

Создание высококачественно аннотированной библиотеки соединений, тестированных на противовирусную активность

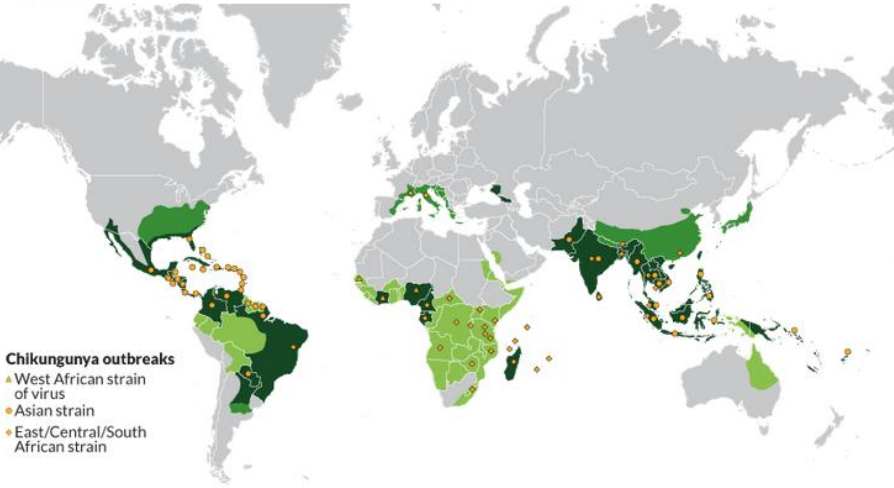
Никитина Анастасия Андреевна,
Орлов А.А., Осолодкин Д.И., Палюлин В.А., Зефирова Н.С.

*Химический факультет МГУ им. М.В. Ломоносова,
ИПВЭ им. М.П. Чумакова*

Chikungunya is on the move

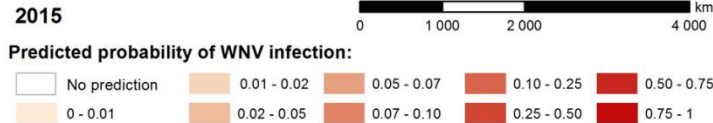
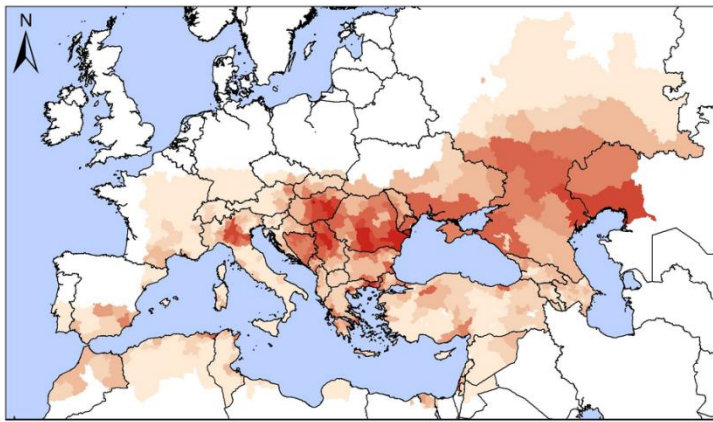
The virus has found a new hemisphere and might get a new latitude

BY NATHAN SEPPA 3:49PM, JUNE 2, 2015



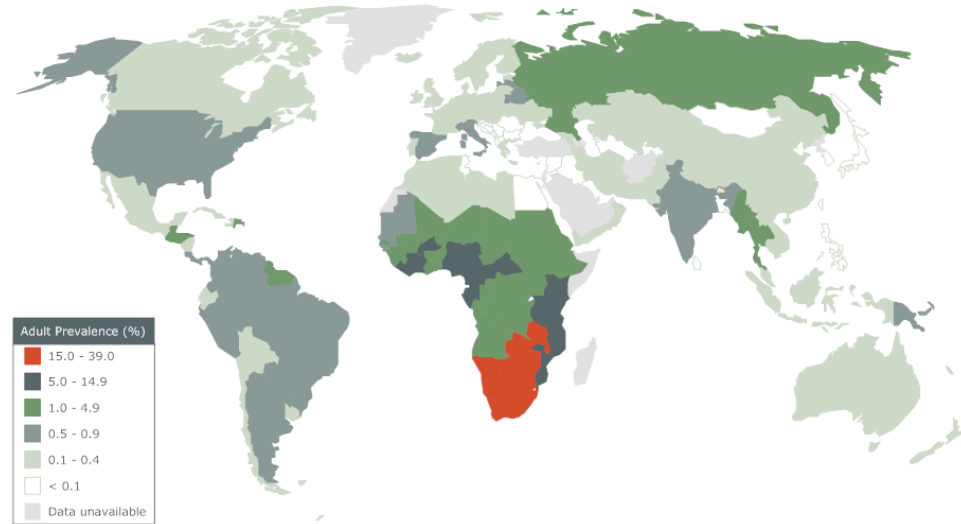
GETTING AROUND The chikungunya virus spreads via mosquitoes in tropical regions. Now it has found a way to hijack a second mosquito, posing a threat to people in Europe, North America and China. (See below for a detailed version of the map.)

