

Protecting and improving the nation's health

# **An Electronic Resource for Mining Viral Genomes**

**NCPV500** 

Karen Buttigieg<sup>1</sup>, Sarah Alexander<sup>1</sup>, Lidia Liberty<sup>1</sup>, Joseph Taylor<sup>1</sup>, Nicholas Grayson<sup>2</sup>, Edward Burnett<sup>1</sup>, Julian Parkhill<sup>2</sup>, Julie E. Russell<sup>1</sup>

<sup>1</sup>Public Health England, Porton Down, UK. <sup>2</sup>Wellcome Trust Sanger Institute, Cambridge, UK.

# INTRODUCTION

NCPV500 is a collaboration between Public Health England and the Wellcome Trust Sanger Institute to produce 500 viral genomes from PHE's National Collection of Pathogenic Viruses (NCPV) using the Illumina sequencing platform. NCPV curates and supplies authenticated human pathogenic viruses for the research community.



## METHODS



Viral DNA was extracted from virus strains within NCPV, using a Maxwell 16 (Promega) automated system. Whole genome sequencing was performed on the Illumina MiSeq platform at the Wellcome Trust Sanger Institute. Assembled and annotated genome sequences will be freely publicly accessible via a comprehensive web-based **Biological Resource Information** Centre (BRIC).

Data mining the sequences of individual genomes helps to catalogue the genes encoded in a particular strain and is a vital step for in-depth characterisation studies. Sequencing of multiple isolates, strains or species enables understanding of the factors responsible for varying virulence using comparative genomics.

#### Figure 1. Extraction and sequencing technology: Maxwell 16 and Illumina MiSeq

R	ESU	LTS

EMBL-EBI	Services Research Training About us
European Nucleotide Archive	Examples: BN000065, histone Advanced Sequence
Home Search & Browse Submit & Update Software About ENA Support	
Study: PRJEB12890	<u>Contact Helpdesk</u> 🔤
Sequencing_extracted_nucleic_acid_from_the_National_Collection_of_Pathogenic_Viruses	
View: Project XML Study XML	Download: Project XML Study XML
Name Sequencing_extracted_nucleic_acid_from_the_National_Collection_of_Pathogenic_Viruses The Wellcome T	
Secondary accession(s) ERP014415	
Description This project aims to sequence the genomes of RNA viruses in the National Collection of Pathogenic Viruse characterised, authenticated human pathogenic viruses in a secure facility, and supplies the agents or no sequenced within this project will be a valuable additional resource that will also be made available to the	ucleic acids derived from them to the scientific community. The genomes
Navigation Read Files Portal Attributes	
🚫 Bulk Download Files 📐 (If the downloader app doesn't open, please try using Firefox to launch it.)	
Download: 1 - 139 of 139 results in TEXT	

# Nucleic acid from 196 ACDP Hazard Group 2 viruses has been extracted and sequenced.

Sequences from 139 RNA virus strains have been uploaded to the European Nucleotide Archive, under Study PRJEB12890 (Figure 2).

Files can be downloaded and analysed by users in FASTQ or

Galaxy formats.

