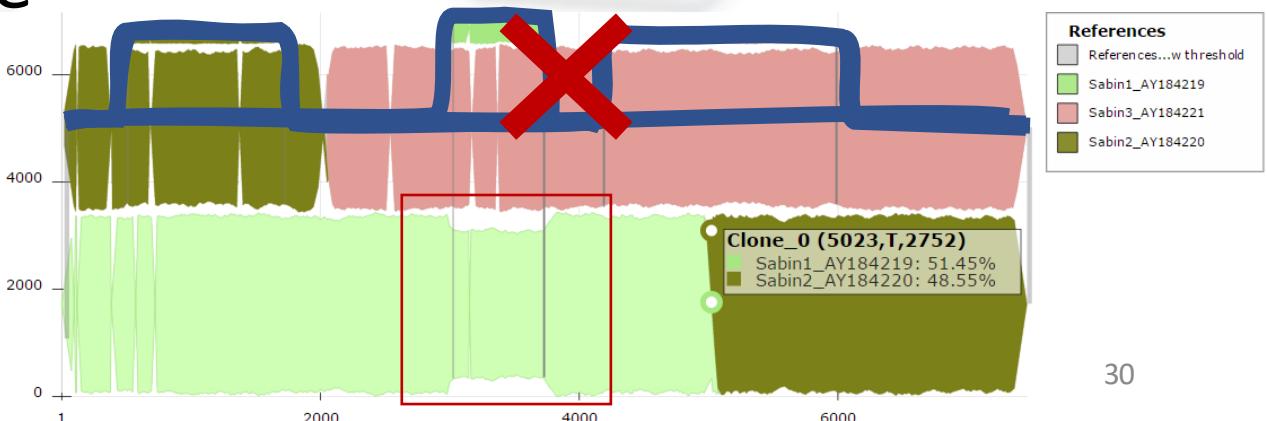


HexaHedron algorithm

- Sankey diagram
 - + Flow diagram
 - Energy between processes
- Constant width
- No axis information



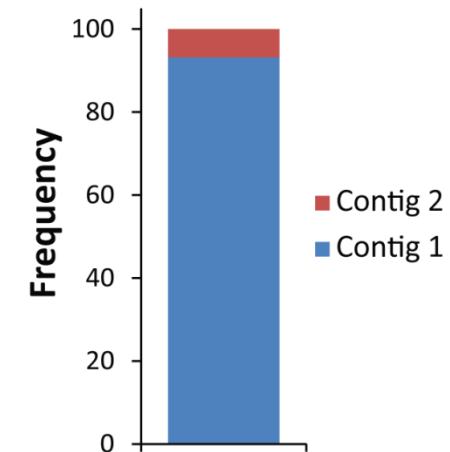
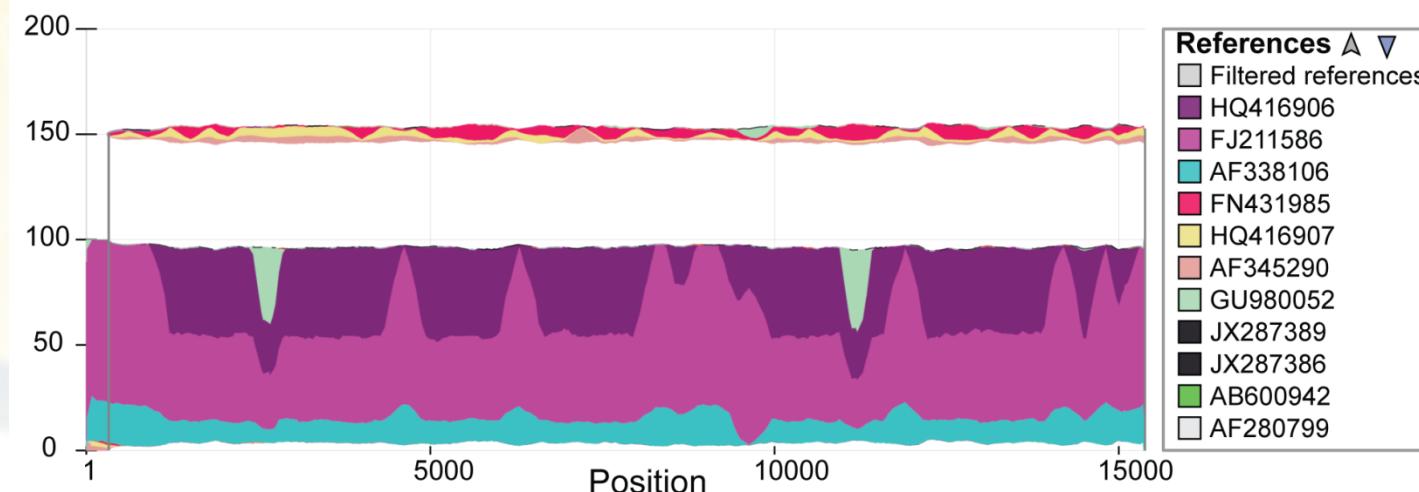
- Nephosome / Graph Genome
 - Width corresponds to the depth of the coverage
 - Trajectories from left to right





Empirical sample (Mumpsvax)

- DB of genomic sequences of 54 strains of mumps
- Alignment produced 688,000 hits
- Predicted frequencies : **93.18% and 6.82%**. Consistent with previous estimate based on quantitative PCR (95%-5%)
- Consensus sequences of these two sub-strains were identical to those determined by conventional sequencing of plaque-purified clones