GEOATLAS/GRAPHI-OGRE, ADAPTED BY J. HIRSHFELD



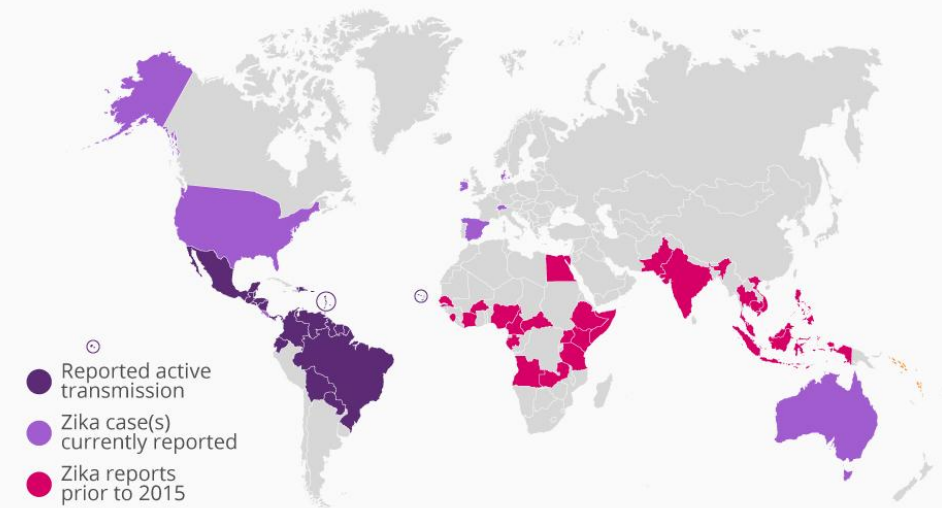
21/08/2015

Adults (15-49 years of age) Living with HIV



The Spread Of The Zika Virus

Countries and territories with active Zika virus transmission* and reported cases





ChEMBL_20
02.02.2015

-
- 1 463 270 соединений (1 456 020 имеют структуры)
 - 13 520 737 результатов измерения активности
 - 1 148 942 биотестов
 - 10 774 мишеней действия
 - 59 610 источников экспериментальных данных
-

