

RABV-GLUE

Rabies virus genome resource

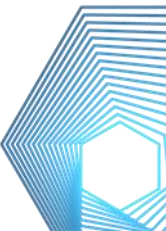
Josh Singer

MRC-University of Glasgow Centre for Virus Research

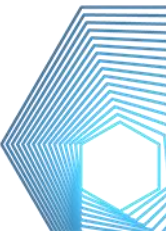
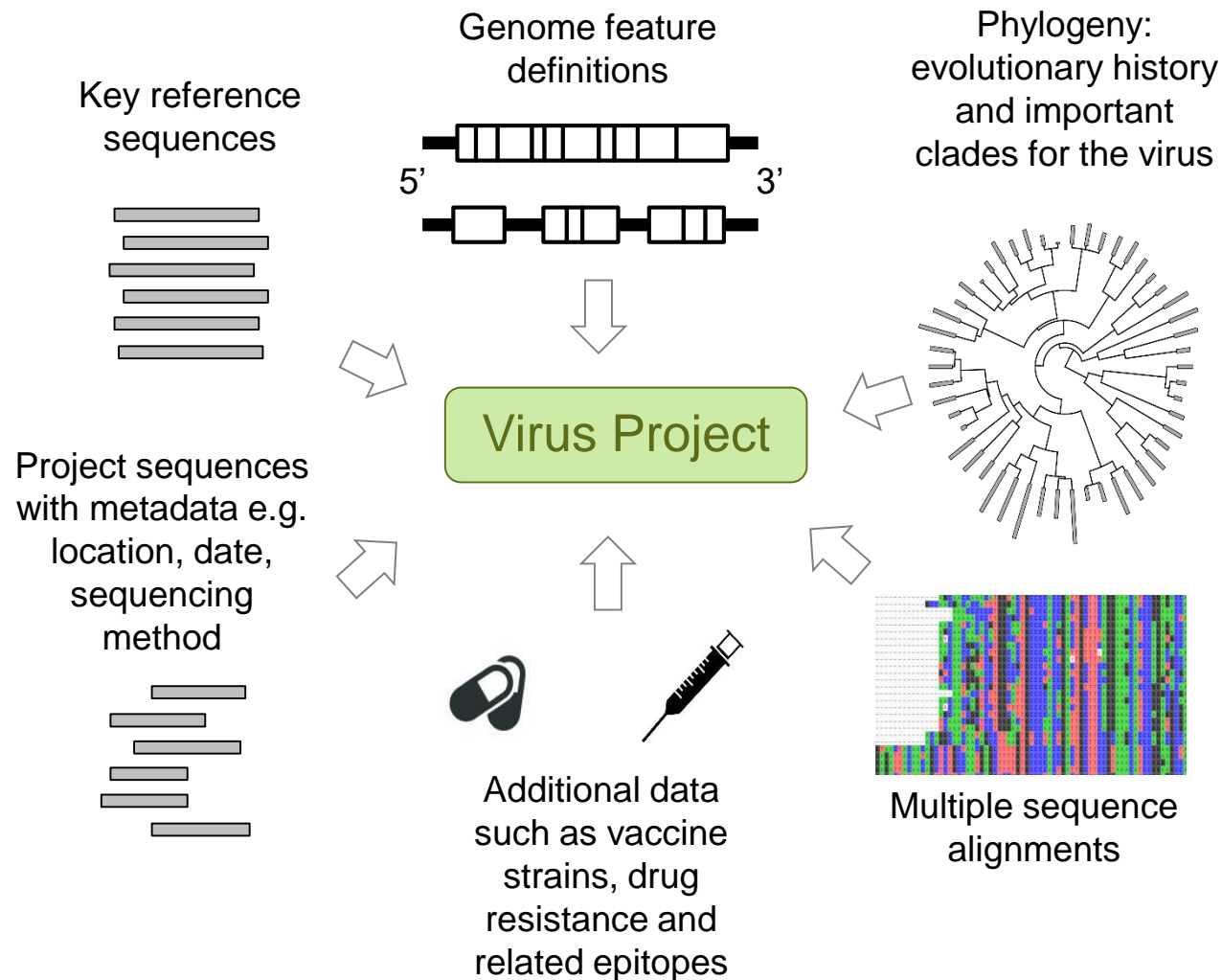
22nd November 2018

What is GLUE?