Source code

HEXAHEDRON



<https://github.com/kkaragiannis/hexahedron>



<https://github.com/kkaragiannis/pan-HIV-ARM-quasispecies-analysis/>



Tutorial

The screenshot displays the HIVE Space interface, a bioinformatics tool for managing and analyzing sequencing data. The left sidebar shows a tree view of the HIVE Space, listing various projects and datasets such as Censoscope, Clonal, CVM, Demo, Demo Genomes, Demo Reads, Diprofiler_publication, Genomes, HIV-pack, Human-HIV-crosslink-project, Influenza-Project, Influenza, MiSeq, Model organisms, MUMPS, N120630_Lunhua, nt.fa, Optimization_of_RNA_library, OPV3 consistency project 2013, and Others. The main area features a search bar at the top with filters for all, folders, genomes, reads, research projects, annotations, files, and computations. Below the search is a table showing two fastq files: Indian_OPV_Types123_TGACCA_L008_R2.fastq and Indian_OPV_Types123_TGACCA_L008_R1.fastq. The table includes columns for ID, Name, Records, Size, and Created. At the bottom, a detailed plot titled 'positionalQC' shows base count versus position in the reads, with a sharp initial peak followed by a stable baseline around 8 million counts.

ID	Name	Records	Size	Created
994	Indian_OPV_Types123_TGACCA_L008_R2.fastq	31,699,866	6.36 GB	12/19/2012
992	Indian_OPV_Types123_TGACCA_L008_R1.fastq	31,699,866	6.37 GB	12/19/2012

base count

position in the reads

Download graph as SVG file

<https://hive.biochemistry.gwu.edu/dna.cgi?cmd=main>



Tutorial

The screenshot shows the MAFFT Aligner interface. On the left, a sidebar lists 'MAFFT Aligner', '> Sabin_123_aligned' (highlighted with a yellow bar), 'Parameters' (with 'General' selected), 'Progress', and 'Results'. The main panel has tabs for 'General', 'System', 'Batch', and 'Advanced'. Under 'General', the 'Name' is set to 'Sabin_123_aligne', 'Alignment Algorithm' is 'MAFFT 7.407 MAFFT multiple alignment algorithm', and 'Reference Genome' is 'ID=6235 | Sabin123.fasta 3 Genomic sequence(s)'. A large 'ALIGN' button is at the bottom right. To the right of the main panel is a sidebar titled 'Mafft Alignment Program' with a help icon, a brief description of MAFFT as a multiple sequence alignment program for unix operating systems, and a link to access the Mafft help document.

<https://hive.biochemistry.gwu.edu/dna.cgi?cmd=main>



Tutorial

Hexagon Aligner

Indian_OPV_Types123_TGACCA_L008_R2.fastq versus Sabin123.fasta

Parameters

- General
- System
- Batch
- Advanced

Progress

Results

General System Batch Advanced

Name Indian_OPV_Types1

Alignment Algorithm HEXAGON 2.0 Native HIVE-hexagon algorithm op...nace Cloud computing environments

Sequence Read ID=994 | Indian_OPV_Types123_TGACCA_L008_...699866 Nucleotide read(s)

Sequence Read (Pair)

Reference Genome ID=6235 | Sabin123.fasta 3 Genomic sequence(s)

Minimum Match

Length	Unit
75	base(s)

Matches to Keep Random vote between equally best alt

Mismatches

Percent Allowed	Computed on
15	Minimum match length

Selected object(s): 1 genome

available to you through your user home directory. Select one or multiple reads by clicking on the corresponding row. To close the window click the X icon in the top right corner.

Service Help Advanced Parameters Help

HIVE-Hexagon

DNA Sequence Alignment

HIVE-hexagon is an alignment algorithm that finds short read alignments by seeding, extension and optimal alignment. The hexagon algorithm runs faster than other industry favorite...

close

home help

Cut Edit

All [6993]

ID Summary Created

ID	Summary	Created
359063	ERR166235.sra Folder	8/22/2017
17653	Inbox System Folder	1/23/2014
6235	Sabin123.fasta 3 Genomic sequence(s)	4/10/2013

Selected object(s): 1 genome



Tutorial

HIVE Seminar Series

Advanced Parameters

alignments stack hit tables downloads help

Start Position End Sequence Repeats

Name Clonal analysis with b

Alignment(s) ID=92551 | Indian_OPV_Types123_TGACCA_L008...>HIVE-hexagon Alignment 100%

Paired End Reads no

Threshold for bifurcation call (percentage) 1

Reference Genome Alignment File > Sabin_123_aligned HIVE-optimized External Alignment 100%

