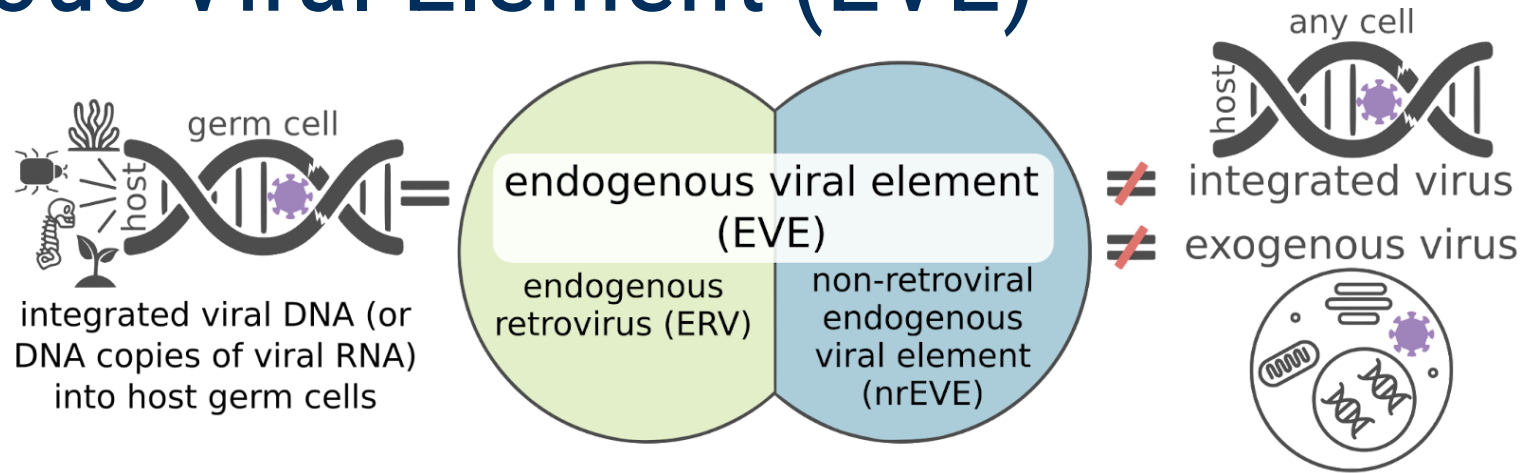

Endogenous Viral Elements in the Shadows: challenges in data availability and accessibility

