

Supplementary Information

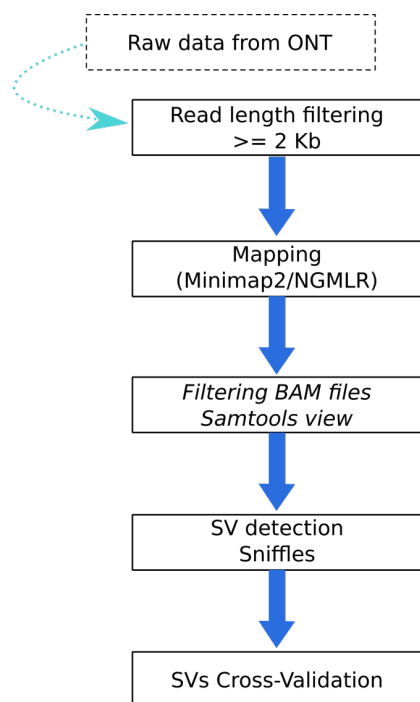


Figure S1. Overview of the bioinformatic workflow to detect SVs in P78 and P99 genomes.

Table S1 Main features of the obtained raw reads for each strain.

num_seq= number of reads, min_len= minimum length, avg_len= average length, max_len= maximum length

a supprimé: lenghht

Sample	num_seqs	min_len	avg_len	max_len	N50	Q20(%)	Q30(%)
P78_all_fast5_pass_basecalled_gup4.fastq	4900000	25	1080.7	91494	1598	59.92	18.65
P99_all_fast5_pass_basecalled_gup4.fastq	2293830	96	3764.5	279443	7643	57.25	17.94
P10_all.fastq	244000	91	1144.4	7819	2022	45.31	14.70
P20_all.fastq	484000	96	1091.9	7146	1938	45.44	14.74
P30_all.fastq	472000	85	1455.1	7519	2430	45.40	14.62
P40_all.fastq	612000	99	1003.4	6197	1635	44.62	14.33
P50_all.fastq	115135	110	1068.3	4596	1632	37.68	9.88
P70_all.fastq	65062	104	1057.3	4824	1614	37.53	9.90
P80_all.fastq	135441	99	1191.7	4592	1742	37.79	9.81
P90_all.fastq	95635	123	1502.9	7823	2624	38.18	10.10

Table S2 SVs detected by each mapper/caller combination for P78 and P99 strains

	P78		P99	
	Minimap2+Sniffles	NGMLR+Sniffles	Minimap2+Sniffles	NGMLR+Sniffles
Deletion	135	125	22	147
Duplication	49	19	2	30
Inversion	538	308	176	210
Insertion	9	8	10	10
Total	731	460	210*	397*

* P99 seems to accumulate complex SVs as Inversion-Duplications, which explains why the combination Minimap2+Sniffles detects fewer SVs than NGMLR+Sniffles while it is the opposite in P78 SVs are mainly Inversions and deletion. (see Sedlazeck et al. 2018 for more details on the mapping/detection algorithms).

Table S2 ORFs included within SVs distribution windows for P78 and P99. Only SVs supported by more than 0,1 % of the reads were considered.

Abbreviation : Perc_Svs_reads= percentage of reads covering the SV ; AF= Allelic frequency

