Case Study 2:
NGS for molecular consistency monitoring of biopharmaceuticals

Konstantin Chumakov

George Washington University

Presentation plan

- Rationale for genetic consistency testing
- History of Oral Polio Vaccine
- Monkey test, PVR-transgenic mouse test
- Molecular Methods
  - MAPREC
  - Next Generation (Deep, High-throughput) sequencing
- Comparing whole-genome SNP profiles
  - Pass-fail decisions

Why do we need to monitor molecular (genetic) consistency?

- Live viral vaccines inevitably mutate during virus growth in cell cultures
- Attenuation reduces virus fitness
  - Mutations that lead to the loss of attenuation have selective advantage
  - Growth in inappropriate conditions can increase virus virulence
  - Mutations can occur in protective epitopes, reducing vaccine efficacy
- Manufacturing consistency is a cornerstone of cGMP
- Ideally, a new vaccine lot should have the same molecular composition as previously released ones
- Breach of molecular consistency is a warning sign calling for investigation
- Whole-genome molecular profiling is ideally suited for this task

Oral Polio Vaccine is the most striking example

Quality of other live vaccines and gene therapy vectors could also benefit from higher molecular consistency

Dr Albert Sabin
1904 - 1993

Oral Polio Vaccine (OPV)

- Weakened “attenuated” virus
- Selected from the pre-exiting attenuated variants within wild-type stocks
- Natural route of administration
- Comprehensive immunity
- “Herd” effect through transmission to contacts

Starting from the early 1960s used throughout the world

(except in Finland, Sweden, and Netherlands)

PROPERTIES OF ATTENUATED POLIOVIRUSES AND THEIR BEHAVIOR IN HUMAN BEINGS*

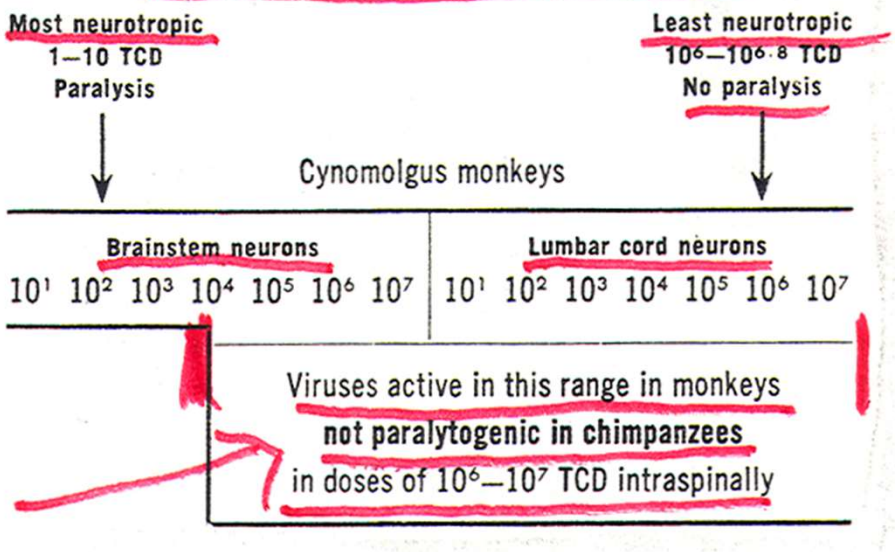
BY

ALBERT B. SABIN

Children's Hospital Research Foundation, University of Cincinnati College of Medicine, Cincinnati, Ohio

TABLE 1

NEUROTROPIC SPECTRUM OF KNOWN POLIOVIRUSES IN RELATION TO DIFFERENT PRIMATE NEURONS



Kostya
– please note
heterogeneous population in Leon KP34 as demonstrated by marked differences in spinal neurovirulence of individual plaque progeny in highly diluted cultures

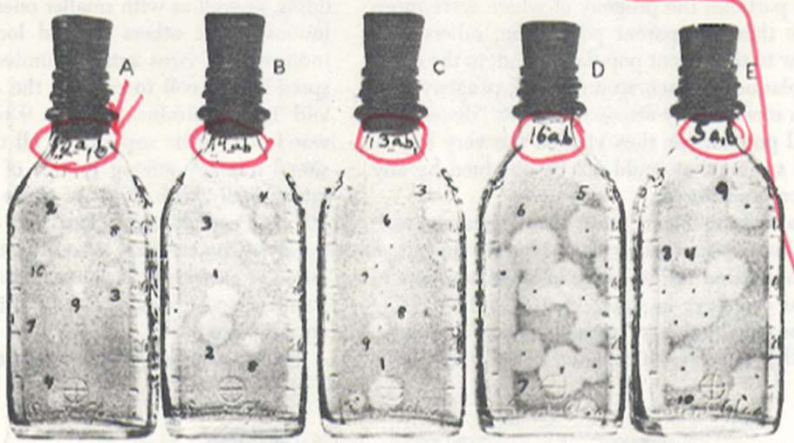
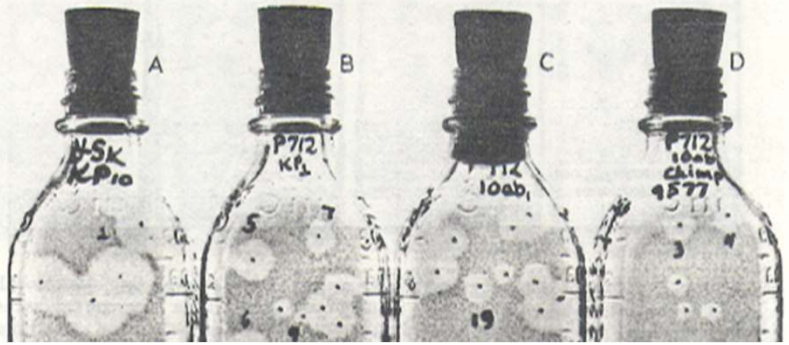


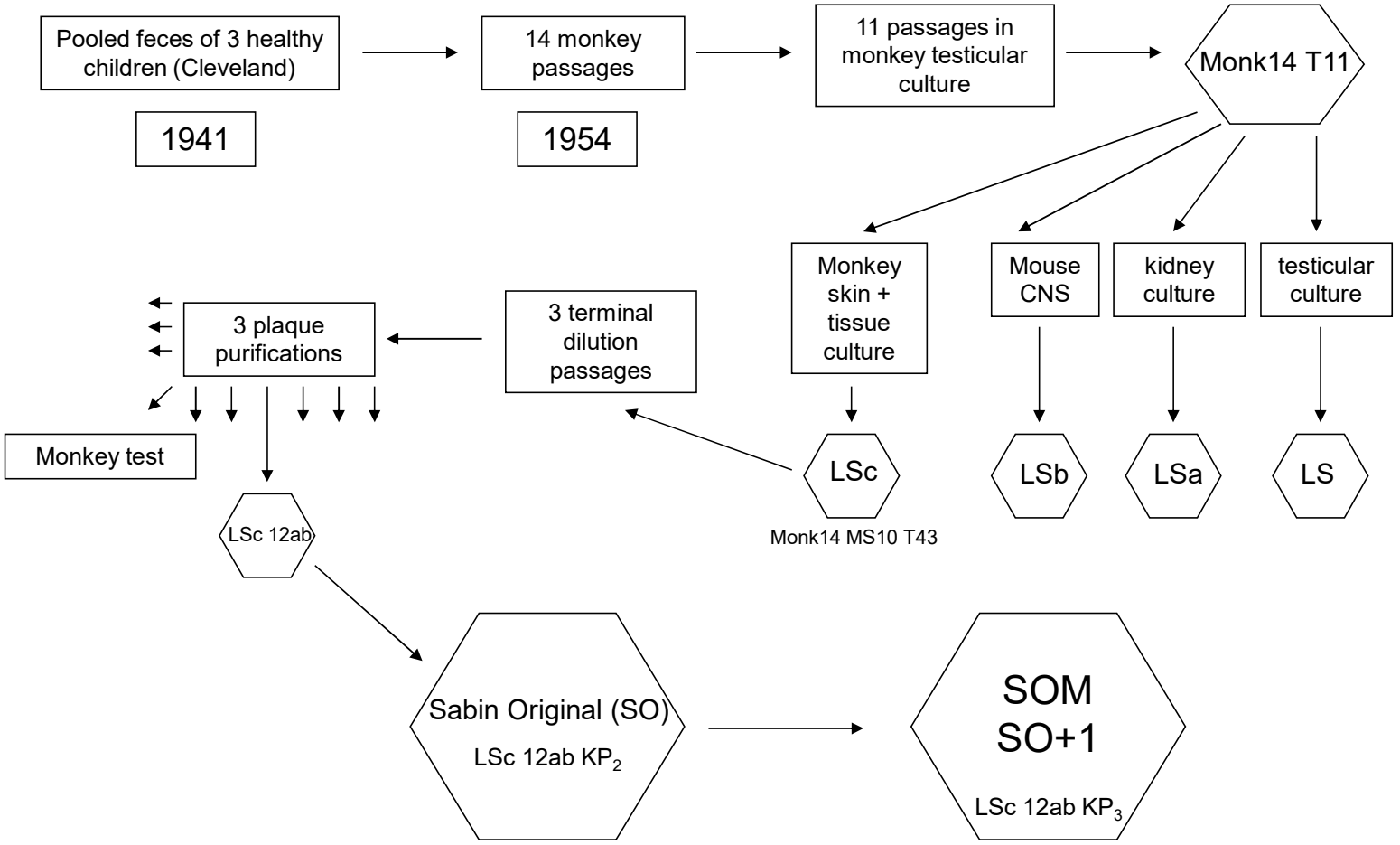
FIGURE 2. The progeny of five different triply purified plaques derived from attenuated Type 3, Leon KP 34 virus previously purified by three terminal dilution passages. All cultures were grown on the same lot of monkey kidney cells and were photographed six days after inoculation. The results of spinal tests in monkeys were as follows:

Dilution of culture fluid	Incidence of paralysis with indicated stock				
	A	B	C	D	E
Undiluted	0/10	0/10	3/5	3/5	5/5
10^{-1}	0/10	0/10	2/5	3/5	4/5

!!



- Virulent particles generate larger plaques than attenuated particles



October 10, 1956

Sabin, A. B., and R. L. Boulger. 1973. History of Sabin attenuated poliovirus oral live vaccine strains. J. Biol. Stand. **1:**115–118.

What goes up must come down

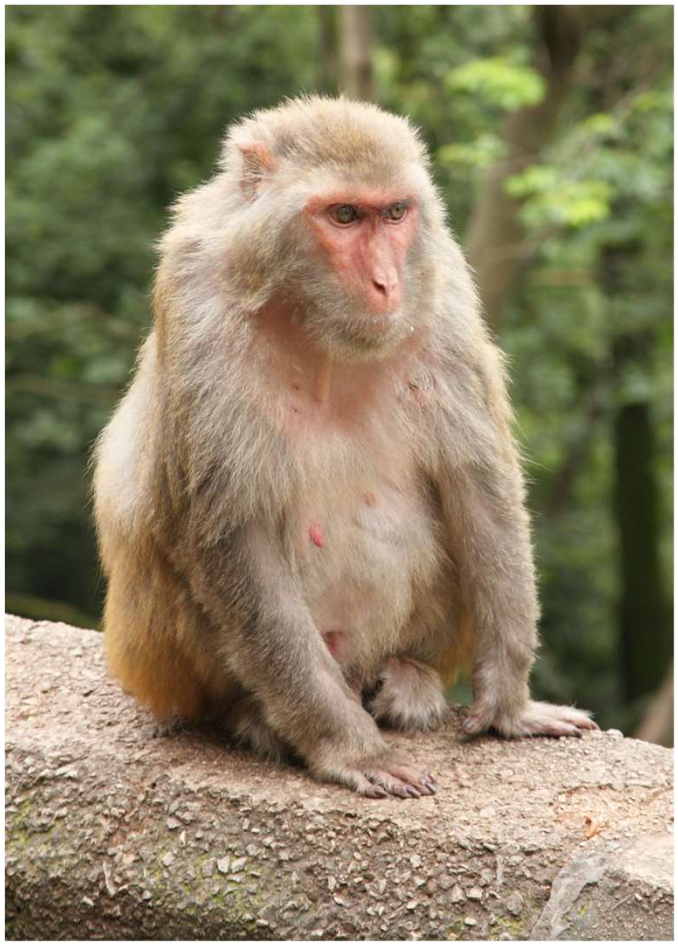
Passage in cell culture and in the gut of vaccine recipients leads to the loss of attenuation

Dr. Sabin required vaccine manufacturers to test every batch of vaccine for neurovirulence in monkeys

WHO Monkey neurovirulence test

To measure residual virulence of Sabin strains

- 24*2=48 monkeys inoculated intraspinally
  - 24 with new vaccine lot and 24 with reference
- Observed for 17 days for signs of paralysis
- All monkeys sacrificed for histological examination
- Lesions in CNS are scored and compared
- Vaccine lot “passes” if lesions are not greater than in reference vaccine
- ~200 monkeys were killed to QC one lot of trivalent vaccine



Monkey neurovirulence test is a product consistency test

- There is no evidence that failure of MNVT leads to unsafe vaccine
- However, it indicates a breach in manufacturing consistency and drift of vaccine virus in the direction of higher neurotropism

- MNVT often yields variable results, is very expensive, takes a lot of time, requires specialized expertise, and is inhumane
- Therefore, there was a strong push to find a surrogate test that could replace MNVT
  - Currently there is an alternative neurovirulence test based on transgenic mice

Biologicals
Volume 24, Issue 2, June 1996, Pages 77-86

Regular article

A Poliovirus-susceptible Transgenic Mouse Model as a Possible Replacement for the Monkey NeurovirulenceTest of Oral Poliovirus Vaccine

Dragunsky E. [a] [f1], Taffs R. [a], Chernokhvostova Y. [a], Nomura T. [b], Hioki K. [b], Gardner D. [a], Norwood L. [a], Levenbook I. [a]

Increased neurovirulence associated with a single nucleotide change in a noncoding region of the Sabin type 3 poliovaccine genome

D. M. A. Evans*, G. Dunn*, P. D. Minor*, G. C. Schild*, A. J. Cann†, G. Stanway†, J. W. Almond†, K. Currey‡ & J. V. Maizel Jr‡

* National Institute for Biological Standards and Control, Holly Hill, Hampstead, London NW3 6RB, UK
† University of Leicester, University Road, Leicester LE1 7RH, UK
‡ Laboratory of Molecular Genetics, National Institutes of Health, Bethesda, Maryland 20205, USA

Most of the small number of cases of poliomyelitis which occur in countries where Sabin's attenuated poliovirus vaccines are used are temporally associated with administration of vaccine and involve polioviruses of types 2 and 3 (ref. 1). Recent studies have provided convincing evidence that the Sabin type 2 and 3 viruses themselves may revert to a neurovirulent phenotype on passage in man[2–6]. We report here that a point mutation in the 5′ noncoding region of the genome of the poliovirus type 3 vaccine consistently reverts to wild type in strains isolated from cases of vaccine-associated poliomyelitis. Virus with this change is rapidly selected on passage through the human gastrointestinal tract. The change is associated with a demonstrable increase in the neurovirulence of the virus.