39th TBI Winterseminar in Bled

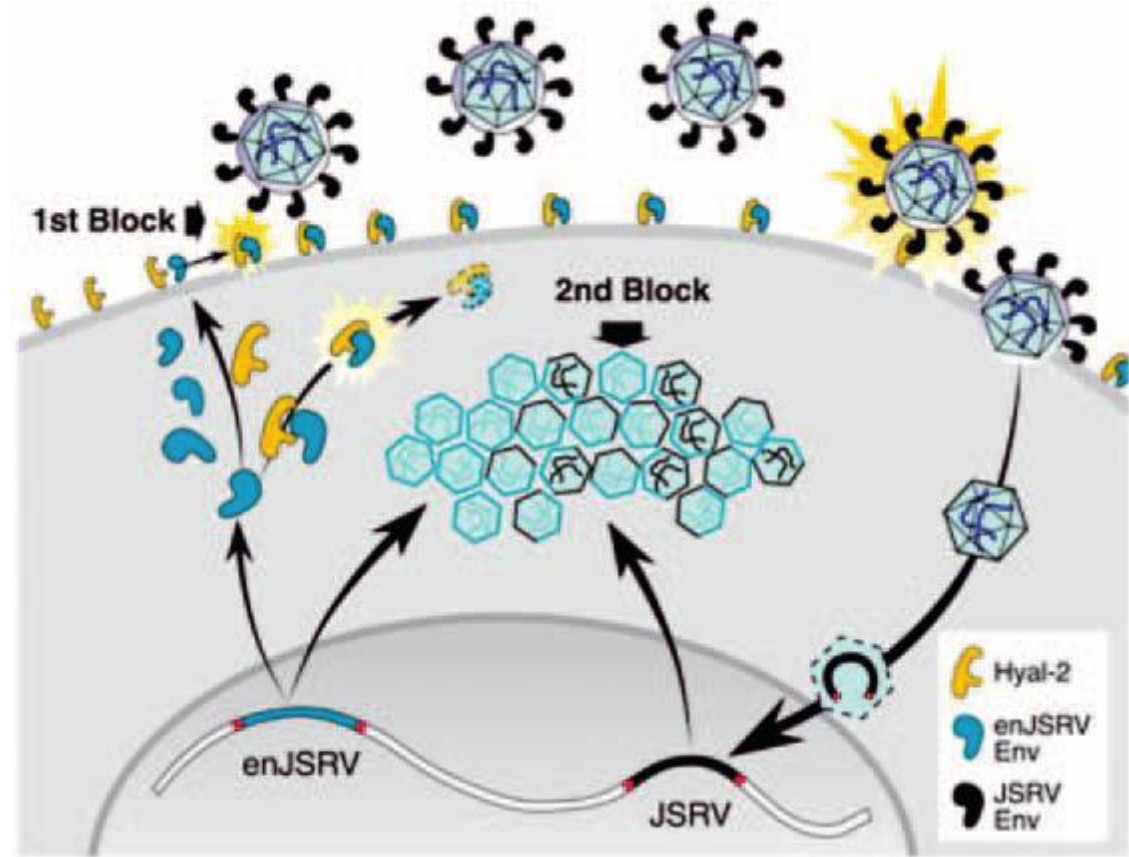
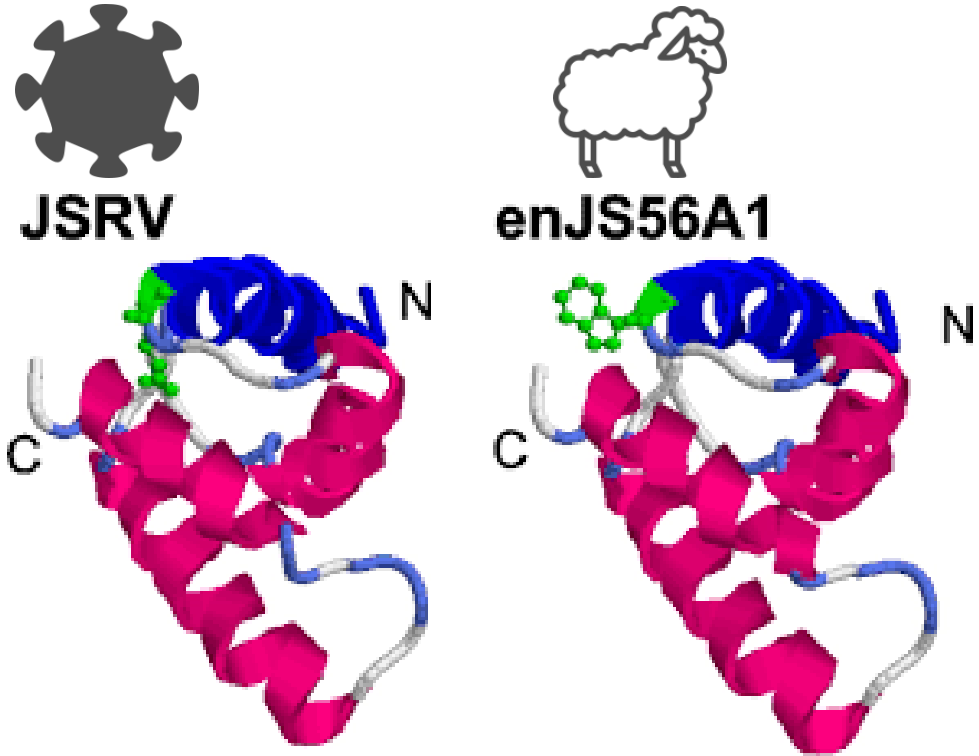


Endogenous Viral Element (EVE)



	Retroviruses	DNA viruses/ RNA viruses
Replication location	nucleus (integration in host genome)	nucleus cytoplasm
Steps to become an EVE	Germ cells	Integration Germ cells Transport Translation Integration Germ cells