Пример использования Web-интерфейса

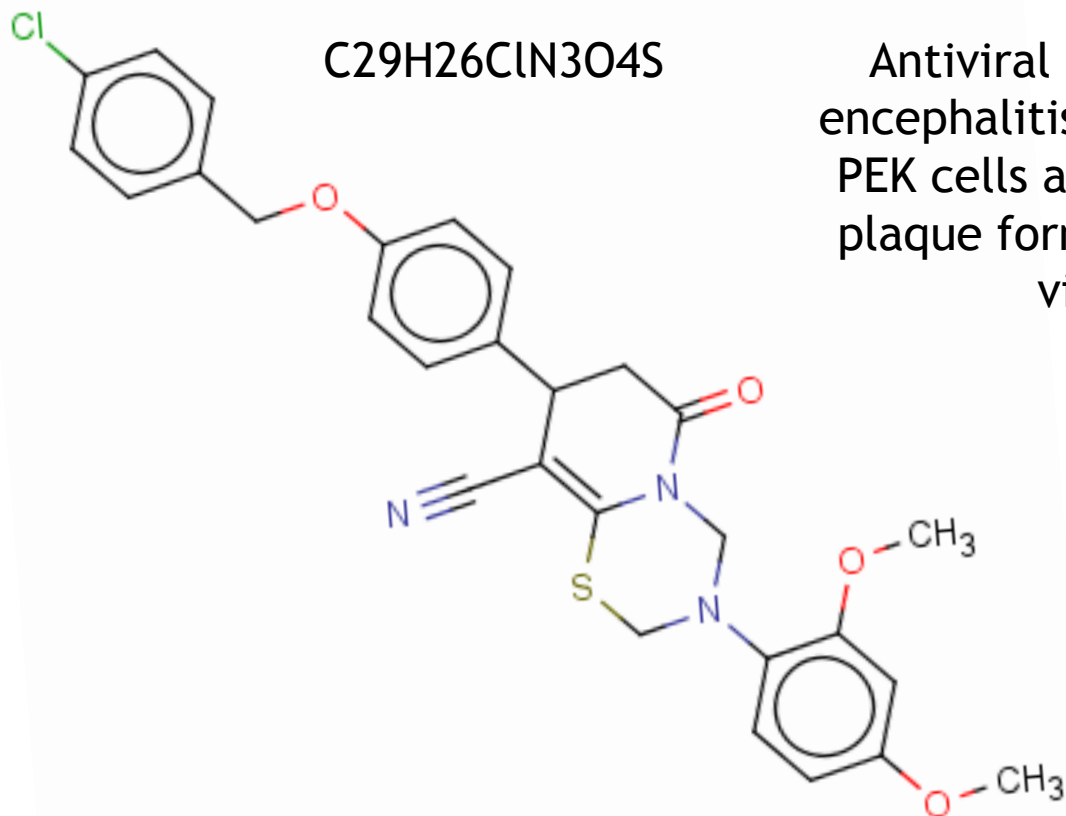
EBI > Databases > Small Molecules > ChEMBL Database > Target Search > Target Classification Hierarchy

 Protein Target Tree Taxonomy Tree

* Click arrows or use keyboard arrows (on selected items) to navigate tree

- Eukaryotes (8140)
- Bacteria (1225)
- Fungi (664)
- Viruses (367)
 - ssRNA (192)
 - dsDNA (98)
 - retro-transcribing (70)
 - dsRNA (7)
- Archaea (6)
- Unclassified (2)

assay_id	организм биотеста	организм мишени	курирование
986472	Tick-borne encephalitis virus	NULL	Autocuration

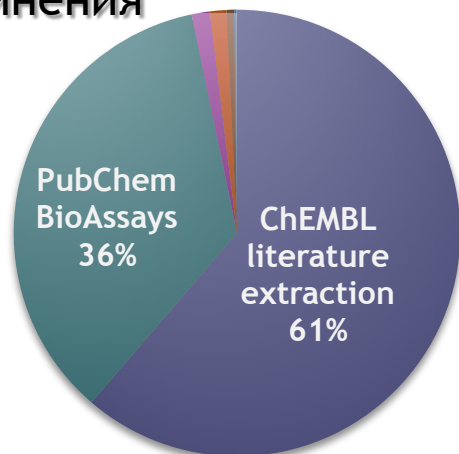


Antiviral activity against Tick-borne encephalitis virus Absettarov infected in PEK cells assessed as inhibition of viral plaque formation after 1 hr by gentian violet staining assay