Table 1 Base at position 472, time of isolation, neurovirulence and temperature sensitivity of Sabin type 3 vaccine-derived strains of poliovirus

Virus	Base at position 472	Time of isolation after vaccination	Mean histological lesion score	rct marker test*
Sabin vaccine	U		0.36	>5.5(rct⁻)
DM1	U	24 h	ND	ND
DM2	U	31 h	1.58	6.13 (rct⁻)
DM3	U/C	35 h	ND	ND
DM4†	C	47 h	2.48	5.71 (rct⁻)
DM38	C	18 days	ND	ND
DM119	C	3–4 weeks	3.34	0.25 (rct⁺)

Mean histological lesion scores were determined using the standard WHO neurovirulence test[14]. The range of mean histological lesion scores of a type III attenuated refer... strain in eight tests carried out over the past 2 yr is 0.43–0.79

ND, not determ

* The rct ma... logarithmic diff... showing reduct... sensitive (rct⁻).

† The sequen... DM4 were dete... was sequenced... RNA, annealin... by the dideoxy... polymerase I[13].

We therefor... to cytidine a... ated cases ... phenotype o...

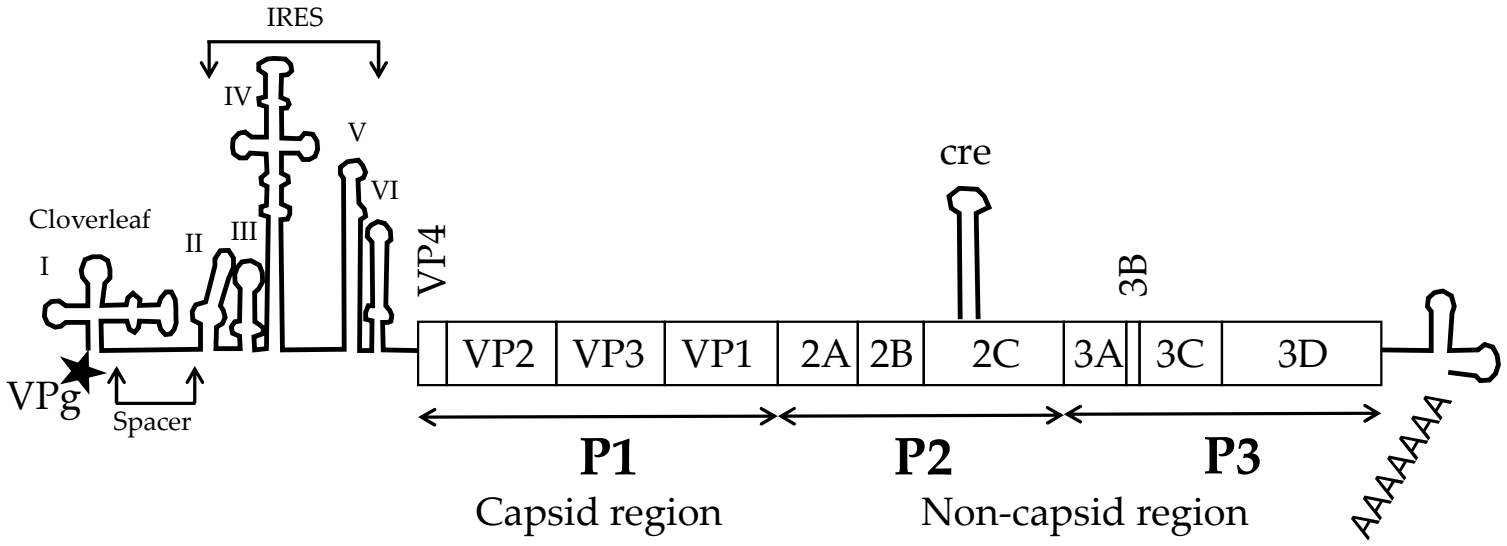
Twelve str...

Table 2 Base at position 472 of the poliovirus genome in primary vaccinees

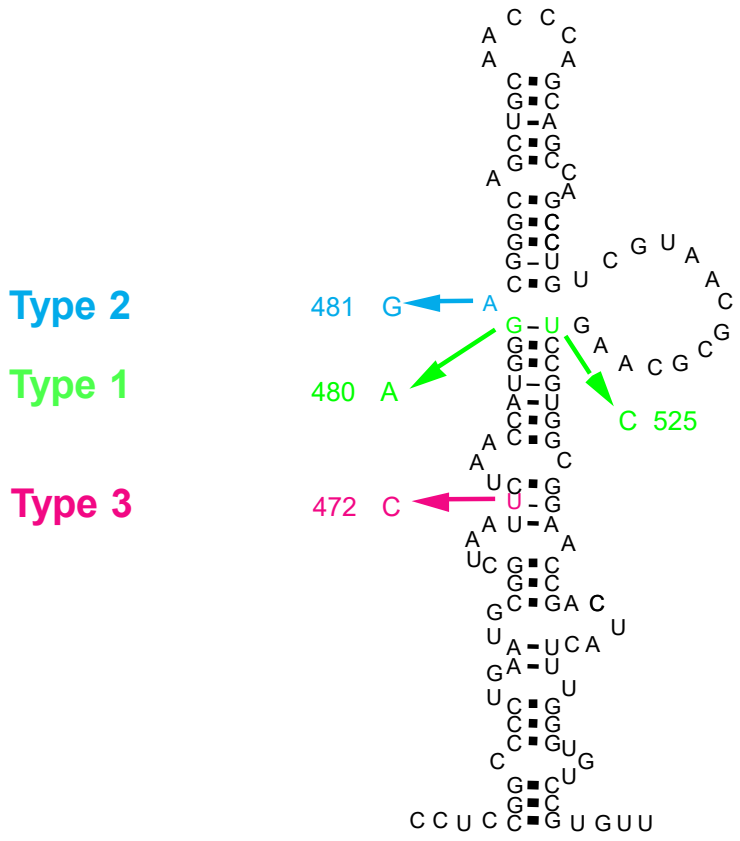
Day post-vaccination	Vaccinee KT1	KT2	KT3
1	U	*	U
2	U/C	*	U
3	C	*	C
4	C	C	C

Faecal samples were taken daily from three vaccinated infants less than 1 yr old, who received vaccine of the same origin as DM (Table 1).
* Isolates of poliovirus type 3 not available.

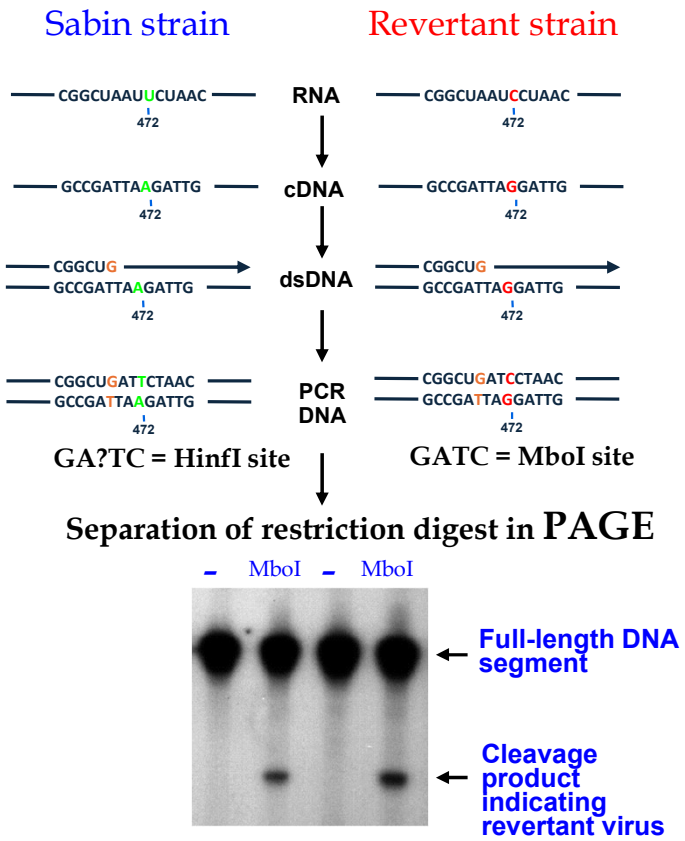
Genome Structure of Enteroviruses



Neurovirulent mutations in domain V of the IRES element



MAPREC assay for neurovirulent revertants in type 3 OPV



Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 199–203, January 1991
Medical Sciences

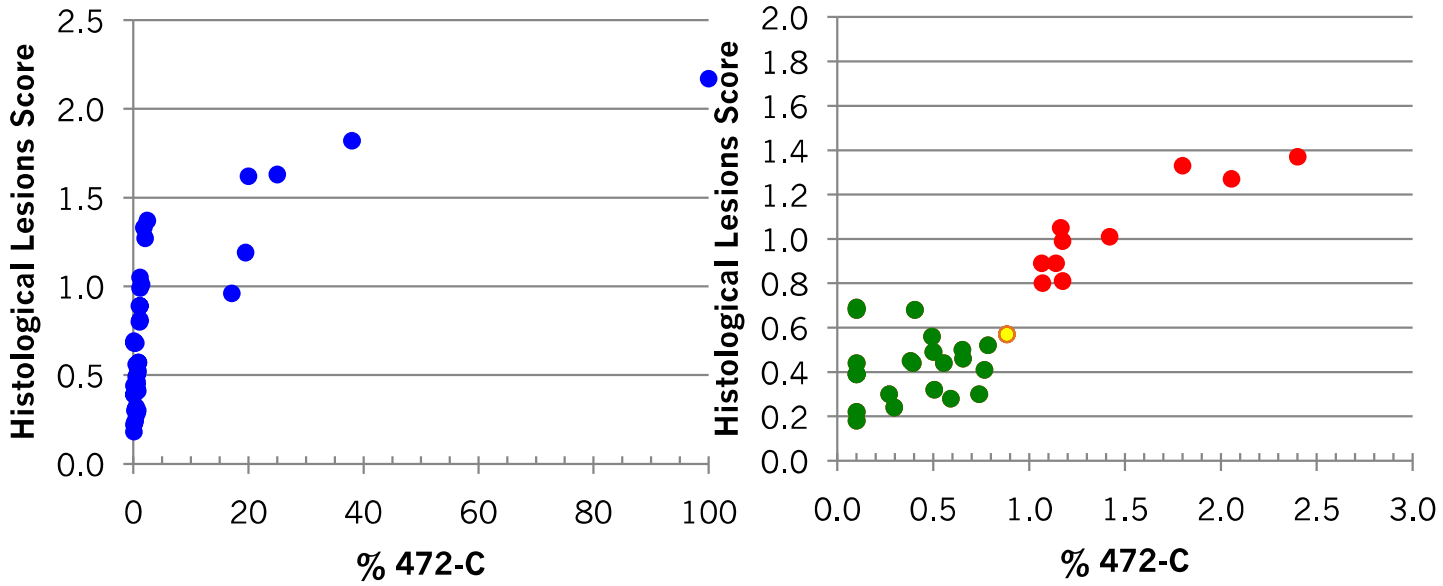
Correlation between amount of virus with altered nucleotide sequence and the monkey test for acceptability of oral poliovirus vaccine

(attenuation/type 3 poliovirus/polymerase chain reaction/restriction enzyme analysis)