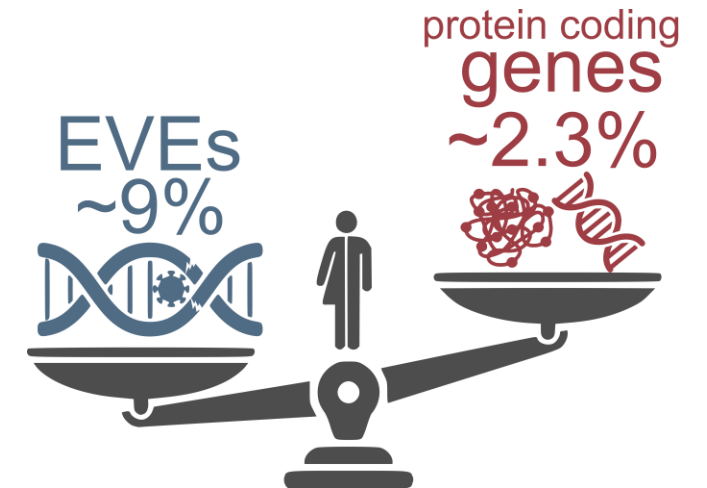
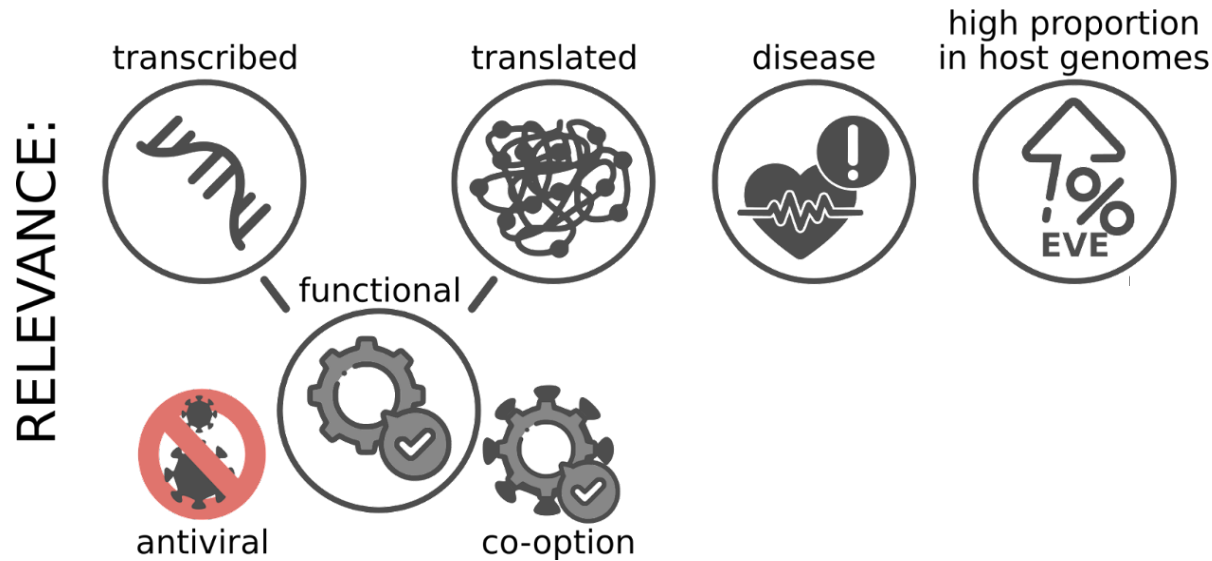
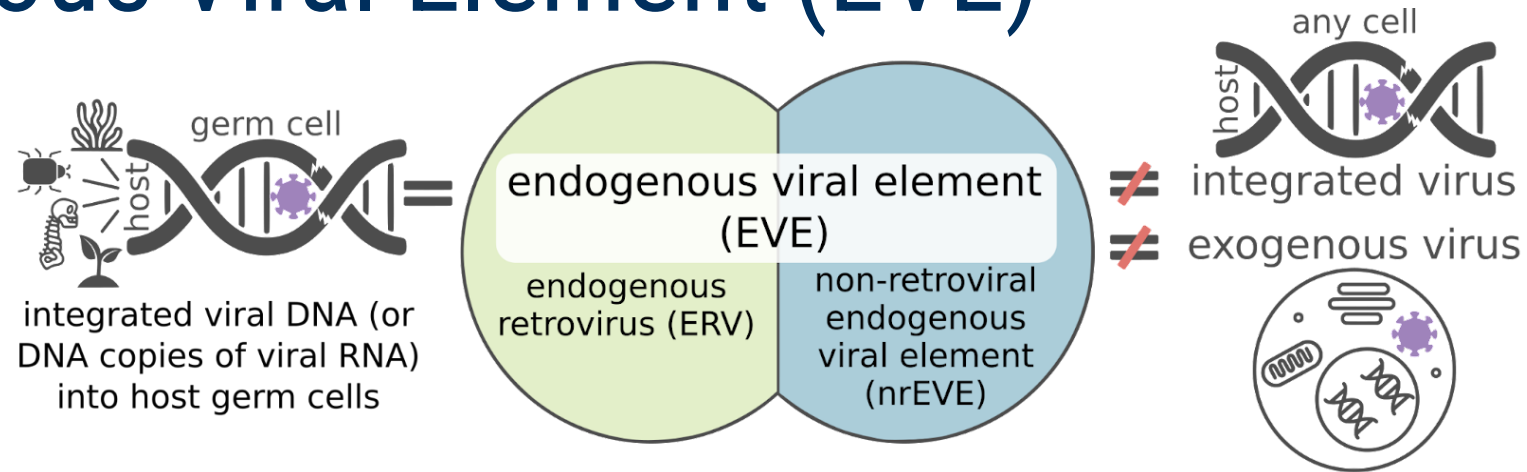
Example: Endogenous retroviruses 'Jaagsiekte sheep retrovirus' (JSRV)