essibility   Contact Us   Site Map   Help			👤 Sig	n In   Register Unit	ted Kingdom (chang
				۲ 🛒 ۲	'our Cart 0 Items
K Public Healt	in England				Search
Home About Us News	Products Ordering	Services Technical Support	Deposit with us	Quality Contact U	s PHE Websit
Culture Collection	S				
You are here: Home > Products > Virus S	earch > Virus Detail				
Products	Virus Detail				
Bacteria and Mycoplasmas LENTICULE Discs	Please Check Stock Avai	lability Prior to Ordering		Formats	Price
LENTICULE Discs Cell Lines and Hybridomas ECACC DNA Products	Please Check Stock Avai	lability Prior to Ordering	NCPV	Formats	Price £263.00
LENTICULE Discs Cell Lines and Hybridomas ECACC DNA Products RNA from Cell Lines cDNA		*	NCPV National Collection of Pathogenet Videos Red by Public Health England	🖉 Ampoule (Virus)	
LENTICULE Discs Cell Lines and Hybridomas ECACC DNA Products RNA from Cell Lines		Орига	National Calection of Path openic Vioses	<ul> <li>Ampoule (Virus) In Stock</li> <li>Virus RNA In Stock</li> </ul>	£263.00
LENTICULE Discs Cell Lines and Hybridomas ECACC DNA Products RNA from Cell Lines cDNA Fungi Search	Print datasheet	Орига	National Calection of Path openic Vioses	<ul> <li>Ampoule (Virus) In Stock</li> <li>Virus RNA In Stock</li> </ul>	£263.00 £314.00 Add to cart
LENTICULE Discs Cell Lines and Hybridomas ECACC DNA Products RNA from Cell Lines cDNA Fungi Search Virus Search	Print datasheet  Virus Collection: H  Catalogue No.: Virus Name: ICTV Species Name: Also Known As: Family:	uman Coronavirus 0310051v Human Coronavirus Alphacoronavirus 229E HCoV-229E Coronaviridae	National Calection of Path openic Vioses	Ampoule (Virus)     In Stock     Virus RNA     In Stock     Quantity	£263.00 £314.00 Add to cart Prices
LENTICULE Discs Cell Lines and Hybridomas ECACC DNA Products RNA from Cell Lines cDNA Fungi Search Virus Search Virus Search	Print datasheet  Virus Collection: H  Catalogue No.: Virus Name: ICTV Species Name: Also Known As:	uman Coronavirus 0310051v Human Coronavirus Alphacoronavirus 229E HCoV-229E	National Calection of Path openic Vioses	Ampoule (Virus) In Stock      Virus RNA In Stock      Quantity      Click for Delivery      Credit Account orders as	£263.00 £314.00 Add to cart Prices

Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	files	FASTQ files (Galaxy)	Submitted files (FTP)	Submitted files (Galaxy)	NCBI SRA file (FTP)	NCBI SRA file (Galaxy)	CRA Inde files (FT
PRJEB12890	SAMEA51860668	ERS1507262	ERX1934202	ERR1873677	11981	Feline calicivirus strain F9	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	<u>CRAM File</u> 1	<u>CRAM File</u> 1			<u>CRA</u> File
PRJEB12890	SAMEA51861418	ERS1507263	ERX1934203	ERR1873678	<u>11137</u>	<u>Human</u> coronavirus 229E	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	<u>CRAM File</u> <u>1</u>	<u>CRAM File</u> <u>1</u>			<u>CRA</u> File
PRJEB12890	SAMEA51862168	ERS1507264	ERX1934204	ERR1873679	<u>185892</u>	<u>Human</u> rhinovirus A10	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	<u>CRAM File</u> <u>1</u>	<u>CRAM File</u> <u>1</u>			<u>CR</u> 4 File
PRJEB12890	SAMEA51862918	ERS1507265	ERX1934205	ERR1873680	<u>185892</u>	<u>Human</u> rhinovirus A10	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	<u>CRAM File</u> 1	<u>CRAM File</u> 1			<u>CR</u> File
PRJEB12890	SAMEA51863668	ERS1507266	ERX1934206	ERR1873681	<u>39767</u>	<u>Human</u> rhinovirus A11	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	<u>CRAM File</u> 1	<u>CRAM File</u> 1			<u>CR</u> File
PRJEB12890	SAMEA51864418	ERS1507267	ERX1934207	ERR1873682	<u>39767</u>	<u>Human</u> rhinovirus A11	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	<u>CRAM File</u> 1	<u>CRAM File</u> 1			<u>CR/</u> File
PRJEB12890	SAMEA51865168	ERS1507268	ERX1934208	ERR1873683	<u>12131</u>	Rhinovirus B14	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	CRAM File	CRAM File			<u>CR</u> File
PRJEB12890	SAMEA51865918	ERS1507269	ERX1934209	ERR1873684	<u>12131</u>	Rhinovirus B14	Illumina MiSeq	PAIRED	<u>File 1</u> File 2	File 1 File 2	<u>CRAM File</u> <u>1</u>	<u>CRAM File</u> <u>1</u>			<u>CR</u> File
PRJEB12890	SAMEA51866668	ERS1507270	ERX1934210	ERR1873685	<u>31708</u>	<u>Human</u> rhinovirus A16	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	CRAM File 1	CRAM File 1			<u>CR</u> File
PRJEB12890	SAMEA51867418	ERS1507271	ERX1934211	ERR1873686	150904	<u>Human</u> rhinovirus B17	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2	CRAM File	CRAM File			<u>CR</u> File