Minimum bifurcation coverage 20

Quality threshold 0

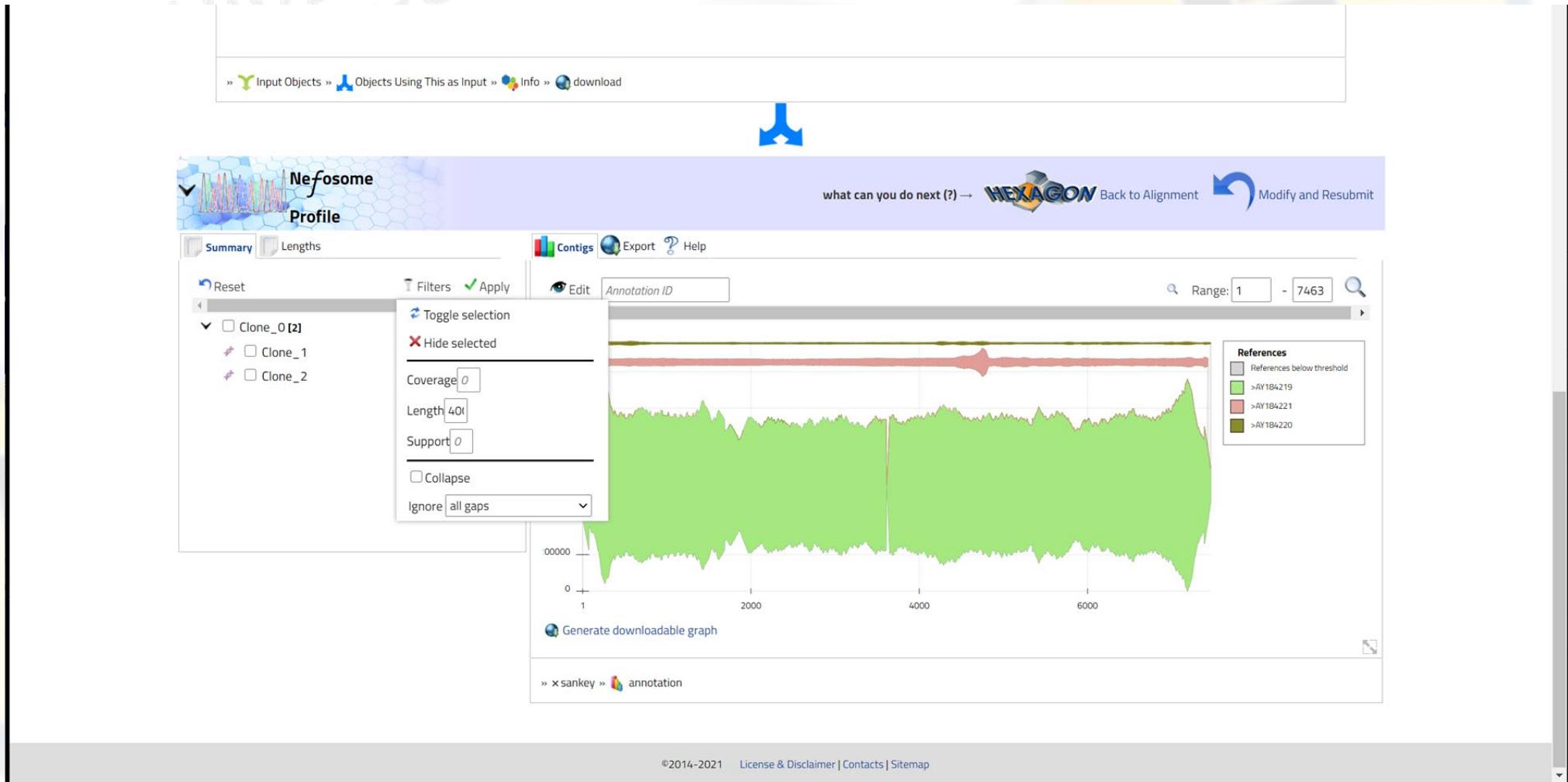
Batch Service

The screenshot shows the HIVE software interface. On the left, there's a table of search results with columns: id, Reference, Hits, Length, RPKM, and Dens. The results include Unaligned reads (4.14e6), several reads from reference AY184219 (54.69e6, 7,441, 125,581.5, 7357), and other reads from AY184221, AY184220, and a total row. On the right, there's a large panel displaying sequence alignments between the user's sample and the reference genome. The alignments are shown as pairs of sequences with vertical lines indicating matches. Below the alignments, various parameters for the analysis are listed, such as the alignment ID, reference genome, and quality thresholds.