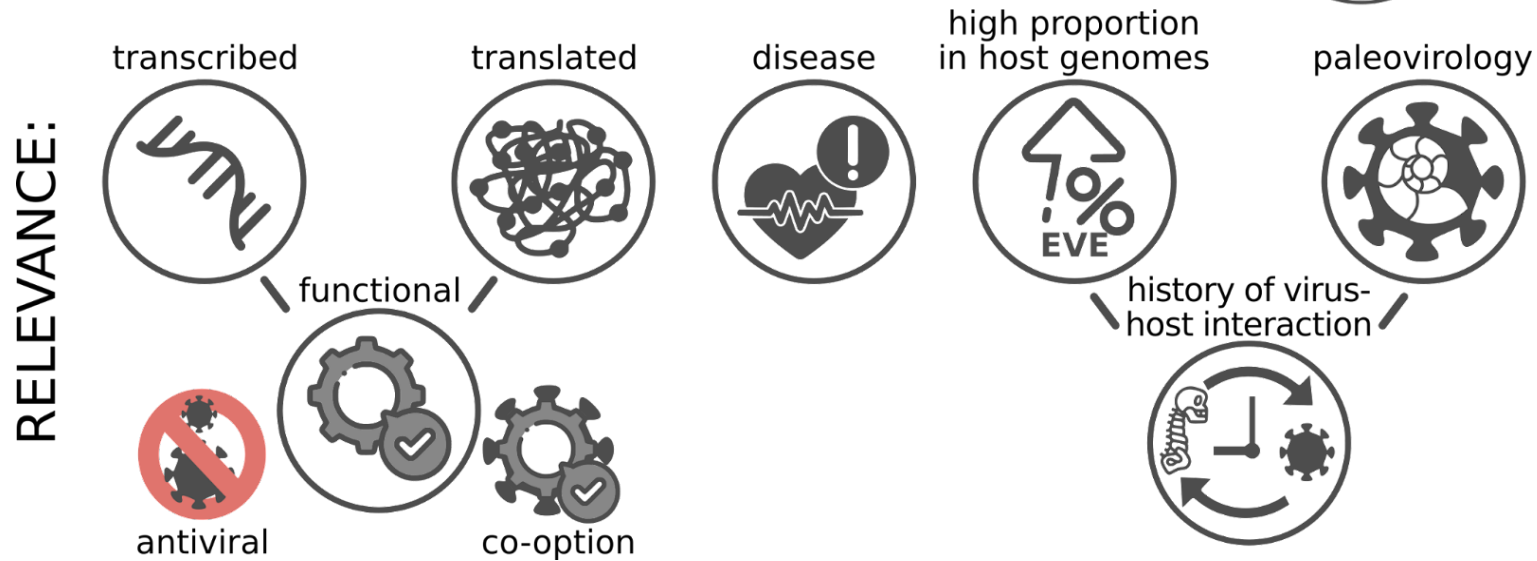
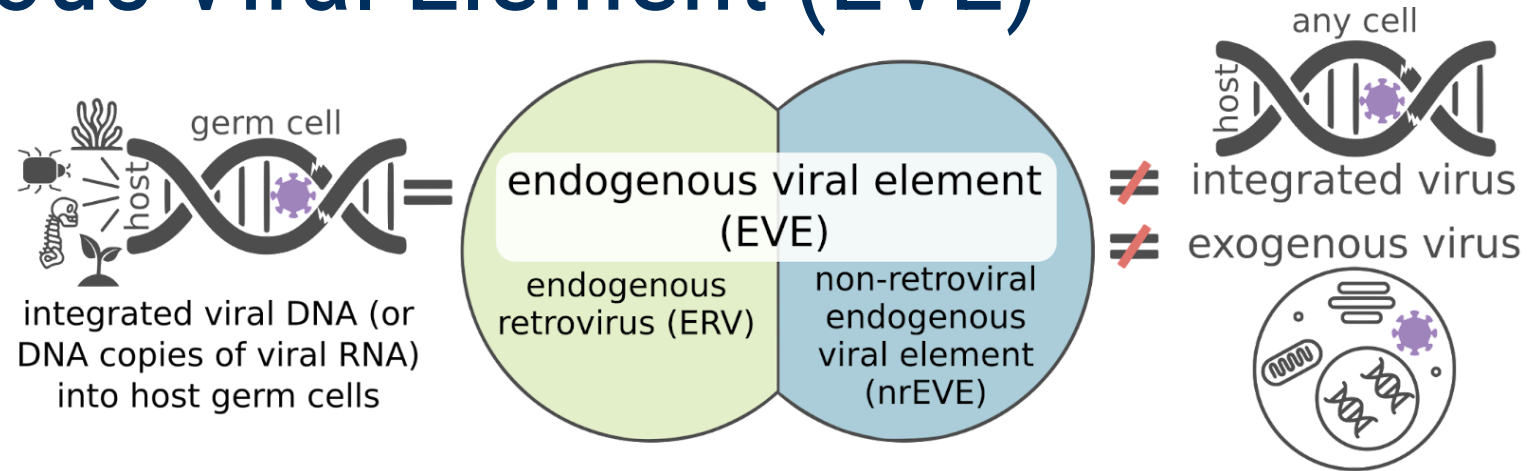
Mura, Manuela, et al. "Late viral interference induced by transdominant Gag of an endogenous retrovirus." *Proceedings of the National Academy of Sciences* 101.30 (2004): 11117-11122

https://www.researchgate.net/figure/The-two-sites-of-enJSRV-block-The-endogenous-JSRV-blocks-genital-invasion-by-the_fig2_38054002

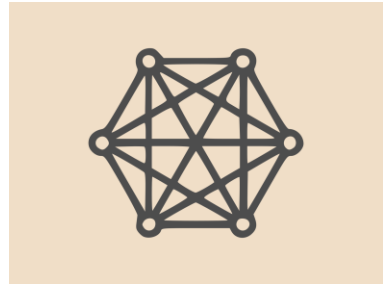
Endogenous Viral Element (EVE)



Endogenous Viral Element (EVE)