Sample	Window	SVs	Breakpoints	Size(bp)	Perc_Svs_reads	AF	Impacted ORF	
P78	[10-40 kb]	Inv	28708-30175		1467	0.13	0.0013 ORF10/ORF11/ORF12	
		Inv	33593-34126		533	0.2	0.002 ORF17	
		Inv	33852-35784		1932	0.17	0.0017 ORF17/ORF18	
		Inv	34303-34560		257	0.24	0.0024 ORF17	
		Inv	34956-35604		648	0.33	0.0033 ORF18	
		Inv	36848-38291		1443	0.11	0.0011 ORF1/ORF19/ORF20	
		SSR(TG) _n	37855		32-309		0.66-1.3	0.0066-0.013 ORF19
		InvDup	38286-39138		852	0.35	0.0035 ORF20	
		Inv	69524-70145		621	0.22	0.0022 ORF33	
		Inv	70305-70796		491	0.33	0.0033 ORF33	
	[70-100 kb]	Inv	71209-75030		3821	0.1	0.001 ORF33	
		Inv	73197-80261		7064	0.12	0.0012 ORF44/ORF45/ORF46/ORF47	
		Inv	73469-73991		522	0.2	0.002 ORF43	
		Inv	74286-75045		759	0.5	0.005 ORF33	
		Inv	76475-76887		412	0.31	0.0031 ORF33	
		Inv	76868-77941		1073	0.32	0.0032 ORF44/ORF45/ORF46	
		Inv	77622-78273		651	0.17	0.0017 ORF33	
		Inv	77973-78931		958	0.28	0.0028 ORF33	
		Inv	78926-79064		138	0.16	0.0016 ORF33	
		Inv	79389-80054		665	0.38	0.0038 ORF33	
		Del	80796-82160		1364	0.12	0.0012 ORF33/ORF44/ORF45/ORF46/ORF47/ORF48	
		Inv	81818-84196		2378	0.11	0.0011 ORF33	
		Inv	82193-83393		1200	0.12	0.0012 ORF33	
		Ins	83040-83070		36	71	0.71	0.0071 ORF33/ORF44/ORF45/ORF46/ORF47/ORF48
		Inv	83337-84329		992	0.25	0.0025 ORF44/ORF45/ORF46/ORF47/ORF48	
		Del	89476-89769		293	0.22	0.0022 ORF49/ORF51	
		InvDup	90924-91243		319	0.46	0.0046 ORF49/ORF52	
		Del	91163-91209		46	0.63	0.0063 ORF49/ORF52	
		Del	91208-91292		84	1	1	1 ORF49/ORF52
		InvDup	95924-96465		541	0.79	0.0079 ORF55	
	Inv	95915-96506		591	0.41	0.0041 ORF55		
	Inv	99788-100692		904	0.33	0.0033 ORF57		
	[245-270 kb]	Inv	245797-246416		619	0.35	0.0035 ORF141/ORF142	
		Inv	250093-250915		822	0.56	0.0056 ORF145	
		Inv	251097-251832		735	0.12	0.0012 ORF145	
		Del	253614-255035		1421	0.1	0.001 ORF147/ORF148	
		Ins	256508-256512		36	0.3	0.003 ORF149	
		Inv	256887-257429		542	0.12	0.0012 ORF149	
		Inv	258154-250518		1363	0.44	0.0044 ORF150	
		Del	258153-259517		1363	100	1 ORF150	
Inv		259518-260584		1066	0.37	0.0037 ORF150/ORF151		
InvDup		261462-262073		611	0.67	0.0067 ORF151/ORF152		
P99	[10-20 kb]	Ins	11334		33	100	1 ORF6,	
		SSR(TG) _n	1650		32-48	100	1 ORF8, 1/ORF7,	
	[205-215 kb]	InvDup	211766-212162		396	0.44	0.0044 ORF118	
		Del	214421-214545		124	0.19	0.0019 ORF118/ORF119/ORF120/ORF121	

Table S3 SVs prevalence and type for P10, P20, P30, P40, P50, P70, P78, P80, P90 and P99.