<https://hive.biochemistry.gwu.edu/dna.cgi?cmd=main>



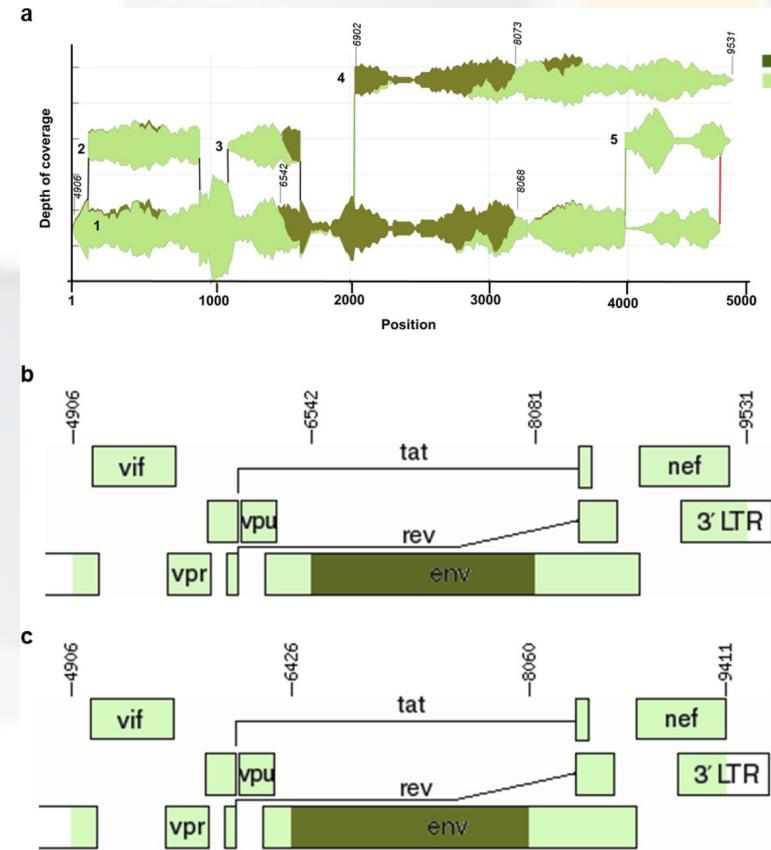
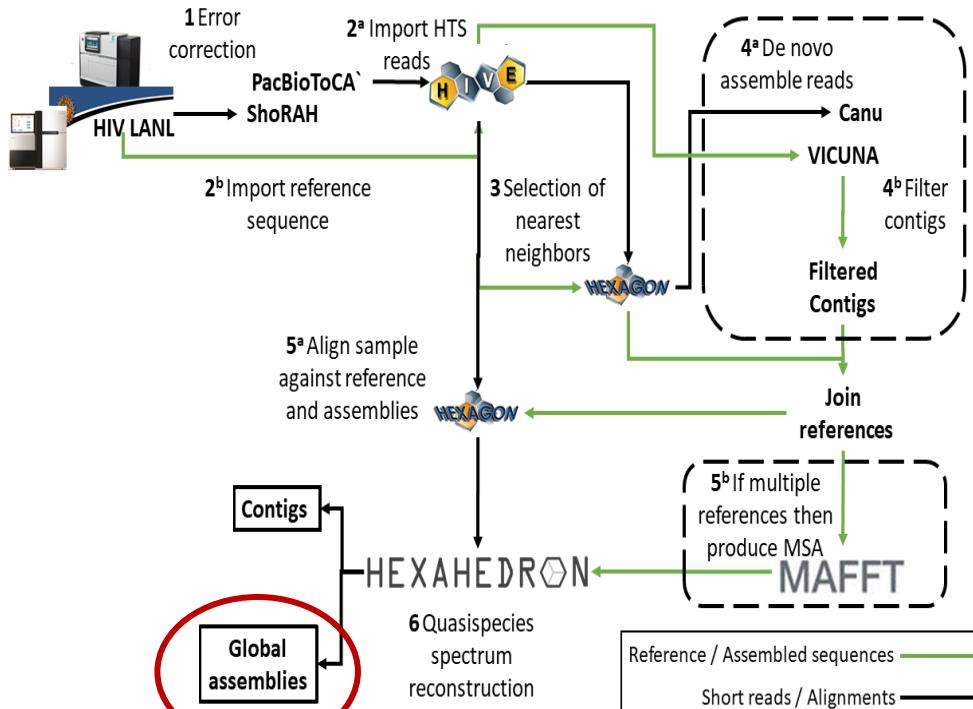
Tutorial