Availability: Complexity



- ERVs: conserved retrovirus-specific sequences



- nrEVEs: no rules

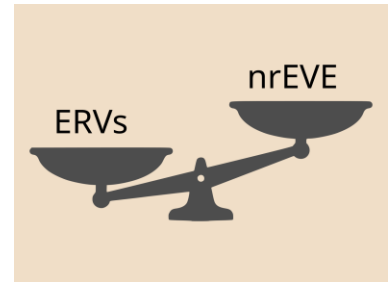


- slower evolutionary rate than their exogenous virus (functional vs. non functional)

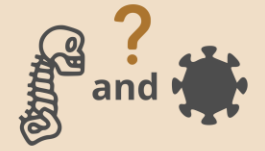
Availability: Unbalanced

ERVs:

- clear structure → less diverse (easier to detect?)
- higher abundance in host genomes (more integration events?)
- more relevant (disease) & were discovered earlier (higher visibility?)



Availability: Taxonomic Assignment



Availability: No good EVE Tool



ERVs:

- tool: RepeatMasker

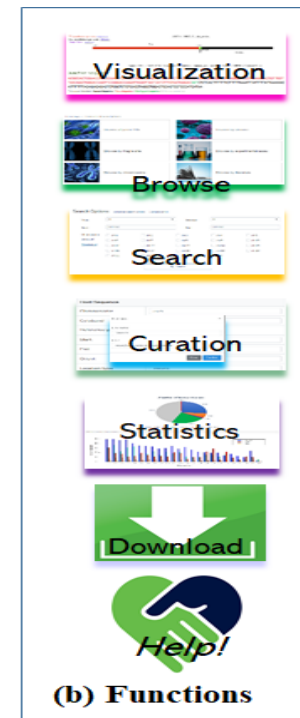
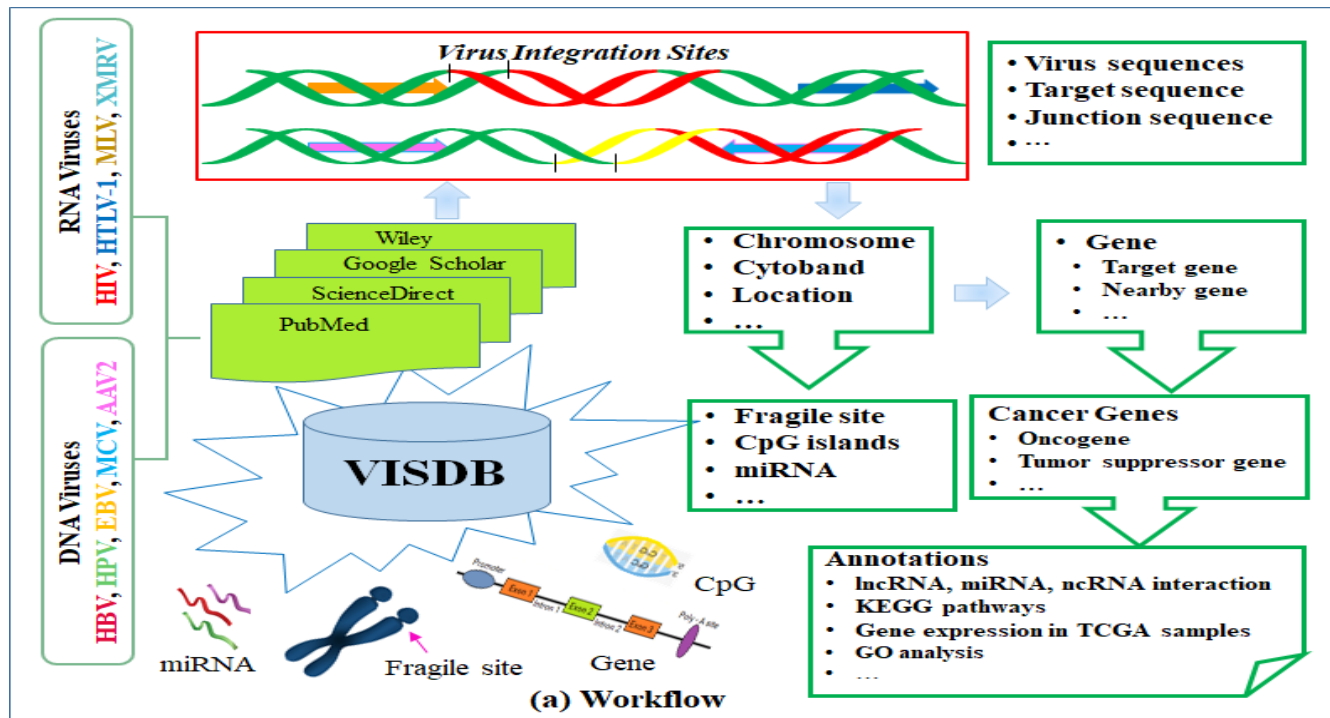
nrEVes:

- no tool
- no standards: everyone is doing slightly different stuff

Availability: lack of Awareness



Example: tool for viral integrations sites



<https://bioinfo.uth.edu/VISDB/index.php/homepage>

Availability Solutions: Tools for EVE identification on...