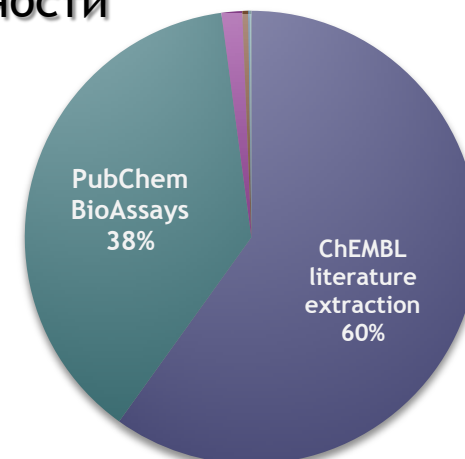
IC₅₀ 70.00 nM

Источники данных и проблемы наполнения ChEMBL

соединения

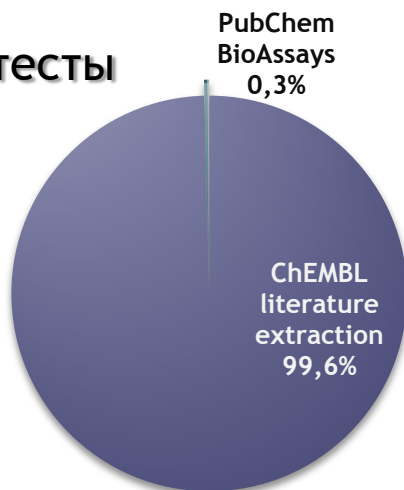


активности



- ChEMBL literature extraction
- PubChem BioAssays
- GSK TCAMS Malaria Data
- PDBe Ligands
- Novartis-GNF Malaria Data
- St Jude Children's Hospital Malaria Data
- Guide to Receptors and Channels
- Sanger Institute Genomics of Drug Sensitivity in Cancer

биотесты



источник ошибок	примеры
эксперимент	Химическая устойчивость и чистота соединения
	Ошибки в каталогах соединений и определении клеточных линий
процесс извлечения данных	Распознавание стереохимии и функциональных групп
	Некорректная и неполная аннотация по мишени действия
исходная публикация	Несоответствующее описание эксперимента, цитирование полученных ранее данных
	Ошибки в единицах измерения и типах активностей

EXPERIMENTAL DATA

DATA_VALIDITY_LOOKUP	
PK	DATA_VALIDITY_COMMENT
	DESCRIPTION

ASSAY_TYPE	
PK	ASSAY_TYPE
	ASSAY_DESC

TARGET INFORMATION

TARGET_TYPE	
PK	ORGANISM

COMPONENT	
PK	PREF_NAME

assay organismорганизм
биотестаdescription

текстовое описание биотеста в свободной форме

organism

целевой организм

pref_nameназвание
мишени

Influenza A virus

In vitro inhibitory activity against influenza A neuraminidase using enzymatic assay

Influenza A virus
(A/Puerto Rico/8/1934(H1N1))

Neuraminidase

Homo sapiens

Antiviral activity against envelope deficient HIV-1 in a single cycle replication assay (experiment 1)

NULL

Unchecked

NULL

Inhibition of HSV-1 DNA polymerase in HSV-1 C42 plaque reduction assay

Herpes simplex virus
(type 1 / strain 17)Human
herpesvirus 1 DNA
polymerase

Hepatitis C virus

Inhibitory activity against the hepatitis C virus NS3 protease was determined

NULL

Unchecked

Tick-borne
encephalitis virus

Antiviral activity against Tick-borne encephalitis virus Absettarov infected in PEK cells assessed as inhibition of viral plaque formation after 1 hr by gentian violet staining assay

NULL

Unchecked

Списки

- Небольшое количество вариантов (< 4000)
- Просматриваются все значения, нужные *отбираются вручную*
- *assays.assay_organism* - 3952
- *target_dictionary.organism* - 2420

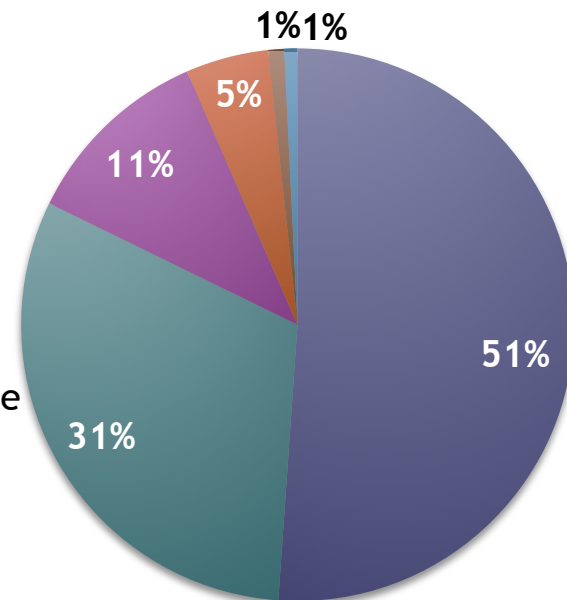
Словарь

- Множество *уникальных* записей (сотни тысяч)
- Составляется на основе наиболее вероятных *ключевых слов* для поиска нужных записей
- *assays.description*
- *target_dictionary.pref_name*