#### Figure 2. ENA uploaded sequences

The BRIC will capture all the information regarding the virus strains, including isolation and deposit details (where known), safety data sheets, production and quality control results. This will be combined with data from external sources, such as sequences from ENA and NCBI, and publications.

Biosafety Responsibility:	It is the responsibility of the customer to ensure that their facilities comply with biosafety regulations for their own country	Co Print datasheet
Product Doce	uments	Related Links
MSDS Human Coron	evinic 220E	NCPV Home
MSDS for frozen live		Search Viruses
		Browse Viruses
Note: Links open in a nev	v window	New viral strains
		NCPV News
🔋 Related Prod	ucte	Culture Collections News
Related 1100	ucto	Exotic Viruses
You might be interested in	n the following related products:	Viral Nucleic Acids
ECACC General Cell	Collection - MRC-5 pd19	Biohazard Risk Assessment Form - Viruses
		Virus deposit form
very effort is made to ensu ulture Collections cannot l	resent deposits of cultures from world-wide sources. While are details distributed by Culture Collections are accurate, be held responsible for any inaccuracies in the data supplied. wen act only as a guide and Culture Collections does not	NCPV Brochure
	ber stated will be the passage number received by the	
Cultures supplied by Culture	e Collections are for research purposes only. Enquiries	

#### Figure 3. Culture Collections catalogue listing

All the sequenced strains are publically available from the National Collection of Pathogenic Viruses (Figure 3).

CPV lational Collection of Pathoaenic Viruse Operated by Public Health England



Cell lines for culturing viruses can be obtained from the European Collection of Authenticated Cell Cultures (ECACC), another Culture Collection of Public Health England.

### DISCUSSION

Next Generation Sequencing is an emerging technology that is becoming more widely used in pathogen microbiology. applications have become more

## ACKNOWLEDGEMENTS

The authors are grateful to ECACC for the provision of quality cell cultures for propagation of virus cultures.

diverse, including real-time epidemiology, clinical diagnosis, microbe discovery, taxonomic classification, quality control of vaccines, tracking adaptation and evolution and understanding the roles of viral genes in infection.

Online availability of the viral genomes, coupled with the biological availability of the virus strains from NCPV, will enable investigation of the interactions between genotype and phenotype of known, emerging and novel viral pathogens.

Specific research questions can be addressed using the historical depth of the Collection:

- compare isolates from before and after vaccine introductions or changes in uptake
- examine changes in vector biology for vector-borne viruses
- examine the effect of serial passage, as an analogy to transmission during an outbreak •
- identify co-infections
- examine the effect of culture in alternative cell lines

Future work will focus on sequencing a further 304 viruses, including Hazard Group 3 organisms.

Researchers are invited to deposit newly isolated virus strains into NCPV free of charge, and have the viral genomes sequenced and added to the BRIC.

The Culture Collections IT team are leading on the design of the BRIC.

## CONTACTS

www.phe-culturecollections.org.uk

#### https://www.ebi.ac.uk/ena/data/view/PRJEB12890