genomic level

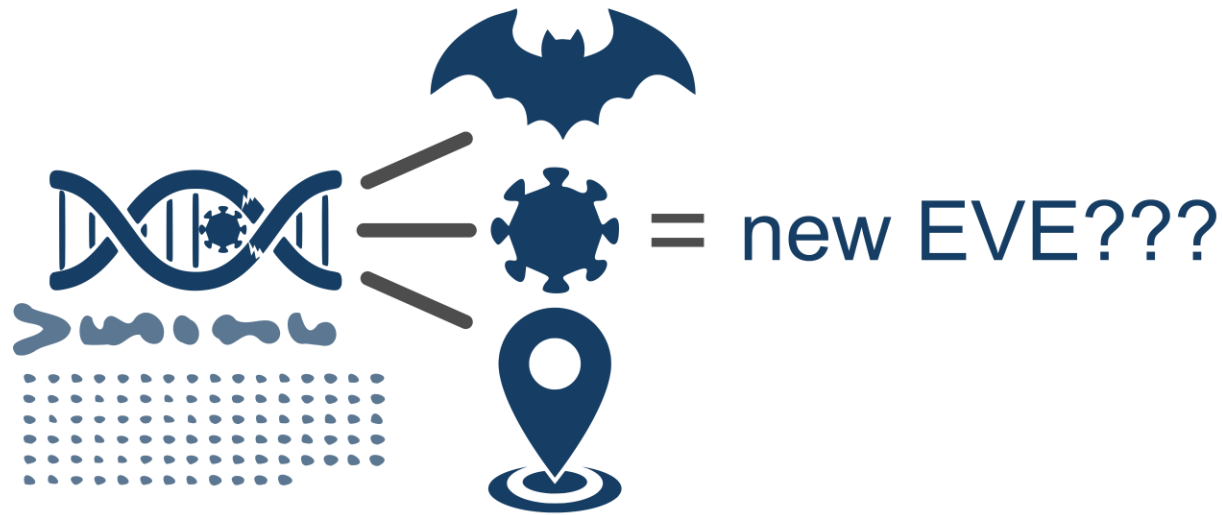


transcriptomic level



→ establishing a common standard

Simple research questions



annotation?

few EVE annotations in host genomes



blast?

few sequences



databases?

no EVE database

no EVEs in general database



paper?

no working keyword search

no standards



Simple research questions

[← Edit Search](#)
[Save Search](#)
[Search Summary ▾](#)
[How to read this report?](#)
[BLAST Help Videos](#)
[Back to Traditional Results Page](#)

Job Title Nucleotide Sequence
RID [WMYHNEKB016](#) Search expires on 02-14 03:06 am [Download All ▾](#)
Program BLASTN [Citation ▾](#)
Database nt [See details ▾](#)
Query ID IcllQuery_2562995
Description None
Molecule type dna
Query Length 696
Other reports [Distance tree of results](#) [MSA viewer](#) [?](#)

Filter Results
Organism only top 20 will appear exclude

[+ Add organism](#)
Percent Identity to
E value to
Query Coverage to

[Descriptions](#) | [Graphic Summary](#) | [Alignments](#) | [Taxonomy](#)

Sequences producing significant alignments
[Download ▾](#)
[Select columns ▾](#)
[Show 100 ▾](#) [?](#)

select all 8 sequences selected
[GenBank](#)
[Graphics](#)
[Distance tree of results](#)
[MSA Viewer](#)