с учетом только информативных значений рассматриваемых полей

Неинформативные значения полей		
<i>Pref_name</i>	<i>TarOrg</i>	<i>AsOrg</i>
Unknown	NULL	Null
Unchecked	Unidentified	Unidentified

- полностью заполненные данные
- организм и название мишени
- только по текстовому описанию
- организм испытаний
- заполнено только название мишени
- организм испытаний и название мишени
- другие варианты заполнения



Списки

Словарь

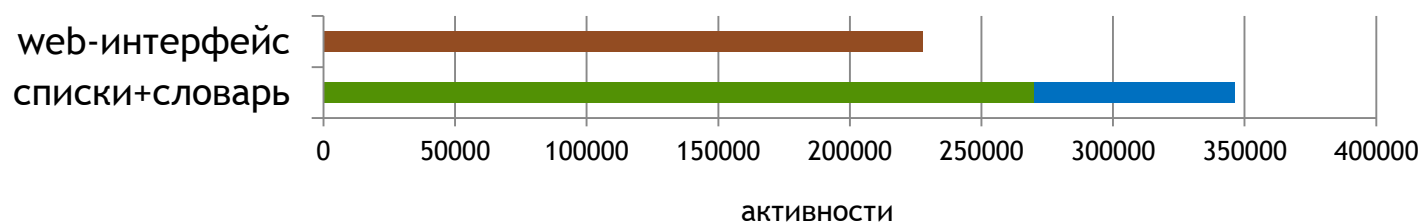
653 организма биотеста из 3952

270 организмов мишени 2420

%virus%, %viral%,
%viridae%, %virinae%, %virales%,



данные различных таксономических
классификаций вирусов, названия
семейств и видов, сокращения



данные	списки	словарь	web-интерфейс
активности	270 018	76 225	227 838
испытания	28 011	5 026	22 629
молекулы	111 025	34 458	101 657

International Committee on Taxonomy of Viruses (ICTV)

Virus Taxonomy: 2014 Release

? How do I use the taxonomy tree?

EC 46, Montreal, Canada, July 2014, Email ratification 2015 (MSL #29)

Order: <i>Caudovirales</i>	(3 Families)	◀ history ▶
Order: <i>Herpesvirales</i>	(3 Families)	◀ history ▶
Order: <i>Ligamenvirales</i>	(2 Families)	◀ history ▶
Order: <i>Mononegavirales</i>	(5 Families)	◀ history ▶
Family: <i>Bornaviridae</i>	(1 Genus)	◀ history ▶
Family: <i>Filoviridae</i>	(3 Genera)	◀ history ▶
Genus: <i>Cuevavirus</i>	(1 Species)	◀ history ▶
Genus: <i>Ebolavirus</i>	(5 Species)	◀ history ▶
Species: <i>Bundibugyo ebolavirus</i>		◀ history ▶
Species: <i>Reston ebolavirus</i>		◀ history ▶
Species: <i>Sudan ebolavirus</i>		◀ history ▶
Species: <i>Tai Forest ebolavirus</i>		◀ history ▶
★ Species: <i>Zaire ebolavirus</i>		◀ history ▶
Genus: <i>Marburgvirus</i>	(1 Species)	◀ history ▶
Family: <i>Nyamiviridae</i>	(1 Genus)	◀ history ▶
Family: <i>Paramyxoviridae</i>	(2 Subfamilies)	◀ history ▶
Family: <i>Rhabdoviridae</i>	(11 Genera)	◀ history ▶
Order: <i>Nidovirales</i>	(4 Families)	◀ history ▶
Order: <i>Picornavirales</i>	(5 Families)	◀ history ▶
Order: <i>Tymovirales</i>	(4 Families)	◀ history ▶
Virus families not assigned to an order	(78 Families)	◀ history ▶

International Committee on Taxonomy of Viruses (ICTV)

? How do I use the taxonomy tree?