<https://hive.biochemistry.gwu.edu/dna.cgi?cmd=main>



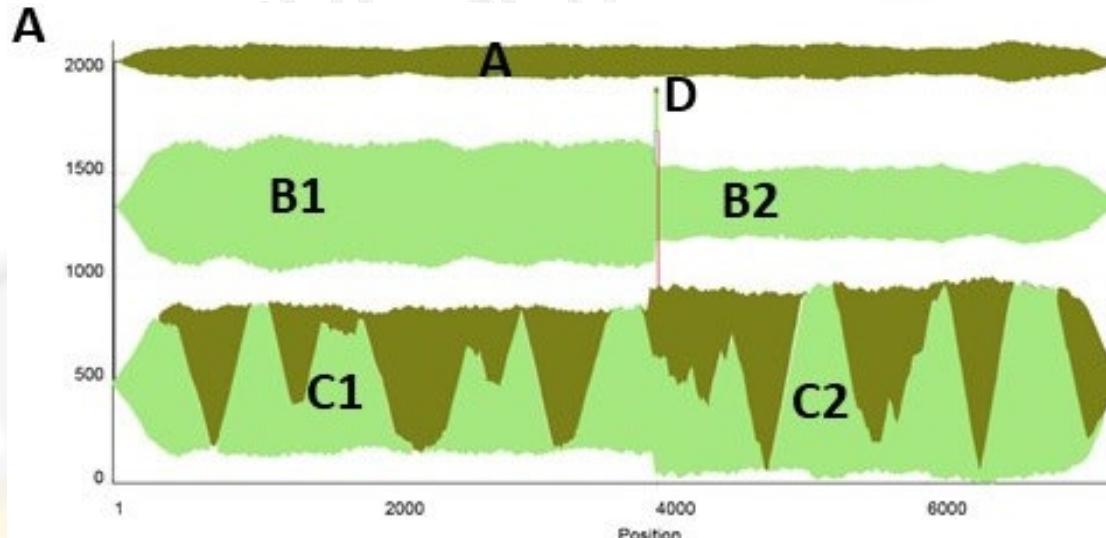
HIV Pipeline



- Subtyping
- Refined reference selection
- Global sequences

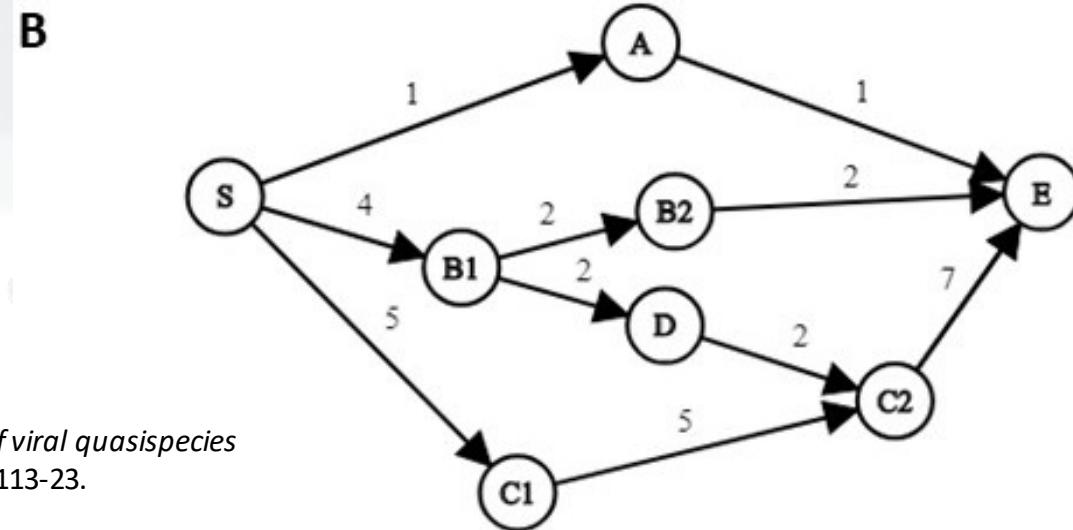


Global sequence inference



Convert Nephosome to a weighted directional graph $G = (V, E)$ where each contig is a vertex $v \in V$

Run Monte Carlo simulations¹ until user specified limit



¹Topfer, A., et al., Probabilistic inference of viral quasispecies subject to recombination. J Comput Biol, 2013. 20(2): p. 113-23.