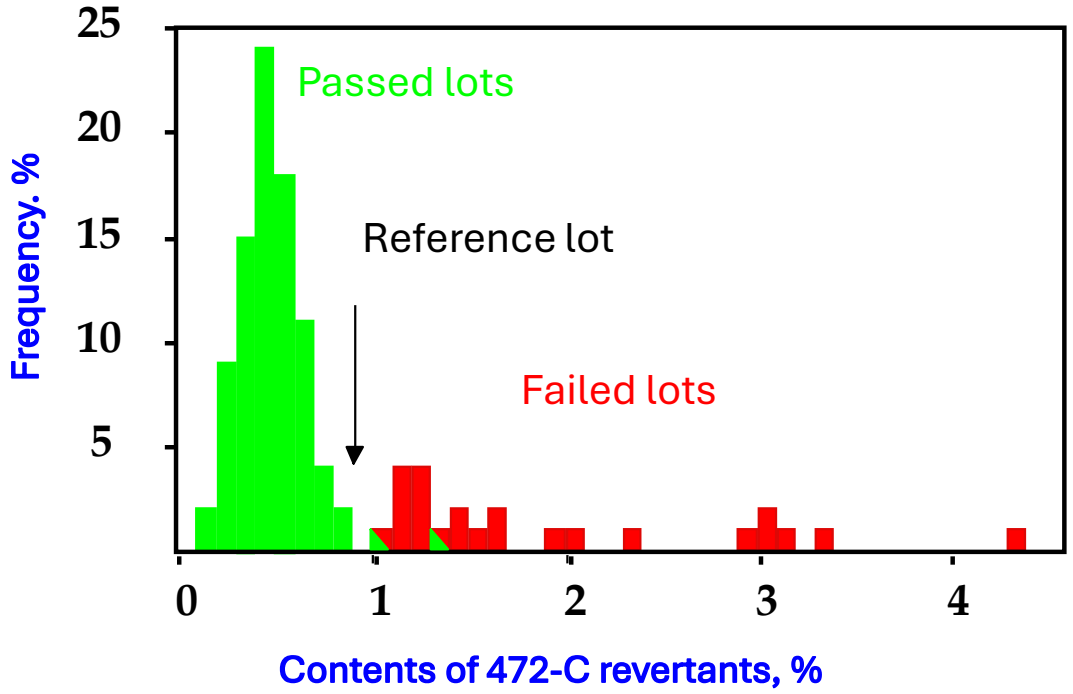
KONSTANTIN M. CHUMAKOV*, LAURIE B. POWERS*, KEVIN E. NOONAN†, IGOR B. RONINSON†, AND INESSA S. LEVENBOOK*

*Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, MD 20892; and †Department of Genetics, University of Illinois at Chicago, Chicago, IL 60612

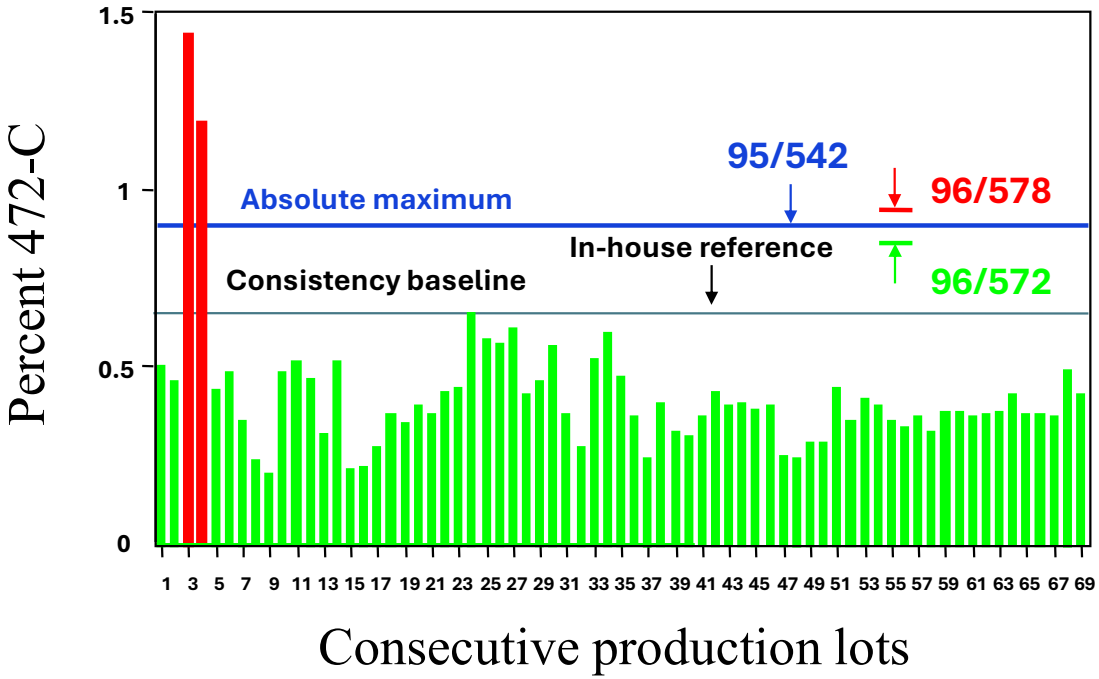
Communicated by Albert B. Sabin, October 10, 1990 (received for review August 16, 1990)



Contents of 472-C revertants in lots of type 3 OPV



Pass-Fail Decision



Regulatory role of MAPREC

- An International Collaborative Studies on MAPREC tests for all three serotypes of OPV were conducted in the 1990s
- WHO Expert Committee on Biological Standardization (ECBS) approved MAPREC as an *in vitro* test of preference for lot release of OPV
- WHO recommendation for manufacture and control of OPV recommend MAPREC in combination with monkey or Tg-mouse neurovirulence test
- If MAPREC is performed rct_{40} marker test can be omitted

Why do we need an alternative to MAPREC?

- MAPREC tests only one genomic position
  - A method testing for all potential mutations would be preferable
- MAPREC test requires a highly skilled personnel and specialized equipment
- MAPREC requires the use of radioactive isotopes
  - An alternative protocol based on fluorescent dyes is available but has a lower dynamic range
- Some labs experience over time an unexplained baseline drift defined by reference materials

www.pnas.org/cgi/doi/10.1073/pnas.1012537107 PNAS | November 16, 2010 | vol. 107 | no. 46 | 20063–20068

Massively parallel sequencing for monitoring genetic consistency and quality control of live viral vaccines

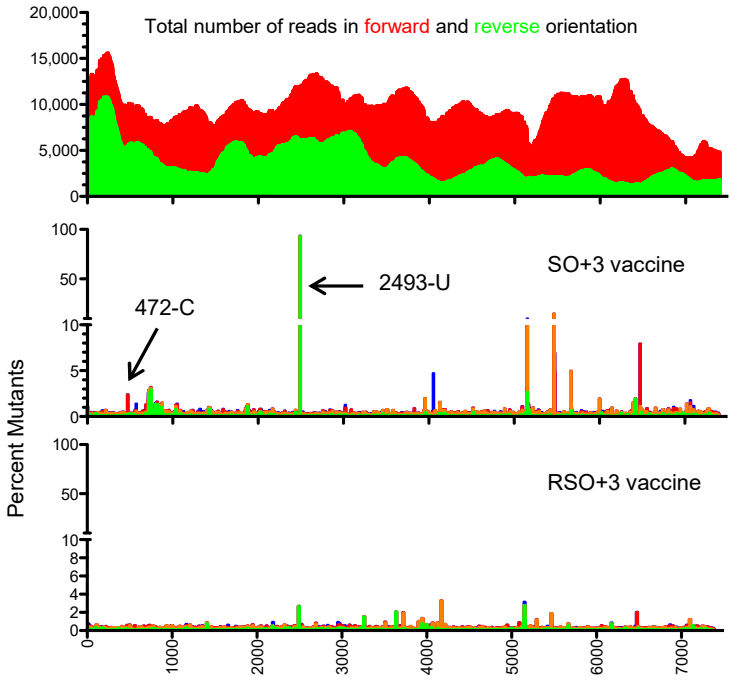
Alexander Neverov and Konstantin Chumakov[1]

Center for Biologics Evaluation and Research, Food and Drug Administration, Rockville, MD 20852

Edited* by Robert H. Purcell, National Institutes of Health, Bethesda, MD, and approved October 6, 2010 (received for review August 24, 2010)

Intrinsic genetic instability of RNA viruses may lead to the accumulation of revertants during manufacture of live viral vaccines, requiring rigorous quality control to ensure vaccine safety. Each

uation. MAPREC is currently recommended by the World Health Organization (WHO) for screening of batches of OPV before they can be released for use in humans (11, 18).

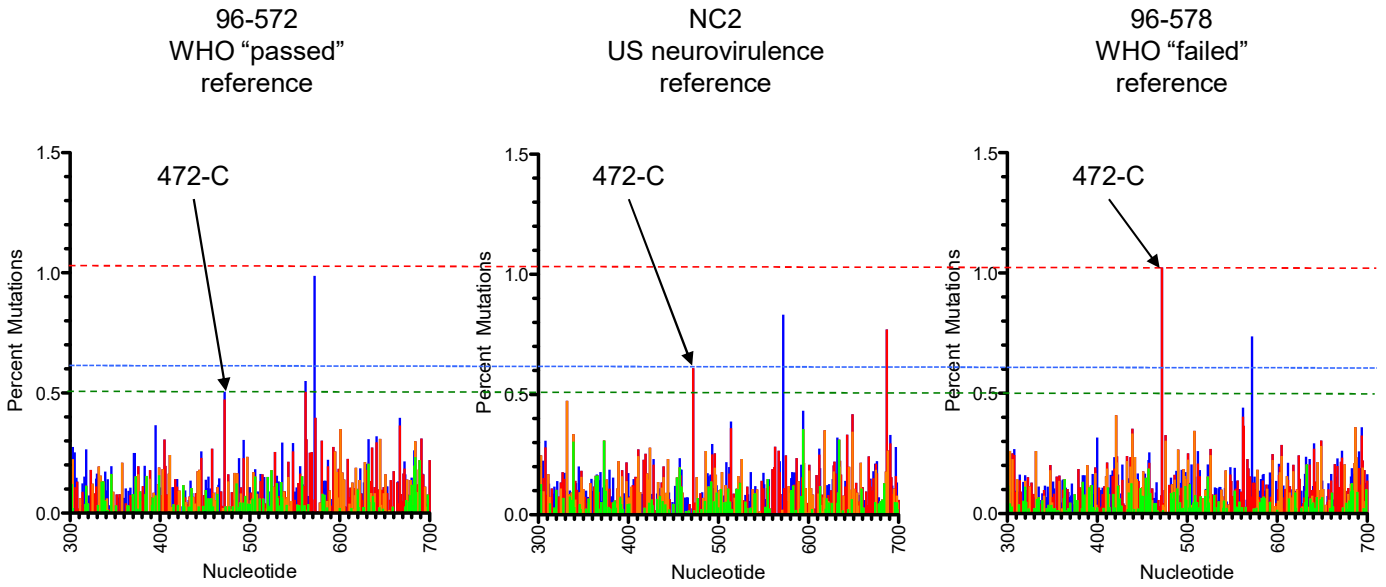


Massively Parallel =

High Throughput =

Next Generation

HTS data for type 3 OPV references

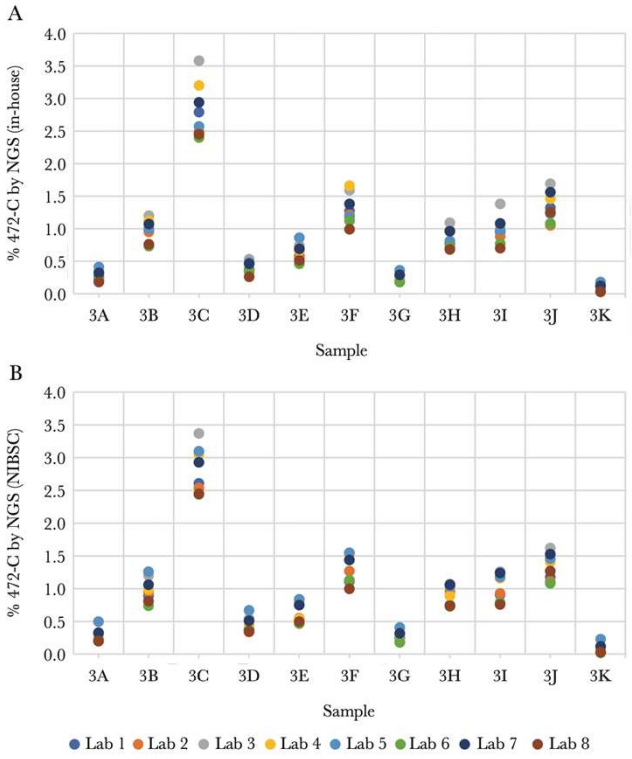


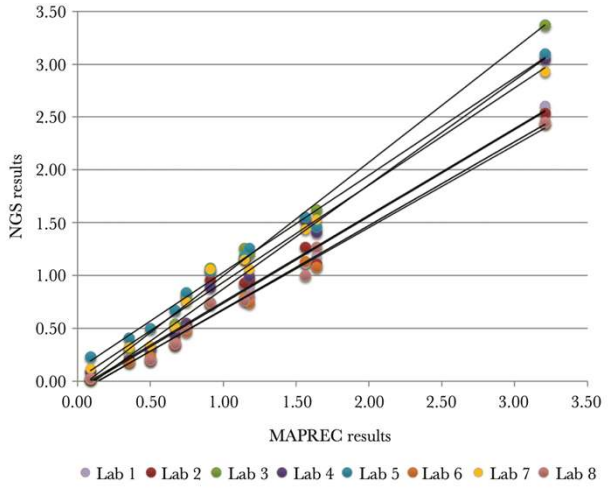
The Journal of Infectious Diseases

ACADEMICSUBJECTS/MED00860

The Use of Next-Generation Sequencing for the Quality Control of Live-Attenuated Polio Vaccines