Virus Taxonomy: 2014 Release

EC 46, Montreal, Canada, July 2014, Email ratification 2015 (MSL #29)

Order: *Caudovirales*

Order: *Herpesvirales*

Order: *Ligamenvirales*

Order: *Mononegavirales*

Family: *Bornaviridae*

Family: *Filoviridae*

Genus: *Cuevavirus*

Genus: *Ebolavirus*

Species: *Bundibugyo ebolavirus*

Species: *Reston ebolavirus*

Species: *Sudan ebolavirus*

Species: *Tai Forest ebolavirus*

★ Species: *Zaire ebolavirus*

Genus: *Marburgvirus*

Family: *Nyamiviridae*

Family: *Paramyxoviridae*

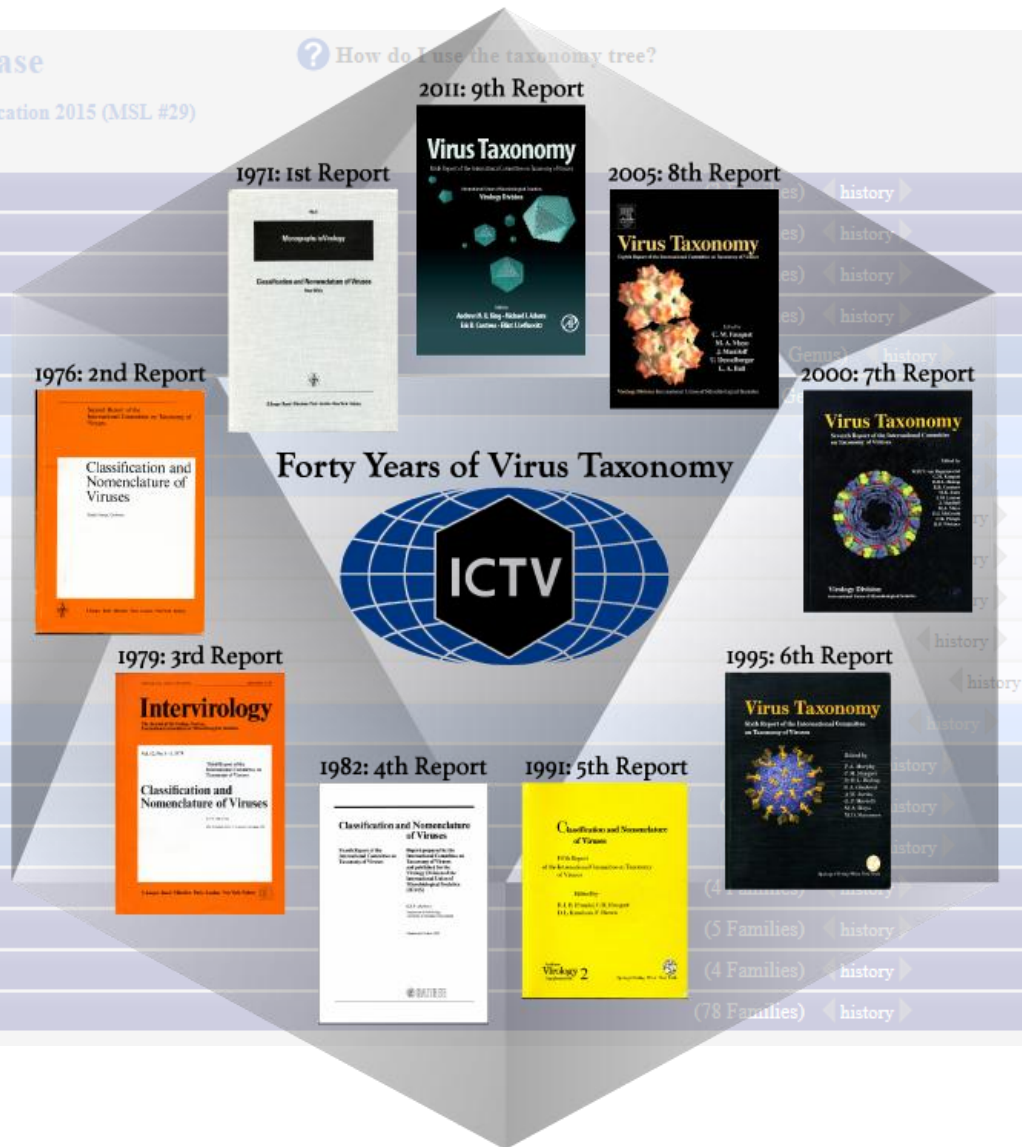