Tutorial

Hive Seminar Series

» Input Objects » Objects Using This as Input » Info » download

Nefosome Profile

what can you do next (?) → **HEXAGON** Back to Alignment **Modify and Resubmit**

Summary **Lengths**

Reset Filters Apply

Clone_0 [2]
Clone_1
Clone_2

Contigs Export Help

Download Digest

Limit 5 Select extended ▾
contig extended

Print consensus Thresholds Apply

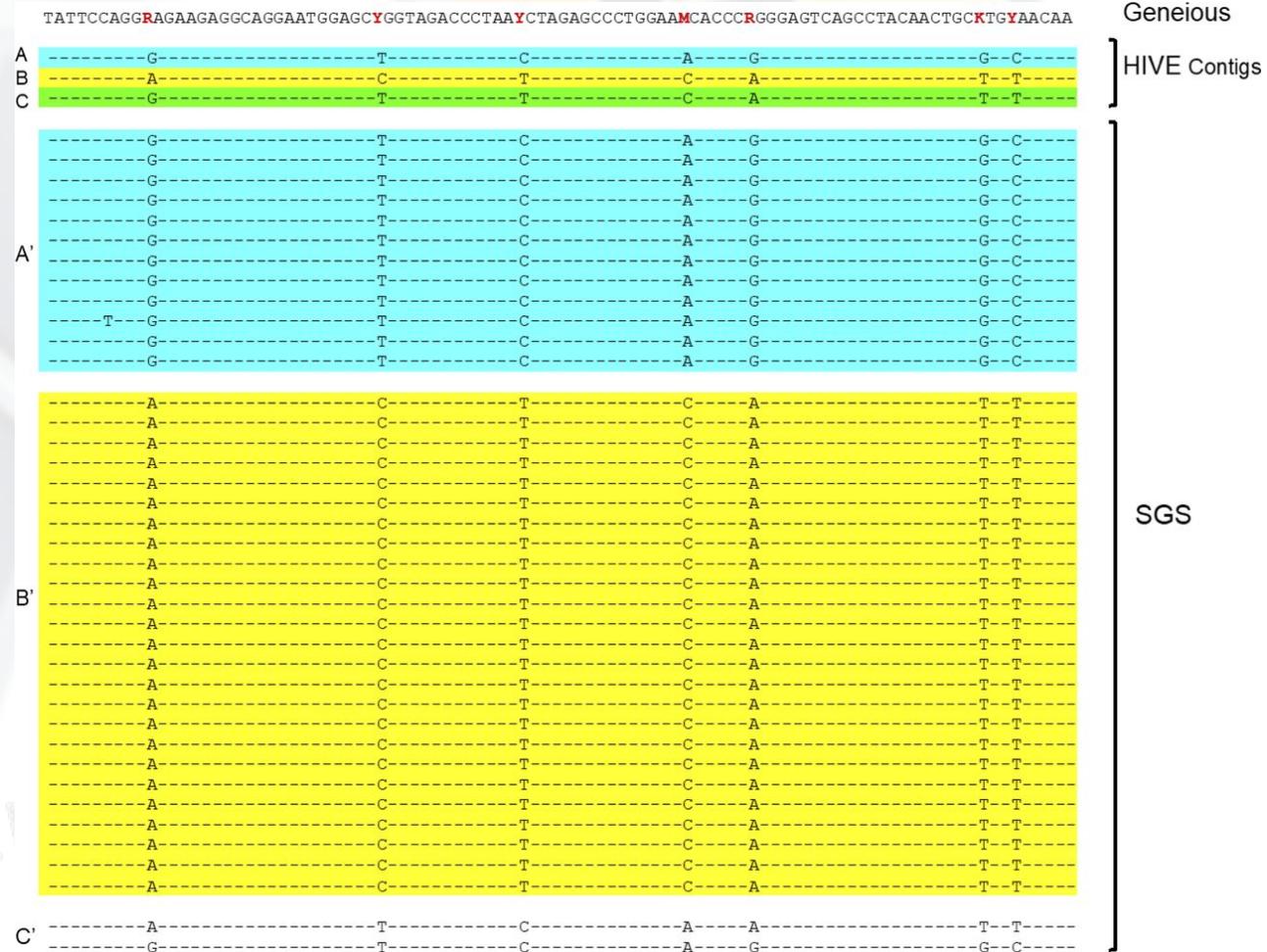
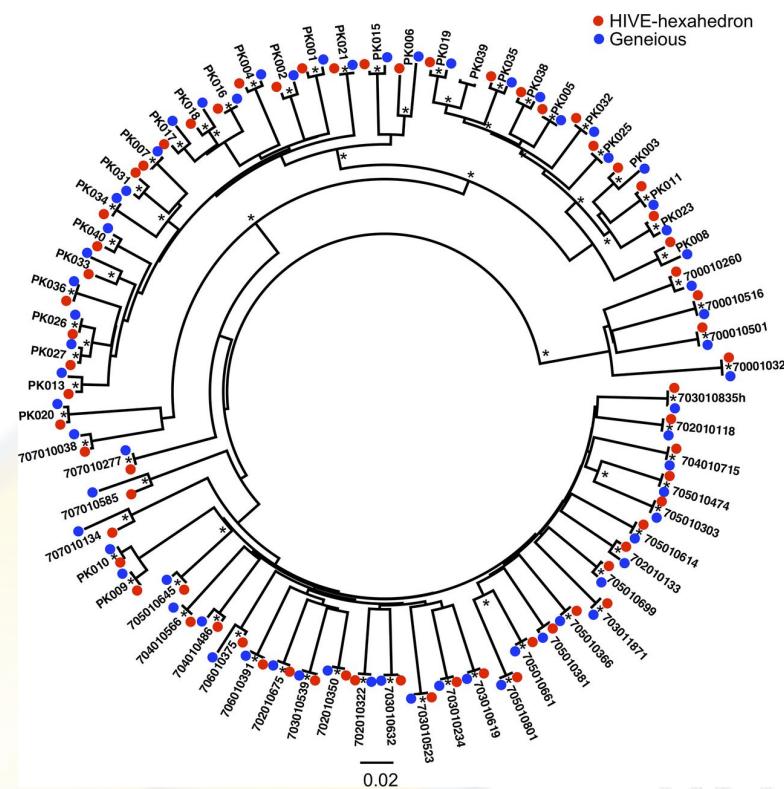
>clone_0 | composition:contig_0[1-7463]|object 92563|Clonal analysis with alignment : Indian_OPV_Types123_TGACCA_L008_R2.fastq versus Sabin12 TTAAACAGCTCTGGGGTGCACCCACCCCAGAGGCCACGTGGCGGCTAGTACTCC CAAGTTCAATAGAAGGGGGTACAACCAAGTACCAACCACGAACAAGCACTCTGTTC ACTTCGAGAACGCCAGTACCAACCTCGGAATCTCGATGCGTTGCGCTCAGCACTCAACCCAGAGTGACTCTAGGCTGATGAGCTGGACATCCCTCA GCGTTGGCGCCTACCATGGCTAACGCCATGGGACGCTAGTTGAACAAGGTGAAGAGCCATTGAGCTACATAAGAACCTCCGGCCCTGAA GCAGGGTGTACAAACCAGTATTGGCTGTGTAACCGCAGTCCGTGGCGAACGACTACTTGGGTGTCGTGTTCTTTATTTATTGTG GATTGTTATCATAAGCGAATTGGATTGGCATCCGGTGAAGTGAAGATTCTATTCTATCTGTTGCTGGATTGCGCTCATTGAGTGTGTTACTCTAA TCAATCAGACAATTGTATCATAATGGGTGCTCAGGTTCTACAGAAAAGTGGGCCACATGAAAACCTCAAATAGAGCGTATGGGGTTACCATTA TAGAGATTAGCTAGTAACCGCCCTCGAACAGGACTCTCTCAAGACCCCTCAAGGTTCAAGCTCAGGACCCATCAAGGATGTCCTGATAAAAACATCC AGAGGCTTGGGGTATAGCGATAGAGTACTGCAATTAACTGGAAACTCCACTATAACCACACAGGAGGCGCTAATTAGTAGTCGTTATGGC CAGCGAAGCCAATCCAGTGGACCAGCCGACAGAACCCAGCAGTCGCTGATGCAAGGTTTACGCTAGACACCGTCTGGACGAAAGAGTCGCA TGCACTGCGGGACATGGGACTCTTGGCCAAATATGACTACCAACTCTAGGTTAGGTCCGGGTACACCGTGCATGACAGTGTACAGCGTAAACGCCCTCAAATT ATTGCGCTGACAGAGATGTGCTGGCCGGGGATAGCAACACCAACTACCATGACACCCAGCTACAAATGCCAATCTGGCGAGAAAGGAGGCACT