Bethany Charlton,[1] Jason Hockley,[2] Majid Laassri,[3] Thomas Wilton,[1] Laura Crawt,[1] Mark Preston,[4] NGS Study Group, Peter Rigsby,[2] Konstantin Chumakov,[3] and Javier Martin[1]





NGS of OPV lots can be interpreted in two ways

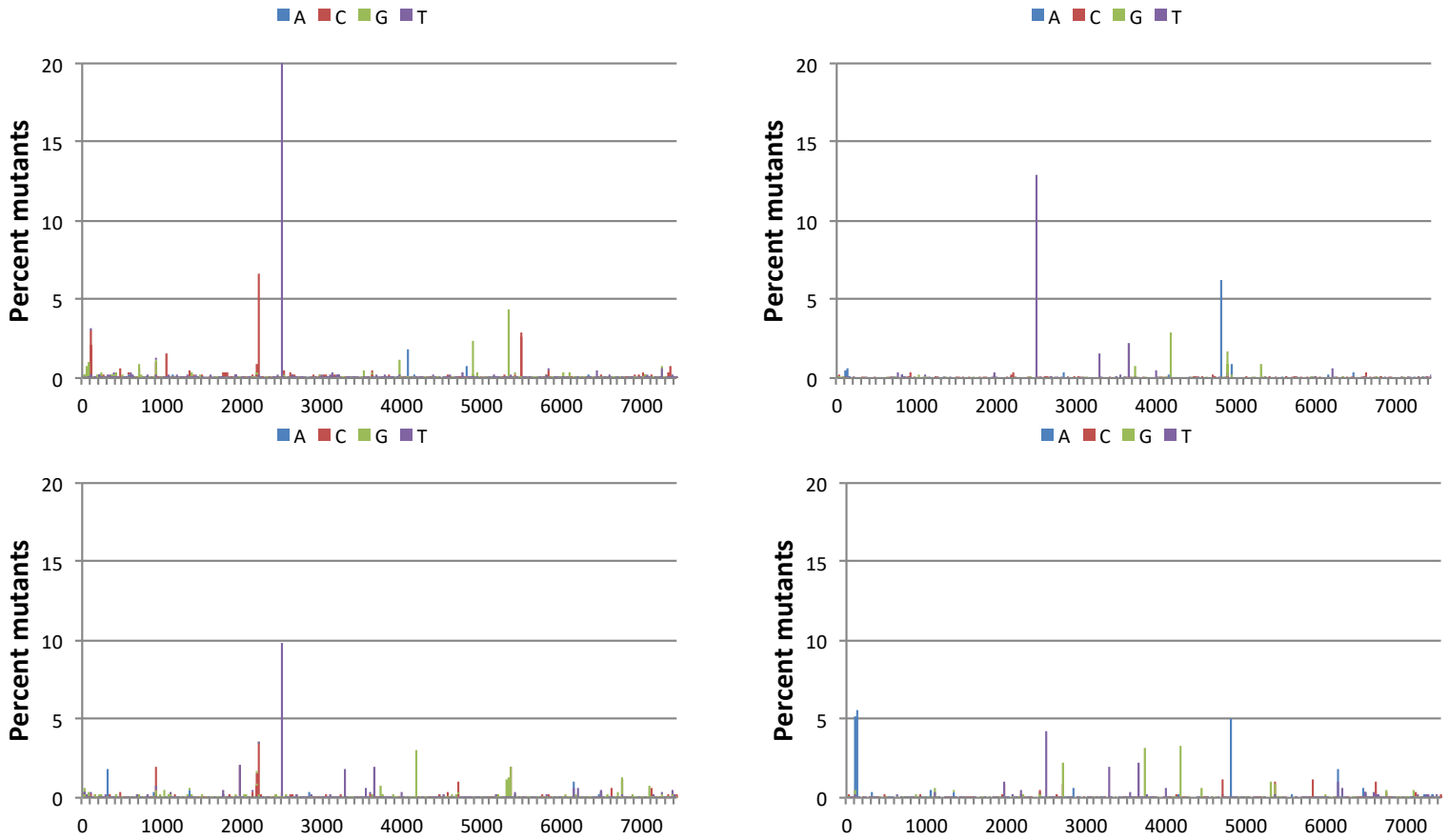
- Quantitative analysis of 5’-UTR mutations
  - An alternative to MAPREC
  - ECBS approved the use of NGS as a replacement of MAPREC
- Whole-genome SNP profiling
  - Comparison of the entire swarm of all possible mutants in vaccine lots
  - The ultimate measure of consistency
  - Replacement of animal neurovirulence testing

What is the purpose of animal testing?

- Monkey and Tg-mouse tests are NOT safety tests, but rather consistency tests
- Contents of 5'-UTR mutations predicts the results of MNVT
- If a lot passes MAPREC or HTS test for these mutations, then the only reason for animal testing is to ensure consistency of manufacture
- Mutational profiles of vaccine virus are more informative than MNVT
  - MNVT measures neurovirulence only, which is easier to test by HTS
  - HTS can detect any breach of manufacturing consistency, including potential changes in antigenic characteristics
  - HTS can also serve as an identity test

ECBS agreed that after appropriate validation whole-genome HTS could be considered as a replacement of animal testing

SNP profiles of OPV3 made form different seed viruses



Unique pattern of mutations in vaccines made by different manufacturers

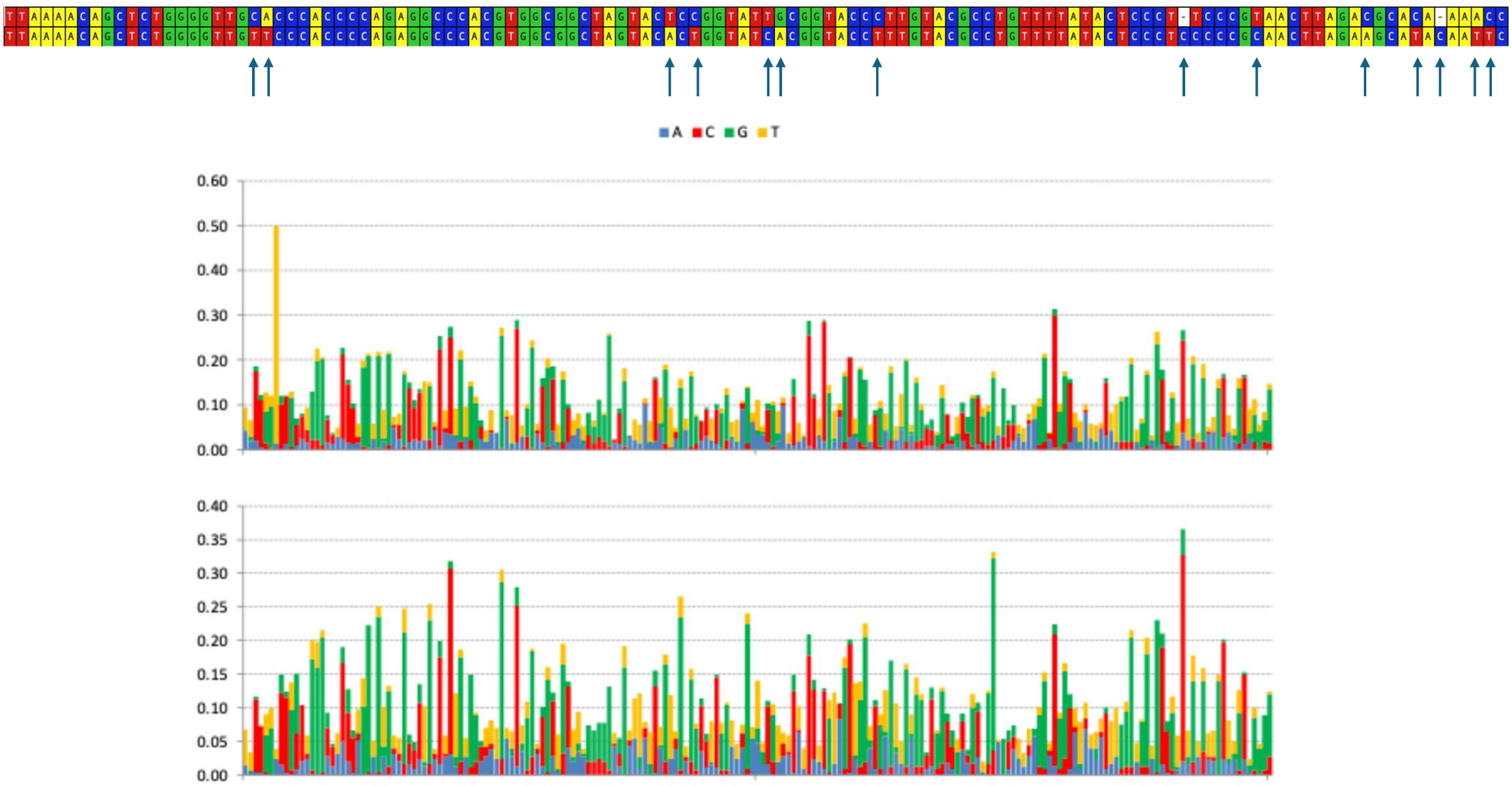
Nucleotide, Protein, Amino Acid					A	B	C
2189	A->C	VP3	143	Thr->Pro	4.7 ± 1.16	0.4 ± 0.09	0.1 ± 0.10
2493	C->T	VP1	6	Thr->Ile	5.2 ± 2.96	13.0 ± 2.84	3.5 ± 1.48
2696	A->G	VP1	74	Thr->Ala	0.6 ± 0.22	0.1 ± 0.01	6.0 ± 3.43
3723	A->G	2A	116	Gln->Arg	1.4 ± 0.26	1.2 ± 0.15	6.2 ± 2.67
4696	G->C	2C	194	Val	0.7 ± 0.50	0.4 ± 0.14	2.9 ± 0.99
4791	G->A	2C	226	Arg->His	0.5 ± 0.51	5.6 ± 0.69	1.0 ± 2.24
5323	T->G	3A	74	Gly	2.4 ± 1.68	0.2 ± 0.02	0.0 ± 0.02