Strain	Supporting_reads_SVs'	Perc_SVs_recs'	SVs	Ref_reads'	Perc_ref_reads'	Total_reads'	Breakpoints
P10	102	0.15	Del	66001	99.84	66103	258153-259517
P10	13	0.023	Inv	56244	99.97	56244	257913-258821
P20	58	0.047	Del	121869	99.97	56257	258153-259517
P20	10	0.01	Inv	93483	99.99	93493	257941-258545
P30	115	0.31	Del	36858	99.68	36973	258153-259517
P30	33	0.089	Inv	36858	99.91	36891	257918-258643
P30	30	0.076	Inv	39293	99.92	39323	258123-260721
P40	39	0.42	Del	9121	99.57	9160	258154-258272
P40	38	0.41	Inv	9114	99.58	9152	257726-259517
P40	47	0.53	Inv	8728	99.46	8775	258155-260208
P50	7468	88.97	Del	925	11.02	8393	258153-259517
P50	16	0.05	Inv	31749	99.94	31765	257893-260117
P50	525	14.34	Ins	3136	85.65	3661	258155
P70	4038	82.25	Del	871	17.74	4909	258153-259517
P70	322	16.01	Ins	1688	83.98	2010	258155
P78	6902	100	Del	0	0	6902	258154-259517
P78	30	0.44	Inv	6704	99.55	6734	258155-259518

P80	10838	30.68	Del	24487	69.31	35325	258153-259517
P80	21	0.18	Del	11196	99.81	11217	258155-259035
P80	18	.0.04	Inv	36044	99.95	36062	257815-259517
P80	12	0.03	Inv	34628	99.96	34640	258076-261050
P80	761	6.79	Ins	10445	93.2	11206	258155
P90	88	0.31	Del	27560	99.68	27646	258153-259517
P90	15	0.054	InvDup	27273	99.94	27288	258386-259486
P99*	-	0	-	-	100	-	-

*At given breakpoints:

Supporting_reads_SVs and Perc_SVs_recs correspond to the number of reads covering a given SV and their percentage.

Ref_reads and Perc_ref_reads correspond to the number of references like reads and their percentage.

Total_reads correspond to the total number of reads covering the breakpoints region.

+As no SVs were detected within the targeted region in P99, 100 % of the mapped reads are reference-like.

a mis en forme : Anglais (E.U.)

a supprimé: s

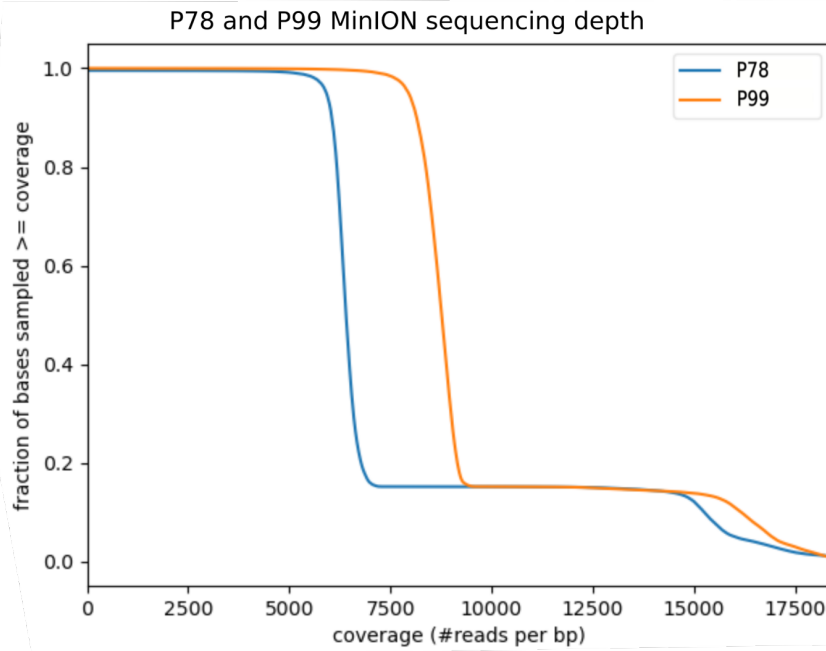


Figure S2. Sequencing depth for KHVP78 and KHVP99 samples. This graph shows how the bases are covered and how many times. For example, 100% of the sampled bp from the KHVP99 genome have at least 7500 overlapping reads and around 10% have at least 15000 overlapping reads.