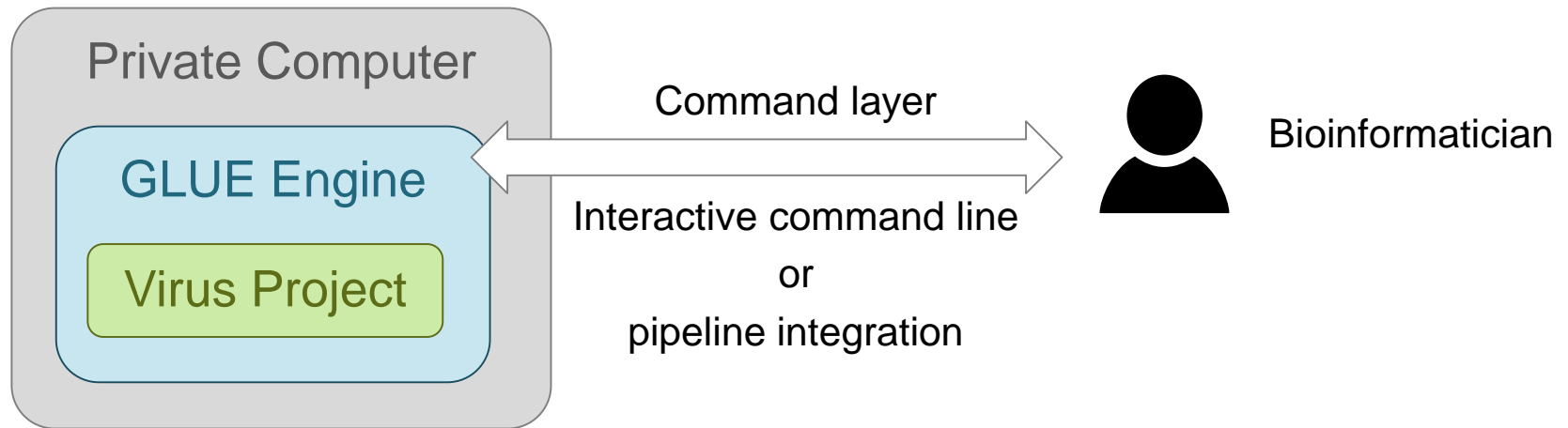
- Bioinformatics software framework
- Focus: analysis of virus genome variation
- Design:
 - Data-centric
 - Organises sequences along evolutionary lines
 - Multiple-sequence alignments are central
- Flexible:
 - Extensible data schema
 - Various built-in analysis modules
 - Scriptable
 - Customisable for different viruses



GLUE projects: virus-specific linked datasets



Offline GLUE for bioinformatics



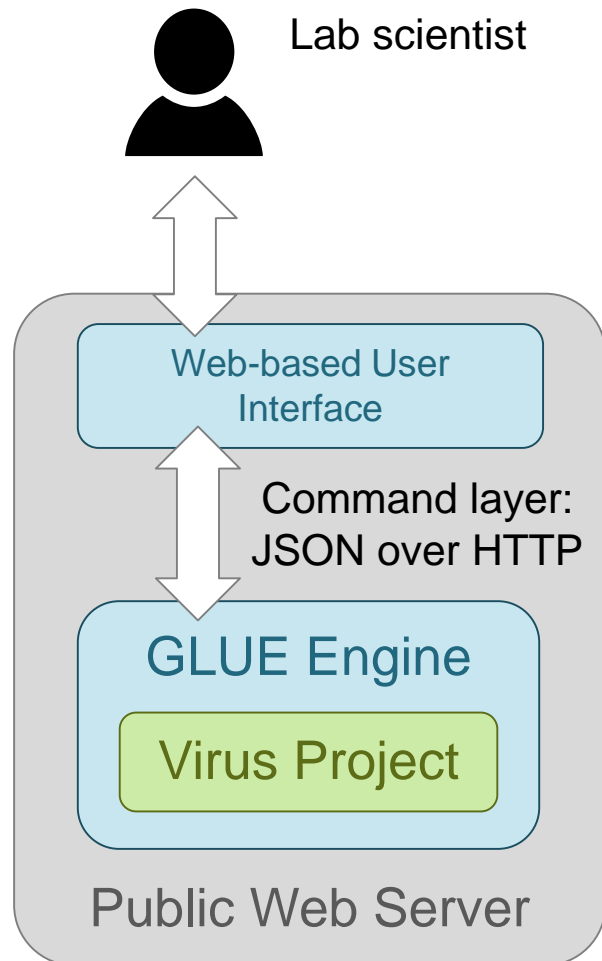
```
GLUE> amino-acid frequency -c -w
"sequence.host_species =
'Sus scrofa'" -r REF_MASTER_M73218 -fMT -l 60 65
=====+=====+=====+=====+
| codon | aminoAcid | numMembers | pctMembers |
+-----+-----+-----+-----+
| 60    | E         | 14         | 100.00     |
| 61    | V         | 14         | 100.00     |
| 62    | L         | 10         | 71.43      |
| 62    | F         | 4          | 28.57      |
| 63    | W         | 14         | 100.00     |
| 64    | N         | 14         | 100.00     |
| 65    | H         | 14         | 100.00     |
+-----+-----+-----+-----+
```