Summary of SNP profiles of type 3 samples

Genomic positions with <1% mutations excluded

Nucleotide	Base	Mutation	Gene	Amino acid	Reference amino acid	Mutant amino acid	O-1	O-2	O-3	O-4	O-5	P-1	P-2	P-3	P-4	P-5	Q-1	Q-2	Q-3	Q-4	Q-5	R-1	R-2	R-3	R-4	R-5	S-1	S-2	S-3	S-4	S-5	T-1	T-2	T-3	T-4	T-5	U-1	U-2	U-3	U-4	U-5	V-1	V-2	V-3	V-4	V-5	W-1	W-2	W-3	W-4	W-5	X-1	X-2	X-3	X-4	X-5	Y-1	Y-2	Y-3	Y-4	Y-5
472	T	C	non-coding				0.03	0.04	0.03	0.05	0.03	0.92	1.09	1.07	1.00	1.00	0.21	0.22	0.31	0.33	0.29	0.85	0.79	0.77	0.86	0.76	0.34	0.50	0.52	0.45	0.53	2.43	2.07	2.59	2.19	2.44	1.08	1.10	1.12	1.14	1.06	1.21	1.03	1.30	1.12	1.06	0.61	0.68	0.72	0.68	0.73	1.02	0.98	0.99	0.87	0.90	0.02	0.06	0.04	0.03	0.03
537	G	A	non-coding				0.02	0.02	0.04	0.02	0.02	0.08	0.09	0.07	0.13	0.08	8.94	9.24	8.97	9.07	9.45	0.09	0.07	0.07	0.05	0.07	0.09	0.08	0.07	0.10	0.08	0.15	0.19	0.12	0.20	0.16	0.07	0.05	0.05	0.06	0.05	0.09	0.07	0.08	0.09	0.09	0.04	0.04	0.08	0.06	0.05	0.13	0.15	0.09	0.18	0.13	0.04	0.04	0.02	0.05	0.03
687	A	C	non-coding				0.01	0.02	0.02	0.01	0.01	0.01	0.03	0.01	0.01	0.02	0.01	0.00	0.02	0.01	0.00	0.00	0.02	0.00	0.00	0.01	0.01	0.02	0.00	0.01	0.01	1.02	0.88	1.10	0.98	1.22	0.02	0.01	0.01	0.01	0.01	0.87	0.86	0.88	0.92	0.94	0.02	0.00	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.03	0.02	0.00	0.01	0.02
699	C	T	non-coding				0.03	0.04	0.01	0.02	0.04	0.04	0.09	0.02	0.04	0.07	1.51	1.55	1.68	1.55	1.63	0.03	0.05	0.05	0.07	0.05	0.06	0.04	0.06	0.04	0.03	0.04	0.02	0.05	0.04	0.06	0.03	0.02	0.05	0.04	0.04	0.04	0.03	0.02	0.03	0.05	0.02	0.02	0.04	0.05	0.04	0.08	0.04	0.07	0.01	0.03	0.07	0.02	0.05
713	A	G	non-coding				0.17	0.16	0.13	0.13	0.13	0.19	0.17	0.15	0.17	0.12	0.12	0.12	0.13	0.10	0.16	0.15	0.11	0.09	0.12	0.09	0.10	0.15	0.13	0.08	0.14	1.66	1.63	1.72	1.51	1.76	0.12	0.12	0.14	0.13	0.13	1.60	1.60	1.84	1.77	1.79	0.09	0.14	0.11	0.10	0.10	0.19	0.15	0.14	0.10	0.11	0.11	0.13	0.24	0.13	0.08
769	A	G	protein VP4	9	Lys		0.17	0.17	0.15	0.09	0.20	1.50	1.64	1.54	1.58	1.63	0.13	0.20	0.19	0.19	0.14	0.73	0.85	0.68	0.72	0.71	0.80	0.77	0.61	0.83	0.83	0.15	0.16	0.14	0.21	0.16	0.14	0.15	0.13	0.15	0.10	0.12	0.11	0.15	0.14	0.09	0.16	0.15	0.12	0.09	0.12	1.53	1.64	1.42	1.47	1.43	0.13	0.15	0.20	0.13	0.12
858	A	G	protein VP4	39	Asn	Ser	0.16	0.15	0.15	0.09	0.14	0.15	0.12	0.07	0.13	0.11	2.22	2.18	2.00	1.98	1.92	0.08	0.13	0.08	0.13	0.12	0.15	0.10	0.14	0.13	0.05	0.11	0.11	0.10	0.11	0.09	0.11	0.11	0.13	0.12	0.15	0.13	0.13	0.12	0.15	0.17	0.12	0.09	0.11	0.06	0.08	0.16	0.16	0.12	0.11	0.14	0.08	0.09	0.06	0.11	0.11
898	A	G	protein VP4	52	Lys		0.13	0.10	0.10	0.07	0.15	0.12	0.16	0.18	0.12	0.09	9.89	9.85	10.08	9.42	10.03	0.09	0.05	0.10	0.06	0.06	0.06	0.08	0.05	0.08	0.07	0.09	0.07	0.10	0.06	0.12	0.08	0.06	0.07	0.09	0.09	0.09	0.10	0.12	0.07	0.14	0.07	0.10	0.09	0.07	0.05	0.20	0.11	0.12	0.08	0.12	0.08	0.09	0.06	0.08	0.07
1003	A	G	protein VP2	18	Leu		0.09	0.06	0.06	0.08	0.06	1.93	1.93	1.82	1.96	1.95	0.06	0.15	0.14	0.11	0.12	0.97	1.03	1.03	1.01	1.04	1.08	1.06	0.89	1.07	0.97	0.05	0.09	0.06	0.06	0.06	0.46	0.39	0.42	0.41	0.42	0.15	0.08	0.11	0.06	0.06	0.43	0.41	0.38	0.45	0.39	2.05	1.76	1.90	1.93	1.73	0.07	0.09	0.05	0.08	0.11
1523	T	C	protein VP2	192	Tyr	His	0.05	0.06	0.03	0.03	0.07	0.04	0.04	0.05	0.06	0.04	0.98	0.93	0.95	1.01	0.94	0.03	0.04	0.05	0.04	0.04	0.05	0.04	0.04	0.03	0.02	0.06	0.07	0.07	0.02	0.03	0.04	0.05	0.02	0.03	0.02	0.02	0.05	0.03	0.04	0.03	0.03	0.06	0.01	0.02	0.03	0.04	0.04	0.04	0.03	0.02	0.04	0.05	0.05	0.03	0.04
1537	C	T	protein VP2	196	Ile		0.05	0.05	0.02	0.04	0.04	0.07	0.04	0.06	0.09	0.05	1.14	1.17	1.21	1.20	1.18	0.08	0.08	0.10	0.06	0.10	0.13	0.07	0.11	0.10	0.07	0.04	0.04	0.06	0.05	0.04	0.05	0.04	0.06	0.08	0.05	0.04	0.06	0.07	0.05	0.08	0.03	0.05	0.06	0.06	0.08	0.10	0.12	0.12	0.10	0.03	0.03	0.03	0.03	0.05
1867	T	C	protein VP3	35	Asp		0.13	0.10	0.11	0.13	0.15	1.16	1.27	1.23	1.27	1.26	0.17	0.11	0.13	0.20	0.21	0.62	0.58	0.60	0.69	0.70	0.67	0.76	0.75	0.65	0.78	0.14	0.13	0.19	0.13	0.17	0.12	0.16	0.12	0.11	0.14	0.18	0.23	0.13	0.18	0.22	0.15	0.14	0.13	0.14	0.14	1.42	1.42	1.43	1.38	1.39	0.10	0.12	0.10	0.08	0.11
2080	T	C	protein VP3	106	Tyr		0.04	0.07	0.05	0.04	0.05	0.08	0.07	0.06	0.08	0.08	1.03	1.14	1.17	1.02	1.10	0.05	0.04	0.04	0.07	0.04	0.03	0.04	0.04	0.05	0.05	0.05	0.05	0.08	0.07	0.06	0.05	0.06	0.05	0.04	0.06	0.08	0.07	0.08	0.06	0.06	0.05	0.06	0.04	0.06	0.06	0.08	0.08	0.09	0.07	0.10	0.07	0.05	0.06	0.05	0.04
2440	A	T	protein VP3	226	Arg		0.01	0.03	0.01	0.03	0.02	0.00	0.01	0.01	0.01	0.01	99.80	99.81	99.82	99.79	99.80	0.01	0.03	0.02	0.02	0.01	0.01	0.02	0.02	0.00	0.02	0.45	0.38	0.47	0.43	0.48	0.01	0.02	0.01	0.01	0.01	0.39	0.42	0.43	0.37	0.42	0.01	0.02	0.01	0.01	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.01	0.02	0.03	0.03
2493	C	T	protein VP1	6	Thr	Ile	99.95	99.91	99.93	99.95	99.92	75.38	75.26	75.55	75.33	74.97	99.91	99.94	99.90	99.90	99.93	47.02	47.05	46.74	46.64	46.01	44.72	44.06	44.20	44.35	43.72	94.66	94.70	94.67	94.72	94.36	21.58	21.68	21.30	21.23	20.97	92.47	92.92	92.69	92.92	92.57	24.84	24.99	24.80	24.43	25.02	74.88	75.34	75.08	75.27	75.23	99.95	99.92	99.90	99.93	99.94
2731	C	T	protein VP1	85	Val		0.05	0.03	0.09	0.04	0.06	0.07	0.08	0.09	0.16	0.09	0.05	0.04	0.07	0.06	0.06	0.07	0.09	0.09	0.13	0.10	0.09	0.06	0.10	0.09	0.08	0.06	0.06	0.06	0.09	0.08	1.32	1.32	1.12	1.30	1.36	0.05	0.10	0.08	0.09	0.07	0.04	0.05	0.07	0.09	0.05	0.07	0.09	0.11	0.12	0.09	0.05	0.06	0.10	0.08	0.05
2950	C	T	protein VP1	158	Ile		0.05	0.06	0.05	0.09	0.04	6.37	6.40	5.93	6.04	6.23	0.05	0.04	0.05	0.06	0.06	2.59	2.51	2.49	2.45	2.70	2.35	2.61	2.67	2.43	2.82	0.03	0.07	0.05	0.07	0.10	0.05	0.06	0.06	0.05	0.04	0.07	0.07	0.06	0.07	0.05	0.07	0.07	0.08	0.10	0.06	6.25	6.22	6.33	6.44	6.27	0.04	0.04	0.06	0.05	0.05
3262	C	T	protein VP1	262	Pro		0.03	0.02	0.01	0.03	0.02	0.47	0.51	0.48	0.47	0.41	0.04	0.04	0.05	0.04	0.04	0.74	0.95	0.78	0.78	0.69	0.89	0.87	0.78	0.83	0.89	0.04	0.05	0.04	0.05	0.03	1.07	1.06	1.00	1.01	1.03	0.05	0.04	0.06	0.04	0.06	1.12	1.07	1.00	0.94	0.97	0.55	0.53	0.56	0.52	0.54	0.02	0.03	0.02	0.03	0.02
3396	A	G	protein 2A	7	Lys	Arg	0.05	0.06	0.05	0.09	0.05	1.04	1.19	1.02	1.15	1.16	0.07	0.07	0.05	0.05	0.05	0.49	0.47	0.50	0.52	0.58	0.21	0.25	0.25	0.21	0.24	0.04	0.03	0.06	0.04	0.06	0.05	0.05	0.05	0.04	0.04	0.04	0.06	0.04	0.06	0.04	0.05	0.05	0.05	0.96	1.08	1.00	1.07	1.04	0.06	0.03	0.05	0.06	0.04
3640	C	T	protein 2A	88	Tyr		0.02	0.02	0.01	0.02	0.03	1.78	1.76	1.72	1.87	1.82	0.05	0.04	0.04	0.04	0.07	1.79	2.01	1.80	1.79	2.09	2.27	2.21	2.10	2.21	2.43	0.04	0.04	0.04	0.05	0.04	2.04	2.12	2.01	2.15	2.10	0.03	0.02	0.04	0.04	0.05	2.17	1.97	1.86	1.81	1.89	1.85	1.52	1.86	1.68	1.75	0.03	0.04	0.01	0.02	0.03
3766	A	G	protein 2A	130	Leu		0.12	0.07	0.10	0.09	0.08	1.36	1.35	1.36	1.38	1.39	0.08	0.11	0.10	0.10	0.13	0.68	0.76	0.59	0.81	0.62	0.84	0.62	0.68	0.67	0.76	0.13	0.11	0.10	0.15	0.12	0.16	0.16	0.15	0.15	0.18	0.11	0.06	0.10	0.09	0.12	0.20	0.17	0.16	0.19	0.20	1.38	1.44	1.41	1.43	1.43	0.08	0.10	0.08	0.06	0.08
3956	A	G	protein 2B	45	Ile	Val	0.06	0.05	0.05	0.03	0.07	0.11	0.12	0.06	0.08	0.07	0.08	0.08	0.07	0.05	0.03	0.06	0.06	0.06	0.06	0.05	0.05	0.07	0.05	0.06	0.07	1.29	1.22	1.22	1.11	1.12	0.08	0.08	0.06	0.07	0.11	1.29	1.06	1.31	1.35	1.28	0.06	0.08	0.11	0.08	0.07	0.08	0.07	0.10	0.05	0.10	0.02	0.06	0.04	0.03	0.04
4054	G	A	protein 2B	77	Pro		0.04	0.06	0.05	0.04	0.04	0.03	0.04	0.05	0.07	0.07	0.05	0.05	0.06	0.04	0.06	0.06	0.07	0.03	0.05	0.04	0.02	0.05	0.05	0.03	0.02	4.16	3.94	4.00	3.83	4.22	0.03	0.06	0.03	0.04	0.05	4.02	4.31	3.83	4.13	3.94	0.04	0.05	0.05	0.04	0.06	0.06	0.07	0.07	0.07	0.05	0.05	0.03	0.03	0.04	0.05
4171	A	G	protein 2C	19	Glu		0.04	0.07	0.06	0.05	0.05	1.28	1.36	1.19	1.17	1.22	0.06	0.06	0.09	0.08	0.06	1.88	2.05	1.83	2.00	1.73	1.99	1.98	2.01	2.05	2.19	0.07	0.07	0.07	0.06	0.10	2.75	2.66	2.57	2.52	2.64	0.08	0.06	0.13	0.08	0.08	2.44	2.36	2.47	2.42	2.32	1.32	1.18	1.27	1.29	1.25	0.08	0.06	0.06	0.05	0.07
4872	A	C	protein 2C	253	Glu	Ala	0.01	0.01	0.01	0.03	0.01	2.62	2.49	2.50	2.72	2.68	0.01	0.01	0.01	0.01	0.01	1.87	1.67	1.67	1.67	1.67	1.02	1.02	0.84	0.91	0.99	0.01	0.02	0.01	0.01	0.01	1.80	1.82	1.59	1.69	1.84	0.02	0.01	0.02	0.02	0.02	2.67	2.72	2.53	2.59	2.52	2.55	2.46	2.59	2.70	2.59	0.01	0.01	0.02	0.02	0.02
4872	A	G	protein 2C	253	Glu	Gly	0.19	0.22	0.24	0.23	0.21	12.40	11.99	12.24	12.36	12.18	0.23	0.25	0.25	0.22	0.22	5.67	5.65	5.46	5.62	5.76	4.62	4.61	4.47	4.43	4.61	0.22	0.20	0.20	0.19	0.20	8.18	8.21	8.14	8.22	8.13	0.21	0.19	0.24	0.26	0.23	4.74	4.66	4.67	4.56	4.62	11.96	11.65	11.83	12.14	12.09	0.20	0.24	0.29	0.26	0.25
4883	G	A	protein 2C	257	Asp	Asn	0.02	0.02	0.03	0.03	0.04	0.71	0.80	0.71	0.80	0.73	0.02	0.03	0.04	0.05	0.04	0.83	0.80	0.92	0.87	0.93	0.87	0.92	0.88	0.90	1.04	0.05	0.06	0.06	0.10	0.07	3.19	2.78	2.93	3.05	2.91	0.04	0.04	0.06	0.05	0.06	1.06	1.06	0.93	1.11	0.96	0.93	0.82	0.86	0.87	0.84	0.02	0.04	0.03	0.02	0.04
4884	A	G	protein 2C	257	Asp	Gly	0.13	0.09	0.11	0.08	0.11	2.77	2.69	2.59	2.83	2.70	0.16	0.16	0.14	0.18	0.20	2.50	2.55	2.65	2.68	2.50	2.03	2.01	2.10	2.05	2.09	0.28	0.19	0.17	0.22	0.24	6.36	6.12	6.32	6.00	6.03	0.18	0.17	0.21	0.21	0.17	3.11	2.98	2.95	3.03	2.92	2.59	2.53	2.37	2.47	2.63	0.09	0.12	0.12	0.08	0.10
4925	G	A	protein 2C	271	Asp	Asn	0.03	0.02	0.03	0.02	0.02	4.85	4.74	4.69	4.75	4.60	0.11	0.10	0.07	0.08	0.11	2.45	2.60	2.44	2.48	2.51	1.74	1.63	1.72	1.58	1.61	0.06	0.08	0.07	0.05	0.03	3.51	3.45	3.43	3.38	3.49	0.03	0.08	0.05	0.05	0.04	2.81	2.89	2.79	2.76	2.89	4.52	4.75	4.61	4.48	4.71	0.03	0.04	0.04	0.04	0.04
4935	A	G	protein 2C	274	Gln	Arg	0.11	0.04	0.06	0.07	0.09	5.85	6.25	6.34	6.18	6.15	0.08	0.07	0.06	0.11	0.09	3.35	3.34	3.37	3.38	3.39	2.75	2.39	2.28	2.30	2.38	0.26	0.28	0.25	0.27	0.22	3.95	3.58	3.95	3.69	3.74	0.37	0.34	0.40	0.34	0.32	1.09	1.13	1.08	1.09	1.05	6.16	6.09	6.21	6.21	6.12	0.09	0.07	0.08	0.09	0.06
5296	A	G	protein 3A	65	Ala		0.05	0.07	0.04	0.04	0.05	0.55	0.52	0.45	0.50	0.49	0.06	0.07	0.05	0.05	0.04	0.93	0.89	0.89	0.95	0.88	0.90	0.96	0.94	0.81	0.98	0.06	0.09	0.07	0.03	0.02	1.13	1.16	1.23	1.20	1.02	0.05	0.05	0.06	0.06	0.08	1.08	1.12	1.03	1.02	1.00	0.51	0.54	0.55	0.54	0.47	0.04	0.03	0.04	0.04	0.05
5473	T	C	protein 3C	15	Ile		0.04	0.03	0.04	0.04	0.02	0.05	0.07	0.06	0.06	0.07	0.04	0.05	0.07	0.02	0.04	0.02	0.03	0.04	0.03	0.04	0.05	0.03	0.03	0.04	0.04	6.32	6.14	6.09	5.89	6.35	0.02	0.05	0.05	0.02	0.04	6.37	6.16	6.09	6.22	6.09	0.04	0.02	0.02	0.04	0.04	0.10	0.06	0.04	0.04	0.05	0.03	0.03	0.03	0.03	0.03
5476	T	C	protein 3C	16	Val		0.05	0.02	0.06	0.02	0.05	0.08	0.08	0.08	0.08	0.10	0.12	0.04	0.09	0.07	0.07	0.04	0.06	0.04	0.05	0.04	0.04	0.06	0.05	0.08	0.06	4.28	3.96	3.98	3.96	3.98	0.06	0.04	0.08	0.06	0.07	4.03	3.98	3.99	4.13	4.11	0.04	0.06	0.07	0.07	0.05	0.07	0.07	0.09	0.05	0.08	0.05	0.05	0.06	0.05	0.06
5832	T	C	protein 3C	135	Ile	Thr	0.06	0.07	0.09	0.07	0.07	34.75	34.96	34.44	34.56	34.20	0.06	0.14	0.12	0.08	0.11	15.00	15.35	15.05	14.54	14.85	15.82	15.52	15.31	15.79	16.03	0.23	0.24	0.24	0.23	0.29	0.11	0.14	0.13	0.13	0.11	0.19	0.17	0.23	0.20	0.19	0.10	0.12	0.06	0.10	0.06	34.39	34.83	34.56	34.65	34.46	0.05	0.06	0.04	0.05	0.06
6001	A	G	protein 3D	8	Pro		0.09	0.10	0.09	0.08	0.09	0.11	0.11	0.13	0.10	0.09	0.07	0.09	0.11	0.10	0.09	0.08	0.08	0.10	0.06	0.03	0.07	0.05	0.08	0.04	0.09	1.70	1.55	1.55	1.60	1.63	0.12	0.08	0.09	0.07	0.07	1.75	1.81	1.77	1.67	1.69	0.08	0.07	0.12	0.06	0.03	0.07	0.13	0.12	0.09	0.10	0.10	0.09	0.09	0.07	0.08
6061	T	C	protein 3D	28	Ser		99.92	99.96	99.94	99.92	99.90	0.13	0.15	0.12	0.07	0.11	0.10	0.08	0.07	0.06	0.06	0.10	0.09	0.09	0.10	0.07	0.11	0.12	0.08	0.08	0.07	0.12	0.13	0.10	0.11	0.13	0.12	0.10	0.07	0.07	0.08	0.11	0.09	0.07	0.06	0.07	0.08	0.10	0.07	0.07	0.06	0.10	0.09	0.09	0.06	0.08	99.93	99.92	99.90	99.92	99.91
6304	G	A	protein 3D	109	Ala		99.84	99.78	99.77	99.80	99.86	0.06	0.06	0.09	0.05	0.06	0.04	0.08	0.06	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.07	0.04	0.09	0.03	0.08	0.06	0.05	0.05	0.09	0.08	0.07	0.06	0.06	0.05	0.05	0.09	0.08	0.11	0.09	0.08	0.06	0.04	0.05	0.07	0.07	0.04	0.07	0.10	0.05	0.05	99.85	99.87	99.81	99.87	99.86
6421	C	T	protein 3D	148	Tyr		0.04	0.03	0.03	0.02	0.03	0.05	0.06	0.05	0.06	0.07	0.06	0.06	0.07	0.08	0.04	0.02	0.04	0.06	0.03	0.04	0.08	0.07	0.04	0.08	0.04	1.46	1.32	1.36	1.33	1.50	0.06	0.09	0.05	0.09	0.06	1.59	1.50	1.46	1.65	1.43	0.05	0.05	0.06	0.05	0.05	0.09	0.09	0.07	0.10	0.06	0.07	0.04	0.04	0.05	0.04
6760	A	T	protein 3D	261	Arg	Ser	0.00	0.02	0.01	0.02	0.01	0.01	0.02	0.02	0.01	0.01	99.80	99.81	99.78	99.79	99.79	0.01	0.00	0.03	0.01	0.01	0.01	0.00	0.03	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.02	0.02	0.02	0.01	0.01
7004	C	T	protein 3D	343	Leu		99.81	99.82	99.81	99.81	99.80	0.02	0.05	0.05	0.02	0.04	0.03	0.02	0.01	0.02	0.03	0.03	0.04	0.04	0.03	0.02	0.00	0.01	0.03	0.02	0.03	0.01	0.02	0.01	0.02	0.02	0.02	0.02	0.00	0.02	0.03	0.03	0.04	0.04	0.02	0.02	0.02	0.04	0.03	0.04	0.01	0.02	0.03	0.01	0.03	0.03	99.83	99.84	99.84	99.86	99.81