©2014-2021 License & Disclaimer | Contacts | Sitemap

<https://hive.biochemistry.gwu.edu/dna.cgi?cmd=main>

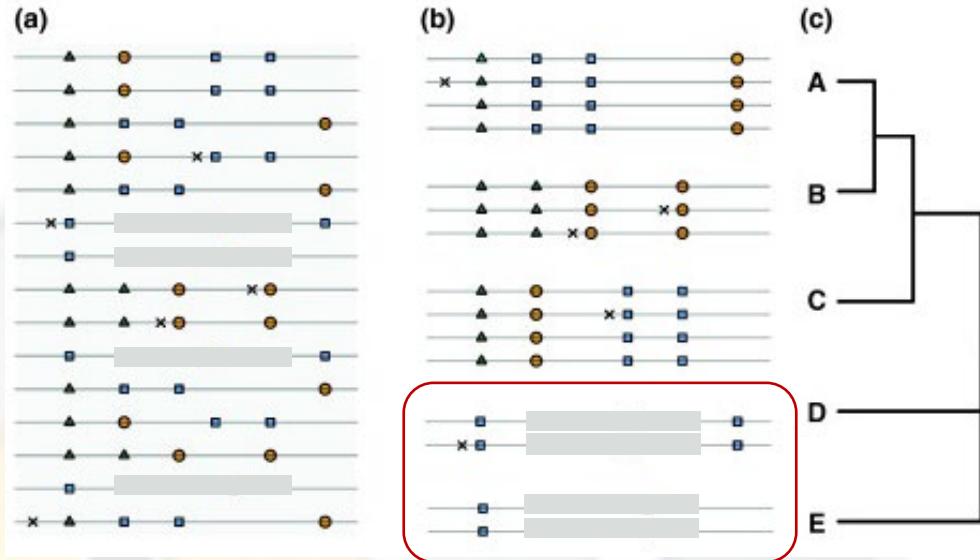


HIV Pipeline





QSR Problem redefined

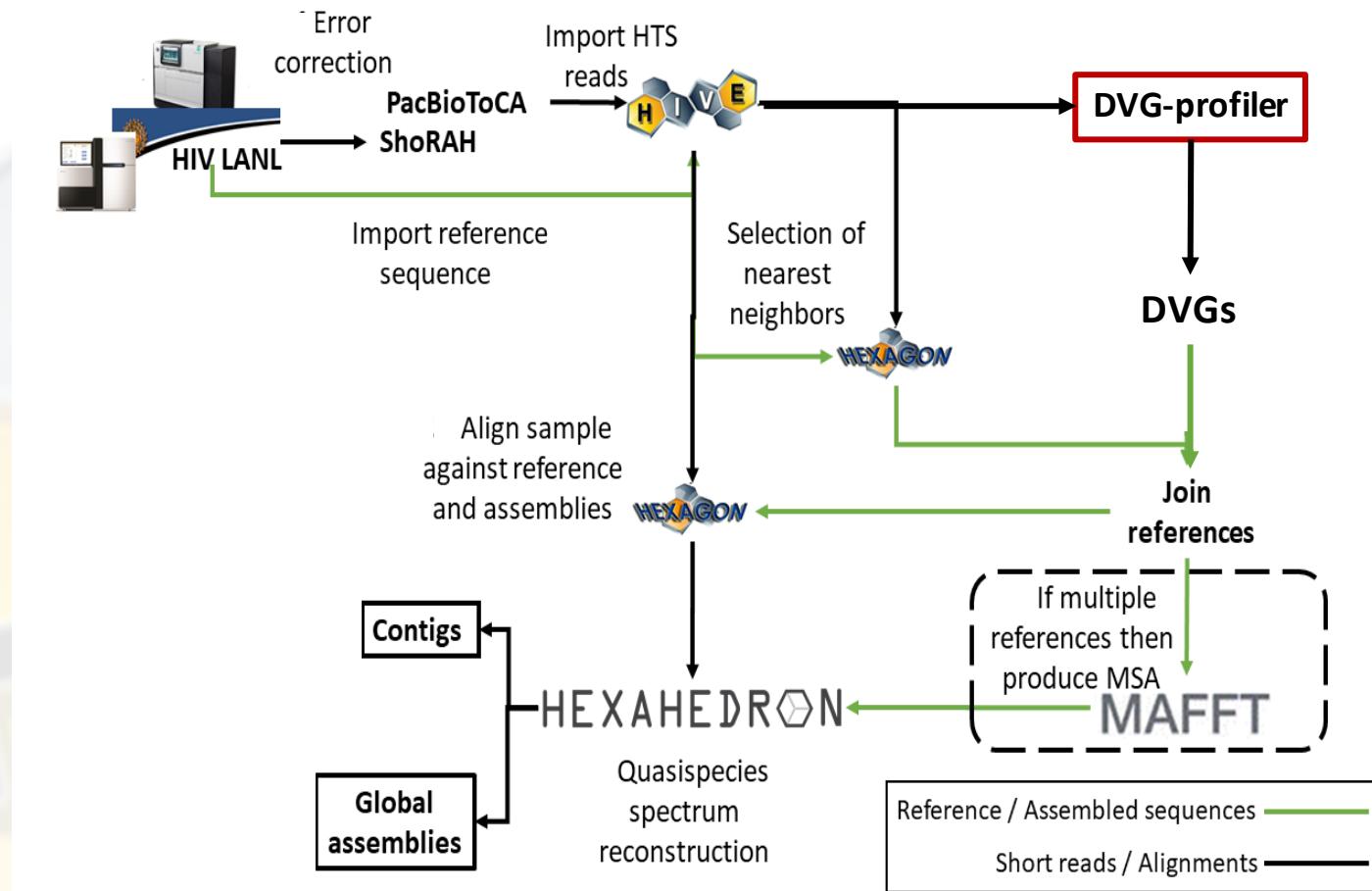


Quasispecies Spectrum Reconstruction (QSR) amended:

- Identify the DVGs
- Reconstruct Defective and non-Defective Viral Genomes consensuses
- Quantitate the relative frequency of each consensus in the sample



Future work





Acknowledgments



FDA/HIVE

Dipankar Chattopadhyay
John Dougherty
Arya Eskandarian
Marianna Faradzheva
Sydney Fenstermaker
Tigran Ghazanchyan
Anton Golikov

Sergey Ivanovsky

Kural Kamil

Alexander Lukyanov

Hasmik Manukyan

Ilya Mazo, PhD

Luis Quintero-Santana, PhD

Krista Smith

Sean Smith, PhD

GWU/HIVE

Raja Mazumder, PhD