Example uses:

- Complex data queries
- Phylogenetics
- Automated genotyping
- Sequence interpretation
- NGS data analysis



GLUE-based web applications



rabv.glue.cvr.ac.uk

RABV-GLUE

A Sequence Data Resource for Rabies Virus

Rabies virus (RABV) is the virus that attacks the nervous system, causing rabies disease. In the wild rabies virus has been found to infect many mammalian species. Rabies results in encephalitis in humans and other mammals and is almost always fatal in unvaccinated humans.

The rabies virus is the type species of the Lyssavirus genus of the Rhabdoviridae family. These viruses are enveloped and have a single stranded, negative-sense RNA genome.

RABV-GLUE is a data-centric bioinformatics resource which organises RABV genome sequence data along evolutionary lines. RABV-GLUE aims to leverage new and existing RABV sequences in order to improve our understanding of the epidemiology and pathology of RABV.

The web version of RABV-GLUE can be used for basic analysis. An offline version of the resource can be installed by bioinformaticians and used for more advanced work.

Web site highlights

- An analysis tool providing genotyping, analysis and visualisation of submitted FASTA sequences.
- A database of RABV sequences and metadata from NCBI updated daily and arranged into major and minor clades.
- Pre-built multiple-sequence alignments of NCBI sequences, which may be downloaded in user defined sections. Use the clade tree to select a clade such as Africa-2. Then under the Download button select "Alignment".

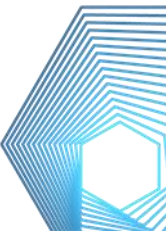
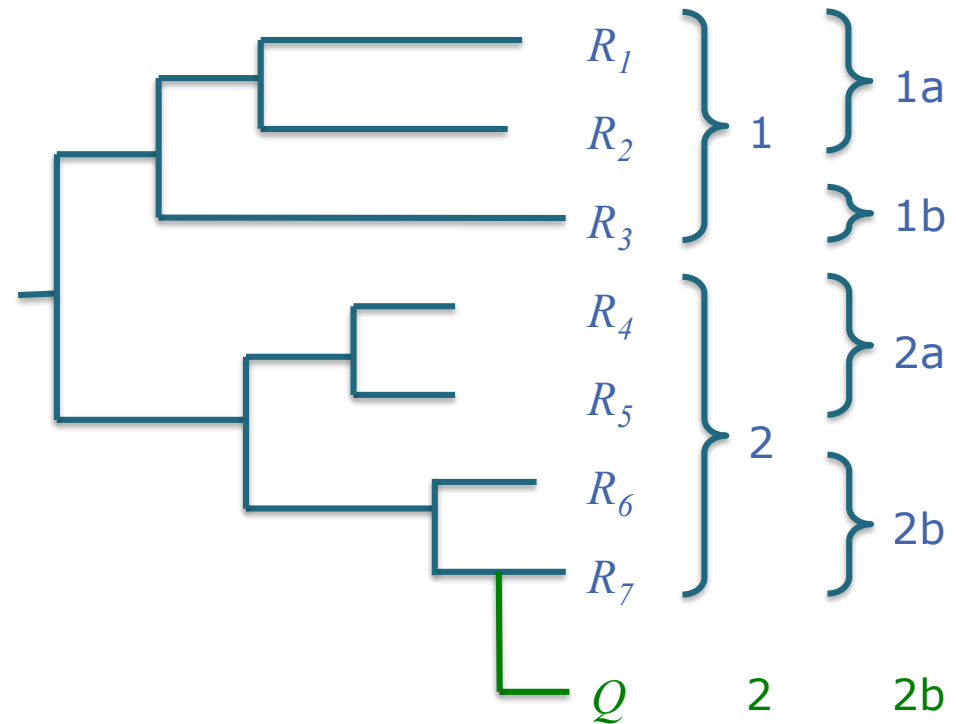
RABV-GLUE is based on the GLUE software framework. GLUE and RABV-GLUE are developed and maintained by the MRC-University of Glasgow Centre for Virus Research, the University of Glasgow Institute of Biodiversity, Animal Health and Comparative Medicine and the UK Government Animal and Plant Health Agency.