How can whole-genome SNP profiles be compared?


A C G T
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

Calculating difference between SNP profiles

Difference for the nucleotide at position *i* in the genome

$\Delta_i = abs(A_1-A_2) + abs(C_1-C_2) + abs(G_1-G_2) + abs(T_1-T_2)$

	Reference	Profile 1	Profile 2	Difference
A	100	99.722	99.859	0.000
C	0	0.086	0.010	0.076
G	0	0.150	0.121	0.029
T	0	0.043	0.010	0.033
			Δ =	0.137

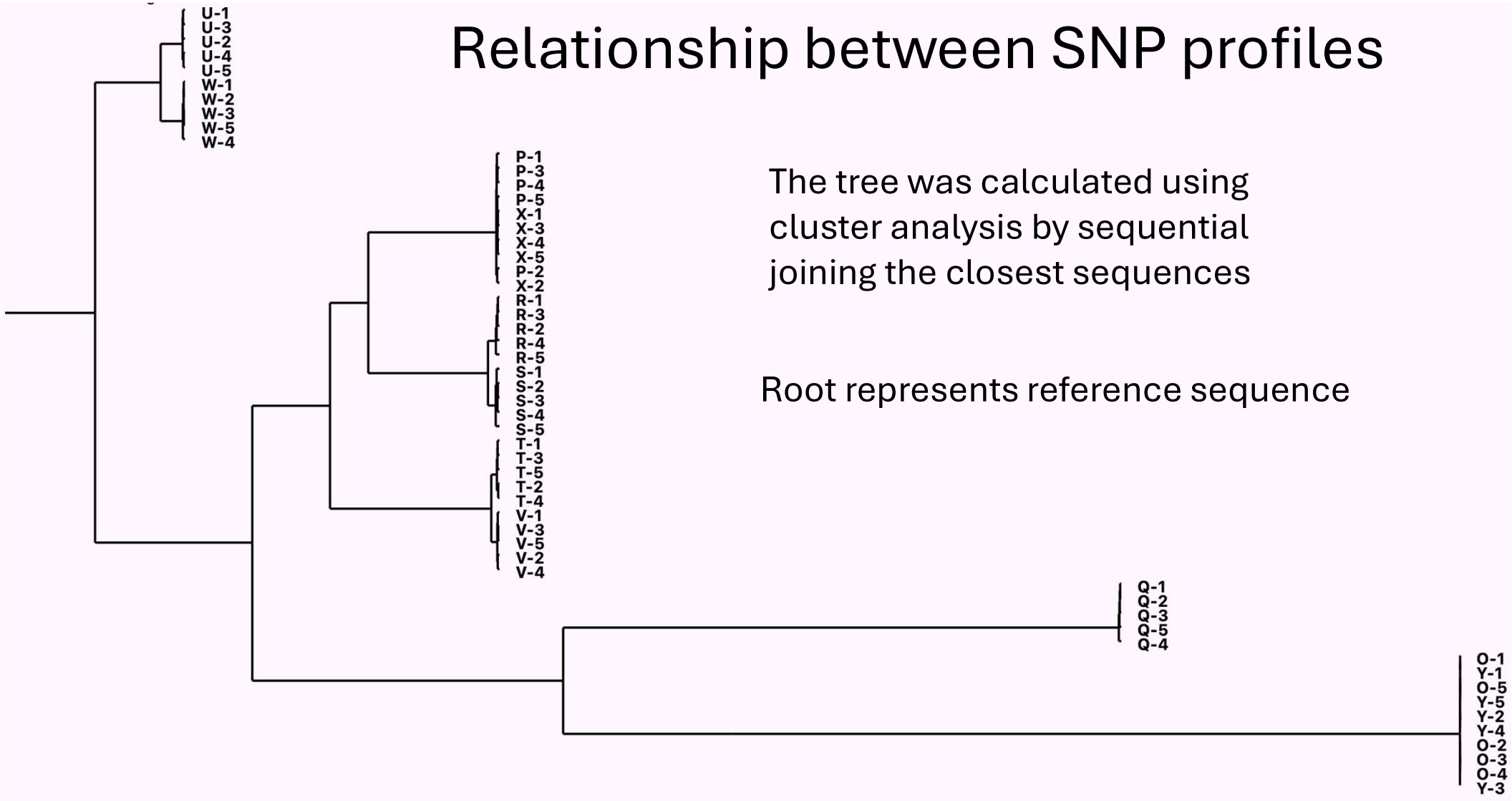
Total difference between profiles of the length N is equal to