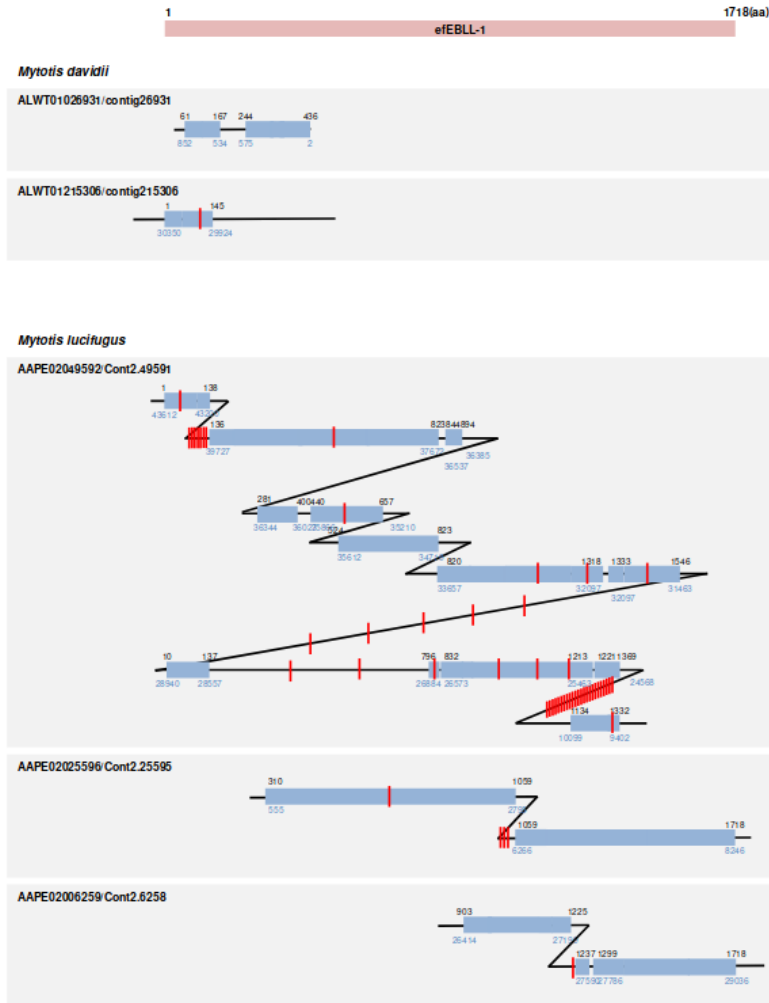
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Pipistrellus pipistrellus genome assembly, chromosome: 5	Pipistrellus pipi...	1024	1024	100%	0.0	93.32%	94929986	LR862361.1
<input checked="" type="checkbox"/>	Myotis daubentonii genome assembly, chromosome: 13	Myotis daubent...	420	420	72%	3e-112	82.08%	64611165	OY725372.1
<input checked="" type="checkbox"/>	PREDICTED: Eptesicus fuscus uncharacterized LOC114227668 (LOC114227668), transcript variant X4... Eptesicus fuscus	Eptesicus fuscus	187	187	14%	3e-42	100.00%	750	XR_003613742.2
<input checked="" type="checkbox"/>	PREDICTED: Eptesicus fuscus uncharacterized LOC114227668 (LOC114227668), transcript variant X3... Eptesicus fuscus	Eptesicus fuscus	187	187	14%	3e-42	100.00%	2088	XR_003613746.2
<input checked="" type="checkbox"/>	PREDICTED: Eptesicus fuscus uncharacterized LOC114227668 (LOC114227668), transcript variant X1... Eptesicus fuscus	Eptesicus fuscus	187	187	14%	3e-42	100.00%	2386	XR_003613743.2
<input checked="" type="checkbox"/>	PREDICTED: Eptesicus fuscus uncharacterized LOC114227668 (LOC114227668), transcript variant X2... Eptesicus fuscus	Eptesicus fuscus	185	185	14%	1e-41	100.00%	810	XR_003613744.2
<input checked="" type="checkbox"/>	PREDICTED: Myotis brandtii uncharacterized LOC102258198 (LOC102258198), transcript variant X2, n... Myotis brandtii	Myotis brandtii	165	165	14%	2e-35	96.04%	622	XR_001359894.1
<input checked="" type="checkbox"/>	PREDICTED: Myotis brandtii uncharacterized LOC102258198 (LOC102258198), transcript variant X1, n... Myotis brandtii	Myotis brandtii	165	165	14%	2e-35	96.04%	489	XR_314510.2

Simple research questions

Database	Problems
ERE	inaccessible
FabriEVEs	inaccessible
gEVE	ORF-related, old (2016)
DbHERV-Res	ERVs, old (2017)
pEVE database	DNA-EVEs, old (2020)
HERVdb	Human ERVs, old (2021)

Supplementary Table 3. tBLASTn screening for EBLNs in vesper bats.

Species	Accession number	Identity (%)	Query		Hit		E-value	Name of EBL	
			Start	End	Start	End			
<i>Myotis lucifugus</i>	AAPE02006259.1	49.3	913	981	26414	26620	4.15E-43	myEBLL-1	
		39.7	982	1136	26619	27077	4.15E-43		
		51.3	1142	1180	27083	27199	4.15E-43		
AAPE02025596.1	AAPE02025596.1	46.9	326	1069	570	2795	0	myEBLL-2	
		40.5	1069	1147	6266	6499	1.89E-24		
		31.1	1152	1312	6510	6983	1.89E-24		
AAPE02049592.1	AAPE02049592.1	52.9	152	202	39727	39575	6.99E-25	myEBLL-3	
		32.7	224	417	39503	38913	6.99E-25		
		50	534	601	38567	38364	4.56E-66		
AAPE02020529.1	AAPE02020529.1	42.1	983	1058	2935	2714	1.42E-26	myEBLL-4	
		31	1062	1296	2718	2038	1.42E-26		
		45.9	392	465	24713	24495	3.66E-28	Inter-family	
AAPE02024702.1	AAPE02024702.1	43.8	487	595	24433	24113	3.66E-28		
		42.4	913	1139	11214	11888	1.97E-41	myEBLL-1	
		45	1141	1180	11890	12009	1.97E-41		
ANKR01212491.1	ANKR01212491.1	46.1	326	1073	41559	43796	0	myEBLL-2	
		ANKR01212492.1	41.3	1070	1149	1625	1861	1.98E-24	
		32.1	1149	1313	1858	2343	1.98E-24		
ANKR01204701.1	ANKR01204701.1	50.8	146	204	1592	1416	5.14E-57	myEBLL-3	
		36.4	223	423	1353	745	5.14E-57		
		44.4	427	534	739	416	5.14E-57		
ANKR01204699.1	ANKR01204699.1	40.2	534	636	416	84	5.14E-57		
		47.9	487	534	40610	40467	5.99E-28		
		45.3	531	615	40477	40220	5.99E-28		
ANKR01204699.1	ANKR01204699.1	52.2	618	663	40214	40077	5.99E-28		
		33.3	357	424	38478	38272	1.43E-73		
		46.3	427	534	38269	37946	1.43E-73		
ANKR01204699.1	ANKR01204699.1	39.4	534	838	37946	37026	1.43E-73		