Please note this is beta software, still undergoing development and testing before its official release.

Logos: CVR, University of Glasgow, Institute of Biodiversity Animal Health and Comparative Medicine, Animal & Plant Health Agency.

GLUE Maximum-likelihood clade assignment

- Assignment to a clade (e.g. genotype) is a necessary first step in sequence interpretation
- Query sequence added as a new row to the reference alignment
- A maximum-likelihood technique then places the new sequence within a reference phylogeny



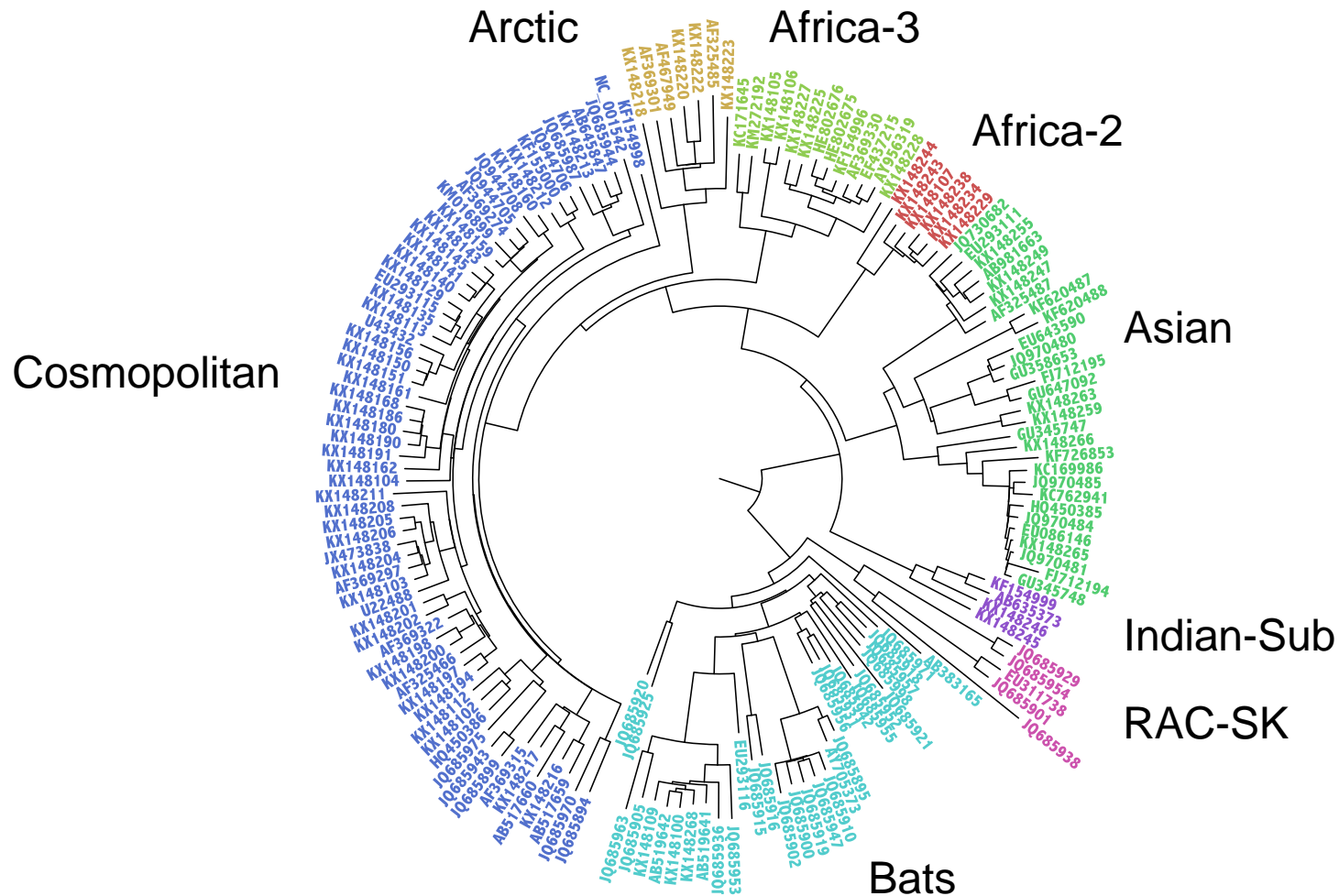
RABV-GLUE

- Clade definitions:
 - 8 major clades, 56 minor clades
 - Sourced from Troupin et al. 2016 [1] and Kuzmin et al. 2012 [2]
- Reference sequences:
 - Initial set from literature of 509
 - Filtered down to representative set of 161
 - 5 CDS regions annotated
- Reference phylogeny generated using RAxML

1. Troupin et al. (2016) Large-Scale Phylogenomic Analysis Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts, PLoS Pathog 12(12): e1006041.
2. Kuzmin et al. (2012) Molecular Inferences Suggest Multiple Host Shifts of Rabies Viruses from Bats to Mesocarnivores in Arizona during 2001–2009, PLoS Pathog 8(6): e1002786.

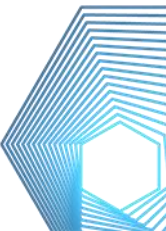


MRC-University of Glasgow Centre for Virus Research



NCBI-RABV-GLUE

- Builds on the RABV-GLUE core project
- Includes all RABV sequences from GenBank, synchronised daily
- Extracts collection date, country and host species metadata.
- Each sequence is genotyped
- Alignments are pre-computed



RABV-GLUE web application (1)

- Filter GenBank sequences:

Coverage of Genome Region	L	is at least	90	%	Remove filter
Global Region	matches		South America		Remove filter
Earliest Collection Year	>=		2005		Remove filter

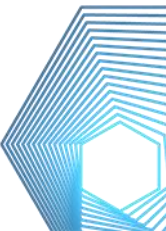