$$\sum_i^N \Delta i$$

Pair-wise distances matrix

	O-1	O-2	O-3	O-4	O-5	P-1	P-2	P-3	P-4	P-5	Q-1	Q-2	Q-3	Q-4	Q-5	R-1	R-2	R-3	R-4	R-5	S-1	S-2	S-3	S-4	S-5	T-1	T-2	T-3	T-4	T-5
O-1		0.001	0.001	0.000	0.000	1.788	1.788	1.786	1.788	1.787	2.236	2.236	2.236	2.236	2.236	1.816	1.816	1.817	1.817	1.819	1.823	1.825	1.824	1.824	1.826	1.732	1.732	1.732	1.731	1.732
O-2	0.001		0.000	0.001	0.001	1.787	1.788	1.786	1.788	1.787	2.235	2.236	2.236	2.235	2.236	1.816	1.816	1.816	1.816	1.819	1.823	1.824	1.824	1.824	1.826	1.732	1.731	1.732	1.731	1.732
O-3	0.001	0.000		0.000	0.001	1.787	1.787	1.786	1.787	1.786	2.235	2.236	2.236	2.235	2.236	1.815	1.816	1.816	1.816	1.819	1.822	1.824	1.823	1.824	1.826	1.732	1.731	1.731	1.731	1.731
O-4	0.000	0.001	0.000		0.001	1.787	1.787	1.786	1.787	1.787	2.235	2.235	2.236	2.235	2.236	1.816	1.816	1.816	1.816	1.819	1.822	1.824	1.823	1.824	1.826	1.732	1.731	1.732	1.731	1.731
O-5	0.000	0.001	0.001	0.001		1.787	1.788	1.786	1.787	1.787	2.235	2.236	2.236	2.235	2.236	1.816	1.816	1.816	1.816	1.819	1.823	1.824	1.824	1.824	1.826	1.732	1.731	1.732	1.731	1.732
P-1	1.788	1.787	1.787	1.787	1.787		0.007	0.008	0.006	0.009	1.489	1.489	1.489	1.489	1.489	0.356	0.354	0.359	0.362	0.365	0.374	0.381	0.382	0.378	0.381	0.438	0.437	0.437	0.437	0.436
P-2	1.788	1.788	1.787	1.787	1.788	0.007		0.009	0.008	0.009	1.490	1.490	1.490	1.489	1.490	0.356	0.354	0.359	0.362	0.365	0.374	0.381	0.381	0.378	0.381	0.439	0.438	0.439	0.438	0.437
P-3	1.786	1.786	1.786	1.786	1.786	0.008	0.009		0.005	0.008	1.488	1.488	1.488	1.487	1.488	0.356	0.353	0.358	0.361	0.364	0.374	0.381	0.381	0.378	0.381	0.434	0.433	0.433	0.433	0.432
P-4	1.788	1.788	1.787	1.787	1.787	0.006	0.008	0.005		0.006	1.489	1.489	1.489	1.488	1.489	0.355	0.353	0.357	0.361	0.363	0.373	0.380	0.380	0.377	0.380	0.437	0.435	0.436	0.435	0.435
P-5	1.787	1.787	1.786	1.787	1.787	0.009	0.009	0.008	0.006		1.488	1.489	1.488	1.488	1.489	0.350	0.348	0.352	0.355	0.358	0.368	0.375	0.375	0.372	0.375	0.435	0.434	0.434	0.434	0.433
Q-1	2.236	2.235	2.235	2.235	2.235	1.489	1.490	1.488	1.489	1.488		0.003	0.004	0.006	0.006	1.523	1.523	1.524	1.524	1.526	1.531	1.533	1.532	1.532	1.535	1.419	1.419	1.418	1.418	1.419
Q-2	2.236	2.236	2.236	2.235	2.236	1.489	1.490	1.488	1.489	1.489	0.003		0.004	0.005	0.004	1.523	1.523	1.524	1.524	1.527	1.531	1.533	1.532	1.533	1.535	1.419	1.419	1.419	1.419	1.419
Q-3	2.236	2.236	2.236	2.236	2.236	1.489	1.490	1.488	1.489	1.488	0.004	0.004		0.007	0.005	1.523	1.523	1.524	1.524	1.526	1.531	1.533	1.532	1.532	1.535	1.419	1.419	1.418	1.418	1.419
Q-4	2.236	2.235	2.235	2.235	2.235	1.489	1.489	1.487	1.488	1.488	0.006	0.005	0.007		0.007	1.522	1.523	1.523	1.523	1.526	1.531	1.533	1.532	1.532	1.535	1.418	1.418	1.418	1.418	1.418
Q-5	2.236	2.236	2.236	2.236	2.236	1.489	1.490	1.488	1.489	1.489	0.006	0.004	0.005	0.007		1.523	1.523	1.524	1.524	1.527	1.531	1.533	1.532	1.533	1.535	1.419	1.419	1.419	1.419	1.419
R-1	1.816	1.816	1.815	1.816	1.816	0.356	0.356	0.356	0.355	0.350	1.523	1.523	1.523	1.522	1.523		0.005	0.005	0.007	0.011	0.030	0.036	0.035	0.035	0.040	0.514	0.513	0.513	0.513	0.511
R-2	1.816	1.816	1.816	1.816	1.816	0.354	0.354	0.353	0.353	0.348	1.523	1.523	1.523	1.523	1.523	0.005		0.006	0.010	0.012	0.029	0.036	0.035	0.035	0.039	0.515	0.514	0.514	0.514	0.512
R-3	1.817	1.816	1.816	1.816	1.816	0.359	0.359	0.358	0.357	0.352	1.524	1.524	1.524	1.523	1.524	0.005	0.006		0.006	0.009	0.027	0.033	0.032	0.032	0.037	0.516	0.516	0.516	0.516	0.513
R-4	1.817	1.816	1.816	1.816	1.816	0.362	0.362	0.361	0.361	0.355	1.524	1.524	1.524	1.523	1.524	0.007	0.010	0.006		0.009	0.029	0.034	0.033	0.034	0.039	0.516	0.516	0.516	0.515	0.513
R-5	1.819	1.819	1.819	1.819	1.819	0.365	0.365	0.364	0.363	0.358	1.526	1.527	1.526	1.526	1.527	0.011	0.012	0.009	0.009		0.024	0.029	0.028	0.029	0.033	0.523	0.523	0.523	0.522	0.520
S-1	1.823	1.823	1.822	1.822	1.823	0.374	0.374	0.374	0.373	0.368	1.531	1.531	1.531	1.531	1.531	0.030	0.029	0.027	0.029	0.024		0.009	0.010	0.007	0.012	0.536	0.535	0.536	0.535	0.533
S-2	1.825	1.824	1.824	1.824	1.824	0.381	0.381	0.381	0.380	0.375	1.533	1.533	1.533	1.533	1.533	0.036	0.036	0.033	0.034	0.029	0.009		0.004	0.005	0.007	0.541	0.541	0.541	0.540	0.538
S-3	1.824	1.824	1.823	1.823	1.824	0.382	0.381	0.381	0.380	0.375	1.532	1.532	1.532	1.532	1.532	0.035	0.035	0.032	0.033	0.028	0.010	0.004		0.006	0.010	0.539	0.538	0.538	0.538	0.536
S-4	1.824	1.824	1.824	1.824	1.824	0.378	0.378	0.378	0.377	0.372	1.532	1.533	1.532	1.532	1.533	0.035	0.035	0.032	0.034	0.029	0.007	0.005	0.006		0.009	0.539	0.538	0.538	0.538	0.536
S-5	1.826	1.826	1.826	1.826	1.826	0.381	0.381	0.381	0.380	0.375	1.535	1.535	1.535	1.535	1.535	0.040	0.039	0.037	0.039	0.033	0.012	0.007	0.010	0.009		0.546	0.545	0.545	0.545	0.543
T-1	1.732	1.732	1.732	1.732	1.732	0.438	0.439	0.434	0.437	0.435	1.419	1.419	1.419	1.418	1.419	0.514	0.515	0.516	0.516	0.523	0.536	0.541	0.539	0.539	0.546		0.006	0.005	0.007	0.005
T-2	1.732	1.731	1.731	1.731	1.731	0.437	0.438	0.433	0.435	0.434	1.419	1.419	1.419	1.418	1.419	0.513	0.514	0.516	0.516	0.523	0.535	0.541	0.538	0.538	0.545	0.006		0.006	0.003	0.007
T-3	1.732	1.732	1.731	1.732	1.732	0.437	0.439	0.433	0.436	0.434	1.418	1.419	1.418	1.418	1.419	0.513	0.514	0.516	0.516	0.523	0.536	0.541	0.538	0.538	0.545	0.005	0.006		0.005	0.005
T-4	1.731	1.731	1.731	1.731	1.731	0.437	0.438	0.433	0.435	0.434	1.418	1.419	1.418	1.418	1.419	0.513	0.514	0.516	0.515	0.522	0.535	0.540	0.538	0.538	0.545	0.007	0.003	0.005		0.008
T-5	1.732	1.732	1.731	1.731	1.732	0.436	0.437	0.432	0.435	0.433	1.419	1.419	1.419	1.418	1.419	0.511	0.512	0.513	0.513	0.520	0.533	0.538	0.536	0.536	0.543	0.005	0.007	0.005	0.008	

Relationship between SNP profiles



The tree was calculated using cluster analysis by sequential joining the closest sequences

Root represents reference sequence

Comparison of samples U and W

Nucleotide	Base	Mutation	Gene	Amino acid	Reference amino acid	Mutant amino acid	Contribution	Set1		Set2		p-value	U-1	U-2	U-3	U-4	U-5	W-1	W-2	W-3	W-4	W-5
Distance between profiles is				0.19																		
4872	A	G	protein 2C	253	Glu	Gly	0.23698	8.18 ±	0.04	4.65 ±	0.07	1E-13	8.18	8.21	8.14	8.22	8.13	4.74	4.66	4.67	4.56	4.62
2493	C	T	protein VP1	6	Thr	Ile	0.18725	21.35 ±	0.29	24.82 ±	0.24	3E-08	21.58	21.68	21.30	21.23	20.97	24.84	24.99	24.80	24.43	25.02
4884	A	G	protein 2C	257	Asp	Gly	0.17111	6.16 ±	0.17	3.00 ±	0.07	2E-10	6.36	6.12	6.32	6.00	6.03	3.11	2.98	2.95	3.03	2.92
4935	A	G	protein 2C	274	Gln	Arg	0.14555	3.78 ±	0.16	1.09 ±	0.03	4E-10	3.95	3.58	3.95	3.69	3.74	1.09	1.13	1.08	1.09	1.05
4883	G	A	protein 2C	257	Asp	Asn	0.10531	2.97 ±	0.15	1.03 ±	0.08	7E-09	3.19	2.78	2.93	3.05	2.91	1.06	1.06	0.93	1.11	0.96
2731	C	T	protein VP1	85	Val		0.06626	1.28 ±	0.09	0.06 ±	0.02	2E-09	1.32	1.32	1.12	1.30	1.36	0.04	0.05	0.07	0.09	0.05
4925	G	A	protein 2C	271	Asp	Asn	0.03382	3.45 ±	0.05	2.83 ±	0.06	1E-07	3.51	3.45	3.43	3.38	3.49	2.81	2.89	2.79	2.76	2.89
472	T	C	non-coding				0.02241	1.10 ±	0.03	0.69 ±	0.05	2E-07	1.08	1.10	1.12	1.14	1.06	0.61	0.68	0.72	0.68	0.73
4171	A	G	protein 2C	19	Glu		0.01221	2.63 ±	0.09	2.40 ±	0.06	0.0014	2.75	2.66	2.57	2.52	2.64	2.44	2.36	2.47	2.42	2.32
3640	C	T	protein 2A	88	Tyr		0.00964	2.09 ±	0.06	1.94 ±	0.14	0.069	2.04	2.12	2.01	2.15	2.10	2.17	1.97	1.86	1.81	1.89
5296	A	G	protein 3A	65	Ala		0.00596	1.15 ±	0.08	1.05 ±	0.05	0.051	1.13	1.16	1.23	1.20	1.02	1.08	1.12	1.03	1.02	1.00
3262	C	T	protein VP1	262	Pro		0.00350	1.04 ±	0.03	1.02 ±	0.08	0.6455	1.071	1.063	1.002	1.013	1.029	1.118	1.071	0.999	0.936	0.966

Comparison of samples T and V

Nucleotide	Base	Mutation	Gene	Amino acid	Reference amino acid	Mutant amino acid	Contribution	Set T		Set V		p-value	T-1	T-2	T-3	T-4	T-5	V-1	V-2	V-3	V-4	V-5
Distance between profiles is				0.04																		
2493	C	T	protein VP1	6	Thr	Ile	0.45301	94.62 ±	0.15	92.71 ±	0.20	2E-07	94.66	94.70	94.67	94.72	94.36	92.47	92.92	92.69	92.92	92.57
472	T	C	non-coding				0.28475	2.34 ±	0.21	1.14 ±	0.11	3E-06	2.43	2.07	2.59	2.19	2.44	1.21	1.03	1.30	1.12	1.06
4054	G	A	protein 2B	77	Pro		0.04212	4.03 ±	0.16	4.05 ±	0.18	0.8794	4.16	3.94	4.00	3.83	4.22	4.02	4.31	3.83	4.13	3.94
5473	T	C	protein 3C	15	Ile		0.03862	6.16 ±	0.19	6.19 ±	0.12	0.7713	6.32	6.14	6.09	5.89	6.35	6.37	6.16	6.09	6.22	6.09
6421	C	T	protein 3D	148	Tyr		0.03404	1.39 ±	0.08	1.53 ±	0.09	0.0428	1.46	1.32	1.36	1.33	1.50	1.59	1.50	1.46	1.65	1.43
6001	A	G	protein 3D	8	Pro		0.03215	1.61 ±	0.06	1.74 ±	0.06	0.0092	1.70	1.55	1.55	1.60	1.63	1.75	1.81	1.77	1.67	1.69
687	A	C	non-coding				0.03158	1.04 ±	0.13	0.89 ±	0.04	0.0363	1.02	0.88	1.10	0.98	1.22	0.87	0.86	0.88	0.92	0.94
713	A	G	non-coding				0.02900	1.65 ±	0.10	1.72 ±	0.11	0.3551	1.66	1.63	1.72	1.51	1.76	1.60	1.60	1.84	1.77	1.79
3956	A	G	protein 2B	45	Ile	Val	0.02852	1.19 ±	0.08	1.26 ±	0.11	0.3072	1.29	1.22	1.22	1.11	1.12	1.29	1.06	1.31	1.35	1.28
5476	T	C	protein 3C	16	Val		0.02621	4.03 ±	0.14	4.05 ±	0.07	0.8346	4.28	3.96	3.98	3.96	3.98	4.03	3.98	3.99	4.13	4.11