c

Table S1. Newly identified endogenous borna-like N (EBLN) elements

Host species	Accession number	BLAST E-value
Bonobo (<i>Pan paniscus</i>)	AJFE01073334	3.00E-74
	AJFE01052936	8.00E-39
	AJFE01104638	1.00E-30
Olive baboon (<i>Papio anubis</i>)	AHZZ01069395	4.00E-60
	AHZZ01098311	2.00E-35
	AHZZ01037919	9.00E-33
Crab-eating macaque (<i>Macaca fascicularis</i>)	AEHL01267442	9.00E-59
	CAEC01365824	1.00E-58
	AEHL01162971	1.00E-35
Black-capped squirrel monkey (<i>Saimiri boliviensis boliviensis</i>)	CAEC01159547	2.00E-35
	CAEC01649023	4.00E-33
	AEHL01409148	2.00E-32
	AEHL01172302	5.00E-26
	CAEC01178005	1.00E-25
	CAEC01649022	3.00E-18
	AGCE01118089	5.00E-49
Aye aye (<i>Dawbentonia madagascariensis</i>)	AGCE01000550	4.00E-48
	AGCE01050342	2.00E-45
	AGCE01075259	2.00E-38
	AGCE01066741	2.00E-30
	AGCE01121515	7.00E-27
Cape golden mole (<i>Chrysochloris asiatica</i>)	AGTM011632853	4.00E-21
	AGTM011653967	4.00E-19
	AGTM011570749	8.00E-13
Star-nosed mole (<i>Candylura cristata</i>)	AMDV01100225	6.00E-112
	AMDV01031468	8.00E-22
	AMDV01073002	7.00E-12
Cape elephant shrew (<i>Elephantulus edwardii</i>)	AMDV01070099	9.00E-12
	AJFV01005318	7.00E-51
	AJFV01039019	2.00E-47
Mouse-eared bat (<i>Myotis davidii</i>)	AMGZ01016176	1.00E-19
	AMGZ01016178	3.00E-19
West Indian manatee (<i>Trichechus manatus latirostris</i>)	ALWT01306233	3.00E-13
	AHIN01118393	2.00E-50
	AHIN01126948	1.00E-45

Mukai, Yahiro, Masayuki Horie, and Keizo Tomonaga. "Systematic estimation of insertion dates of endogenous bornavirus-like elements in vesper bats." *Journal of Veterinary Medical Science* 80.8 (2018): 1356-1363.

Horie, Masayuki, et al. "An RNA-dependent RNA polymerase gene in bat genomes derived from an ancient negative-strand RNA virus." *Scientific reports* 6.1 (2016): 25873.

Horie, Masayuki, et al. "Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes." *Philosophical Transactions of the Royal Society B: Biological Sciences* 368.1626 (2013): 20120499.

Complex research questions

What are all identified EVEs within a species of virus?



Which part integrates most frequently?

Are there patterns?

Which viruses (families) integrate most frequently?

What are all known EVEs in a host?

Are there patterns?



Are they transcribed?

What is the distribution of EVEs?

Where does integration occur most often?

Are there correlations between the host's evolutionary history and the diversity of integrated EVEs?

Accessibility: Solutions

integration in existing databases



EVE specific database



Minimum:

- EVE-Tag
- sequence
- TaxID from host
- TaxID from virus

Wishes:

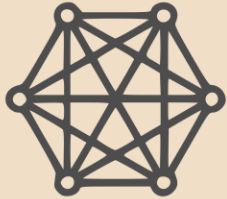
- link to viral sequences
- approach

Availability

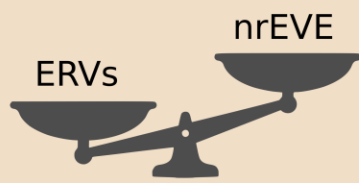
Accessibility

Challenges

complexity



unbalanced



taxonomic assignment



few EVE annotations in host genomes



few sequences



no EVE database



no good EVE tools



Virus tools ignore EVEs



lack of awareness



no EVEs in general database



no working keyword search



no standards



Solutions

Tools for EVE identification on

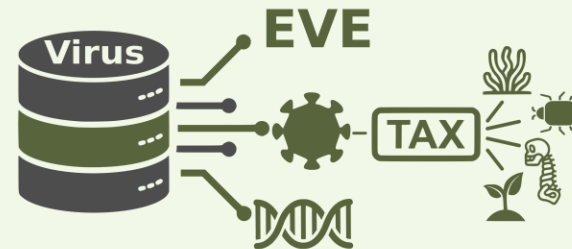


genomic level



transcriptomic level

Integration of EVE data into existing databases



OR EVE specific database



Take home message

Do EVEs play a role in your research, and how can EVEs be integrated?



Thank You for your Attention!