Naila Gulzar
Charles Hadley King

FDA/OVRR

Christian Sauder, PhD
Trent Bosma, PhD
Konstantin Chumakov, PhD
Majid Laassri, PhD

DUKE/EQUAPOL

Feng Gao, MD
Bhavna Hora, PhD

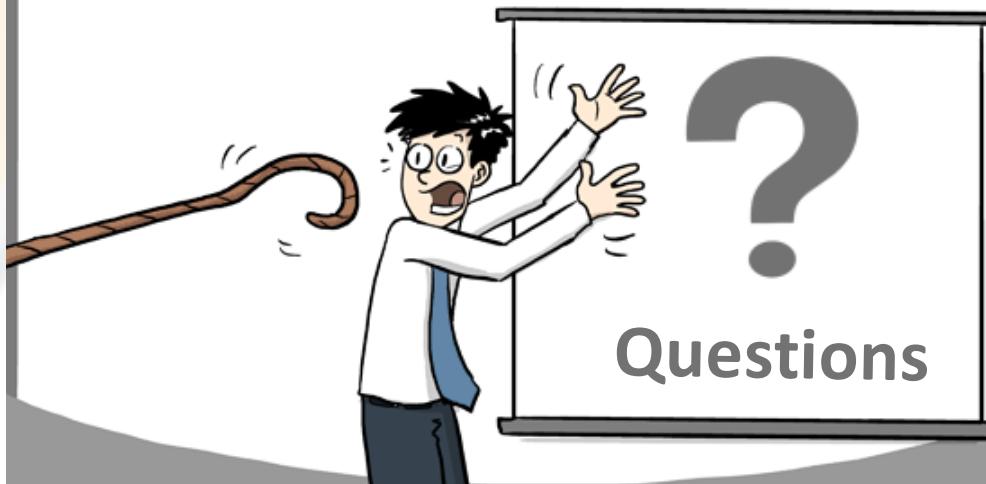
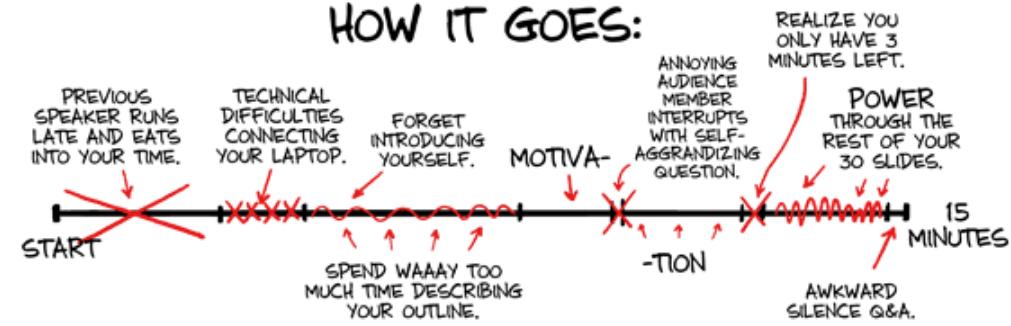


YOUR CONFERENCE PRESENTATION

HOW YOU PLANNED IT:



HOW IT GOES:





Resources

Title	Link
Publications	
HIVE	https://pubmed.ncbi.nlm.nih.gov/26989153/
DVG-profiler	https://pubmed.ncbi.nlm.nih.gov/31100083/
HIVE-Hexahedron	https://pubmed.ncbi.nlm.nih.gov/28977510/
HexaHedron HIV Quasispecies	https://pubmed.ncbi.nlm.nih.gov/33055255/
Tools	
DVG-profiler	https://github.com/kkaragiannis/DVG-profiler
HIVE-HexaHedron	https://github.com/kkaragiannis/hexahedron
GWU-HIVE	https://hive.biochemistry.gwu.edu/home
FDA-HIVE	https://hive.fda.gov/