Comparison of samples P and X

Nucleotide	Base	Mutation	Gene	Amino acid	Reference amino acid	Mutant amino acid	Contribution	Set P		Set X		p-value	P-1	P-2	P-3	P-4	P-5	X-1	X-2	X-3	X-4	X-5
Distance between profiles is				0.02								0.177										
4872	A	G	protein 2C	253	Glu	Gly	0.18076	12.23 ±	0.16	11.93 ±	0.20	0.0322	12.40	11.99	12.24	12.36	12.18	11.96	11.65	11.83	12.14	12.09
5832	T	C	protein 3C	135	Ile	Thr	0.10579	34.58 ±	0.29	34.58 ±	0.17	0.9843	34.75	34.96	34.44	34.56	34.20	34.39	34.83	34.56	34.65	34.46
2493	C	T	protein VP1	6	Thr	Ile	0.09717	75.30 ±	0.21	75.16 ±	0.18	0.3046	75.38	75.26	75.55	75.33	74.97	74.88	75.34	75.08	75.27	75.23
4884	A	G	protein 2C	257	Asp	Gly	0.08477	2.72 ±	0.09	2.52 ±	0.10	0.0112	2.77	2.69	2.59	2.83	2.70	2.59	2.53	2.37	2.47	2.63
2950	C	T	protein VP1	158	Ile		0.07872	6.19 ±	0.21	6.30 ±	0.09	0.3023	6.37	6.40	5.93	6.04	6.23	6.25	6.22	6.33	6.44	6.27
1867	T	C	protein VP3	35	Asp		0.07180	1.24 ±	0.05	1.41 ±	0.02	7E-05	1.16	1.27	1.23	1.27	1.26	1.42	1.42	1.43	1.38	1.39
4925	G	A	protein 2C	271	Asp	Asn	0.05919	4.72 ±	0.09	4.62 ±	0.12	0.1392	4.85	4.74	4.69	4.75	4.60	4.52	4.75	4.61	4.48	4.71
4935	A	G	protein 2C	274	Gln	Arg	0.05674	6.15 ±	0.19	6.16 ±	0.05	0.9547	5.85	6.25	6.34	6.18	6.15	6.16	6.09	6.21	6.21	6.12
3640	C	T	protein 2A	88	Tyr		0.04878	1.79 ±	0.06	1.73 ±	0.14	0.4065	1.78	1.76	1.72	1.87	1.82	1.85	1.52	1.86	1.68	1.75
1003	A	G	protein VP2	18	Leu		0.04777	1.92 ±	0.06	1.87 ±	0.13	0.5102	1.93	1.93	1.82	1.96	1.95	2.05	1.76	1.90	1.93	1.73
769	A	G	protein VP4	9	Lys		0.04580	1.58 ±	0.06	1.50 ±	0.09	0.1311	1.50	1.64	1.54	1.58	1.63	1.53	1.64	1.42	1.47	1.43
3396	A	G	protein 2A	7	Lys	Arg	0.04284	1.11 ±	0.08	1.03 ±	0.05	0.0703	1.041	1.194	1.018	1.152	1.16	0.958	1.075	0.997	1.067	1.035
4171	A	G	protein 2C	19	Glu		0.03043	1.24 ±	0.08	1.26 ±	0.05	0.6722	1.275	1.36	1.193	1.171	1.218	1.324	1.181	1.266	1.287	1.249
472	T	C	non-coding				0.02869	1.02 ±	0.07	0.95 ±	0.07	0.1675	0.92	1.086	1.07	1.005	0.997	1.025	0.981	0.989	0.871	0.896
3766	A	G	protein 2A	130	Leu		0.02074	1.37 ±	0.02	1.42 ±	0.02	0.0046	1.357	1.354	1.362	1.384	1.385	1.378	1.436	1.414	1.426	1.427

Statistical significance

vs.

Biological significance

Mutations present in acceptable OPV3 lots

Nucleotide	Base	Mutation	Gene	Amino acid	Reference amino acid	Mutant amino acid	Max
537	G	A	non-coding				14.05
699	C	T	non-coding				1.67
713	A	G	non-coding				25.02
773	G	A	protein VP4	11	Gly	Ser	4.92
858	A	G	protein VP4	39	Asn	Ser	2.78
876	A	G	protein VP4	45	Asp	Gly	2.14
898	A	G	protein VP4	52	Lys		14.34
1341	G	T	protein VP2	131	Cys	Phe	6.51
1485	T	C	protein VP2	179	Leu	Pro	4.49
1537	C	T	protein VP2	196	Ile		4.31
1681	T	C	protein VP2	244	Val		5.63
1699	G	A	protein VP2	250	Val		98.36
2440	A	T	protein VP3	226	Arg		95.62
2493	C	T	protein VP1	6	Thr	Ile	96.38
2504	G	T	protein VP1	10	Ala	Ser	3.41
2696	A	G	protein VP1	74	Thr	Ala	7.97
2702	G	C	protein VP1	76	Glu	Gln	2.08
2703	A	G	protein VP1	76	Glu	Gly	3.34
2731	C	T	protein VP1	85	Val		4.74
3256	G	T	protein VP1	260	Met	Ile	4.08
3262	C	T	protein VP1	262	Pro		2.03
3278	T	C	protein VP1	268	Trp	Arg	2.47
3353	T	C	protein VP1	293	Ser	Pro	98.57
3357	A	G	protein VP1	294	Glu	Gly	99.10
3640	C	T	protein 2A	88	Tyr		2.87
3700	C	T	protein 2A	108	Asp		98.40
3723	A	G	protein 2A	116	Gln	Arg	7.47
3723	A	G	protein 2A	116	Gln	Arg	1.82
3956	A	G	protein 2B	45	Ile	Val	92.76
4054	G	A	protein 2B	77	Pro		4.30
4171	A	G	protein 2C	19	Glu		4.43
4202	G	T	protein 2C	30	Asp	Tyr	5.58
4872	A	G	protein 2C	253	Glu	Gly	8.84
4883	G	A	protein 2C	257	Asp	Asn	2.93
4884	A	G	protein 2C	257	Asp	Gly	6.19
4925	G	A	protein 2C	271	Asp	Asn	3.54
4935	A	G	protein 2C	274	Gln	Arg	3.95
5075	G	T	protein 2C	321	Gly	Cys	5.94
5137	C	T	protein 3A	12	Ile		3.66
5296	A	G	protein 3A	65	Ala		2.72
5473	T	C	protein 3C	15	Ile		6.78
5476	T	C	protein 3C	16	Val		6.34
5767	T	C	protein 3C	113	Tyr		98.25
5787	C	A	protein 3C	120	Thr	Asn	4.73
5832	T	C	protein 3C	135	Ile	Thr	97.94
6001	A	G	protein 3D	8	Pro		4.97
6178	C	T	protein 3D	67	Ile		1.04
6421	C	T	protein 3D	148	Tyr		98.36
6505	T	A	protein 3D	176	Ile		98.27
6760	A	T	protein 3D	261	Arg	Ser	95.44
6819	G	T	protein 3D	281	Cys	Phe	4.91

Comparing a new vaccine lot with the historical baseline

Comparison of Lot X with historical baseline

Nucleotide	Base	Mutation	Gene	Amino acid	Reference amino acid	Mutant amino acid	Contribution	Historical		New lot X		p-value	Lot 1.1	Lot 1.2	Lot 1.3	Lot 1.4	Lot 1.5	Lot 2.1	Lot 2.2	Lot 2.3	Lot 2.4	Lot 2.5	Lot 3.1	Lot 3.2	Lot 3.3	Lot 3.4	Lot 3.5	Lot 4.1	Lot 4.2	Lot 4.3	Lot 4.4	Lot 4.5	Lot 5.1	Lot 5.2	Lot 5.3	Lot 5.4	Lot 5.5	Lot X1	Lot X2	Lot X3	Lot X4	Lot X5
Distance between profiles is				0.1								0.000000																															
2493	C	T	protein VP1	6	Thr	Ile	0.75001	12.99	± 1.15	20.47	± 0.60	0.000000	12.19	12.91	12.74	12.88	12.58	14.26	14.68	14.40	14.53	14.40	13.41	13.94	14.11	14.25	14.19	12.04	12.55	12.68	12.50	12.57	11.15	11.38	11.44	11.59	11.34	21.30	20.59	19.75	20.66	20.04
4171	A	G	protein 2C	19	Glu		0.04291	3.28	± 0.18	3.34	± 0.63	0.844356	3.77	3.26	3.11	3.24	3.12	3.27	3.27	3.23	3.26	3.31	3.70	3.10	3.50	3.14	3.56	3.24	3.29	3.12	3.33	3.15	3.11	3.22	3.23	3.16	3.29	4.43	2.93	3.13	2.92	3.28
3262	C	T	protein VP1	262	Pro		0.02828	1.61	± 0.06	1.66	± 0.37	0.771601	1.69	1.64	1.56	1.60	1.69	1.68	1.62	1.60	1.58	1.60	1.53	1.69	1.60	1.59	1.54	1.51	1.69	1.58	1.74	1.50	1.59	1.64	1.66	1.62	1.61	1.13	1.60	2.01	1.55	2.03
5296	A	G	protein 3A	65	Ala		0.02542	1.34	± 0.08	1.24	± 0.31	0.439196	1.37	1.39	1.36	1.37	1.38	1.27	1.34	1.29	1.28	1.33	1.25	1.37	1.61	1.40	1.54	1.25	1.31	1.28	1.35	1.26	1.33	1.32	1.36	1.31	1.30	1.56	1.39	0.93	1.41	0.89
3723	A	G	protein 2A	116	Gln	Arg	0.02442	1.17	± 0.11	1.30	± 0.27	0.302031	1.26	1.10	1.14	1.11	1.07	1.06	1.07	1.09	1.07	1.07	1.02	1.36	1.05	1.39	1.01	1.20	1.32	1.21	1.31	1.17	1.24	1.18	1.26	1.21	1.23	0.95	1.21	1.60	1.18	1.54
3640	C	T	protein 2A	88	Tyr		0.01980	2.29	± 0.12	2.35	± 0.29	0.667086	2.75	2.27	2.16	2.25	2.17	2.17	2.31	2.28	2.33	2.34	2.10	2.19	2.20	2.26	2.29	2.29	2.38	2.39	2.37	2.32	2.17	2.34	2.32	2.30	2.29	2.87	2.24	2.16	2.25	2.21

Lot X against historical baseline

Nucleotide	Base	Mutation	Gene	Amino acid	Reference amino acid	Mutant amino acid	Contribution	Historical		New lot X		p-value
Distance between profiles is				0.1								0.000000
2493	C	T	protein VP1	6	Thr	Ile	0.75001	12.99 ±	1.15	20.47 ±	0.60	0.000000
4171	A	G	protein 2C	19	Glu		0.04291	3.28 ±	0.18	3.34 ±	0.63	0.844356
3262	C	T	protein VP1	262	Pro		0.02828	1.61 ±	0.06	1.66 ±	0.37	0.771601
5296	A	G	protein 3A	65	Ala		0.02542	1.34 ±	0.08	1.24 ±	0.31	0.439196
3723	A	G	protein 2A	116	Gln	Arg	0.02442	1.17 ±	0.11	1.30 ±	0.27	0.302031
3640	C	T	protein 2A	88	Tyr		0.01980	2.29 ±	0.12	2.35 ±	0.29	0.667086

Conclusion: New lot X is acceptable

Lot Y against historical baseline

Nucleotide	Base	Mutation	Gene	Amino acid	Reference amino acid	Mutant amino acid	Contribution	historical		New lot Y		p-value
Distance between profiles is				0.07								0.000000
2493	C	T	protein VP1	6	Thr	Ile	0.79075	12.99 ±	1.15	7.44 ±	0.11	0.000000
4171	A	G	protein 2C	19	Glu		0.02289	3.28 ±	0.18	3.30 ±	0.12	0.776733
3640	C	T	protein 2A	88	Tyr		0.01676	2.29 ±	0.12	2.34 ±	0.09	0.295683
3723	A	G	protein 2A	116	Gln	Arg	0.01419	1.17 ±	0.11	1.16 ±	0.06	0.767156
5296	A	G	protein 3A	65	Ala		0.01117	1.34 ±	0.08	1.38 ±	0.05	0.189924
3262	C	T	protein VP1	262	Pro		0.00917	1.61 ±	0.06	1.66 ±	0.01	0.002605

Conclusion: New lot Y is acceptable

Lot Z against historical baseline

Nucleotide	Base	Mutation	Gene	Amino acid	Reference amino acid	Mutant amino acid	Contribution	historical		New lot Z		p-value
Distance between profiles is				0.04								0.001
2493	C	T	protein VP1	6	Thr	Ile	0.63017	12.99 ±	1.15	10.43 ±	0.14	0.0000
4872	A	G	protein 2C	253	Glu	Gly	0.12053	0.96 ±	0.64	0.49 ±	0.01	0.0011
4925	G	A	protein 2C	271	Asp	Asn	0.07370	0.51 ±	0.50	0.14 ±	0.01	0.0004
4884	A	G	protein 2C	257	Asp	Gly	0.05943	0.65 ±	0.41	0.33 ±	0.02	0.0002
4171	A	G	protein 2C	19	Glu		0.03074	3.28 ±	0.18	3.24 ±	0.05	0.5946
3723	A	G	protein 2A	116	Gln	Arg	0.02792	1.17 ±	0.11	1.24 ±	0.03	0.0050
3640	C	T	protein 2A	88	Tyr		0.02760	2.29 ±	0.12	2.35 ±	0.07	0.1443
5296	A	G	protein 3A	65	Ala		0.01637	1.34 ±	0.08	1.32 ±	0.05	0.3895
3262	C	T	protein VP1	262	Pro		0.01356	1.61 ±	0.06	1.61 ±	0.03	0.7642

Conclusion: New lot Z is acceptable

Pass-Fail decisions based on wg-SNPP

- During the establishment of OPV production first several batches of vaccine should be tested in animals as well as by generating whole-genome single-nucleotide polymorphism (SNP) profiles by HTS
  - new manufacturer or major change in production conditions, new seed virus, etc.
- After consistency of manufacture is established, only HTS can be performed
- If a breach of consistency is detected:
  - Careful review of the specific sequencing data should be conducted
  - Based on the results, animal testing may be recommended
  - If the conclusion is that the lot is acceptable, the SNP database is updated

Summary

- This approach determines how close two SNP profiles are
- It agnostically identifies the SNPs that contribute most to the difference
- It enables statistical evaluation of the difference
- It can be used to compare a new vaccine batch to a set of previously produced lots and decide about its consistency
- It is applicable to other biopharmaceutical products

Questions?