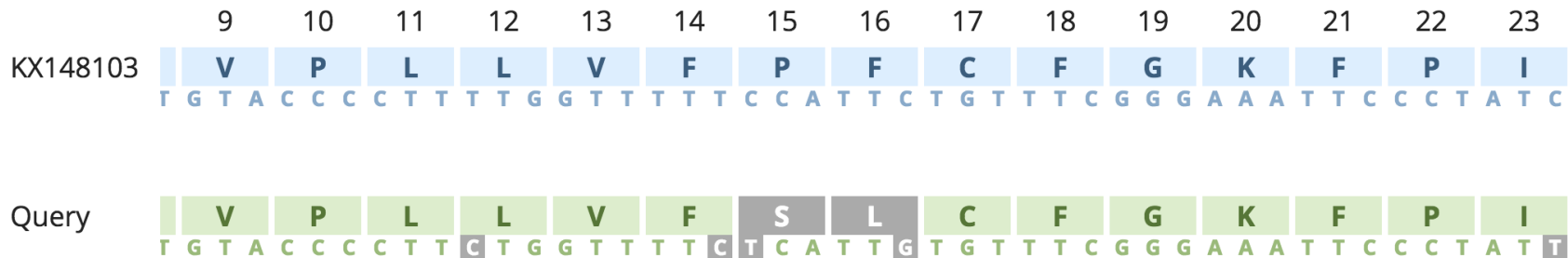
- Download unaligned sequences, metadata or pre-computed NT or AA alignments of specific genome regions

AL_Cosmopolitan_AF1b.ncbi-curated.KR906740	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906741	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906742	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906743	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906744	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906745	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906746	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906747	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906748	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906749	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906750	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906751	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906752	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906753	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906754	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906755	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906756	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V
AL_Cosmopolitan_AF1b.ncbi-curated.KR906757	R	P	I	P	D	A	C	R	A	A	Y	N	K	M	A	G	D	P	R	Y	E	E	S	L	H	N	P	Y	P	D	Y	H	W	L	R	T	V	K	T	T	K	E	S	L	V	I	I	S	P	S	V	A	D	L	D	P	Y	D	K	S	L	H	S	R	V	F	T	G	G	K	C	S	G	I	T	V



RABV-GLUE web application (2)

- Genotyping of submitted FASTA file
- Visualise genome



RABV-GLUE for SPEEDIER

- Store SPEEDIER genome sequence data in a project-specific extension
- Extend database schema to capture detailed time and location data
- Develop a fine-grained lineage genotyping system to detect incursions
- Integrate genomic data into SPEEDIER web applications