Family: *Rhabdoviridae*

Order: *Nidovirales*

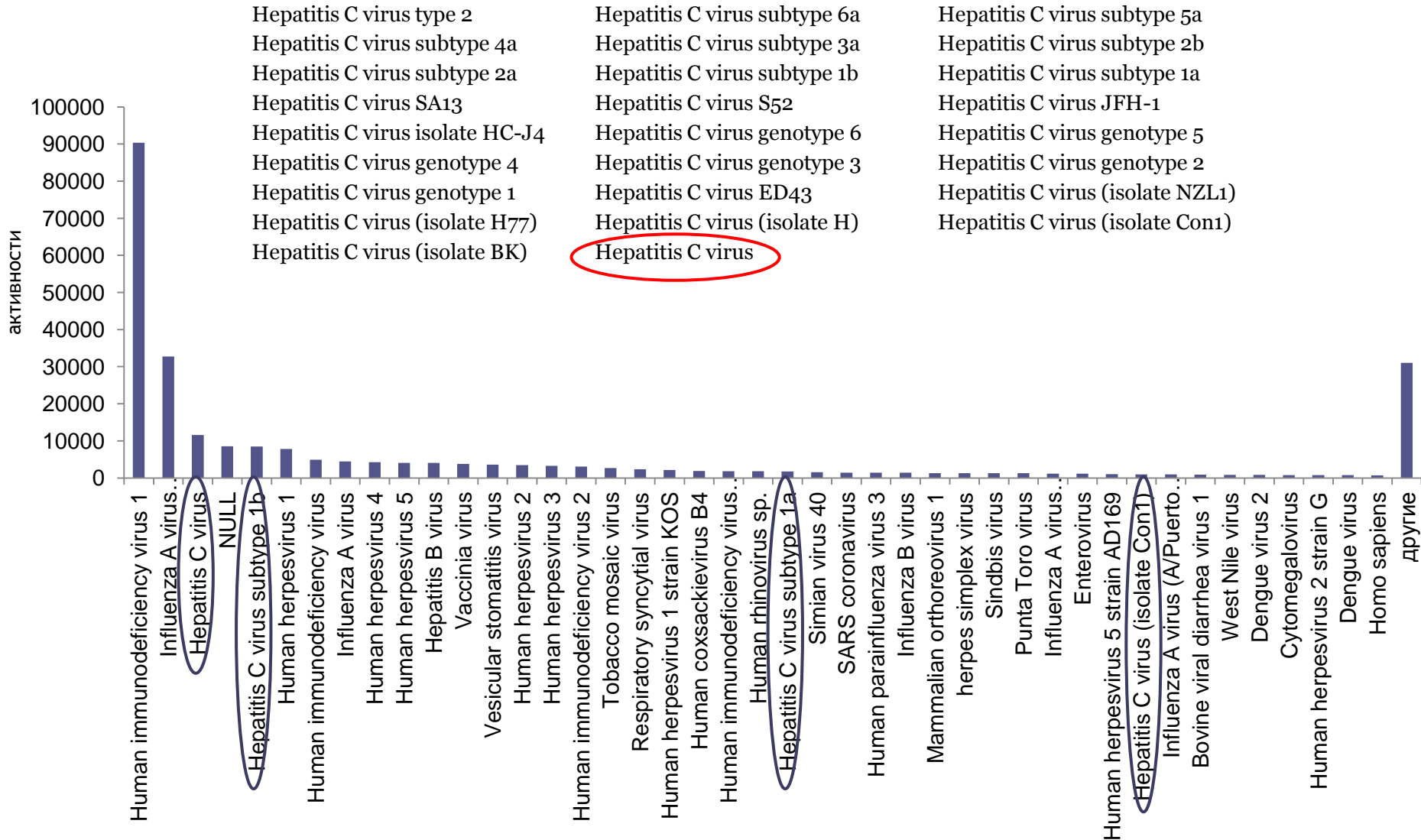
Order: *Picornavirales*

Order: *Tymovirales*

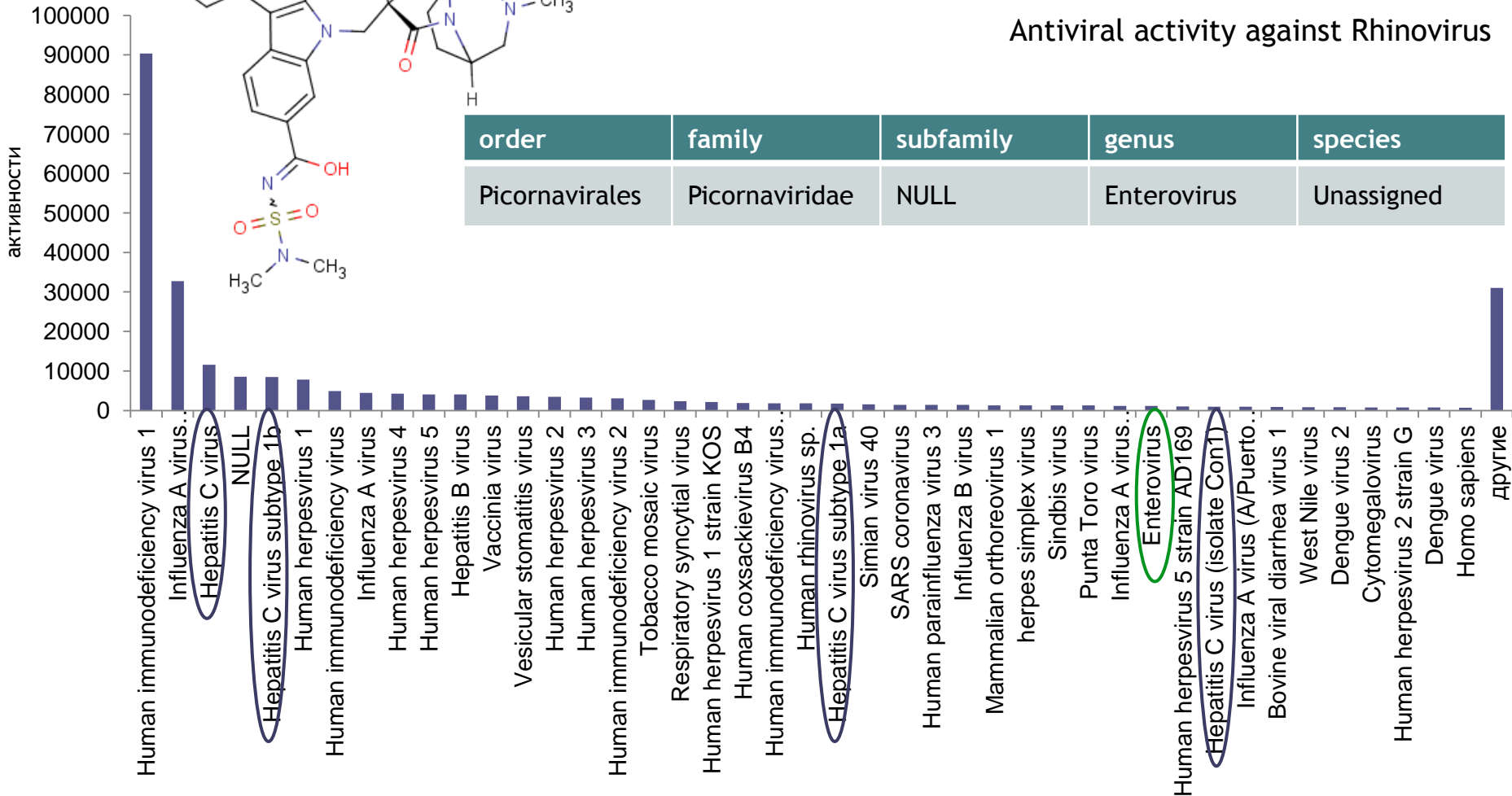
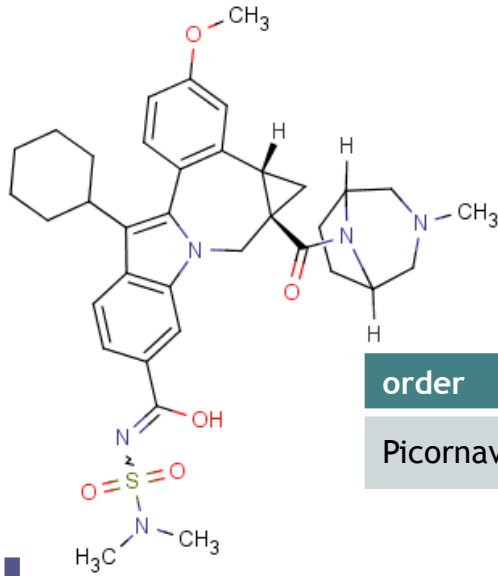
Virus families not assigned to an order



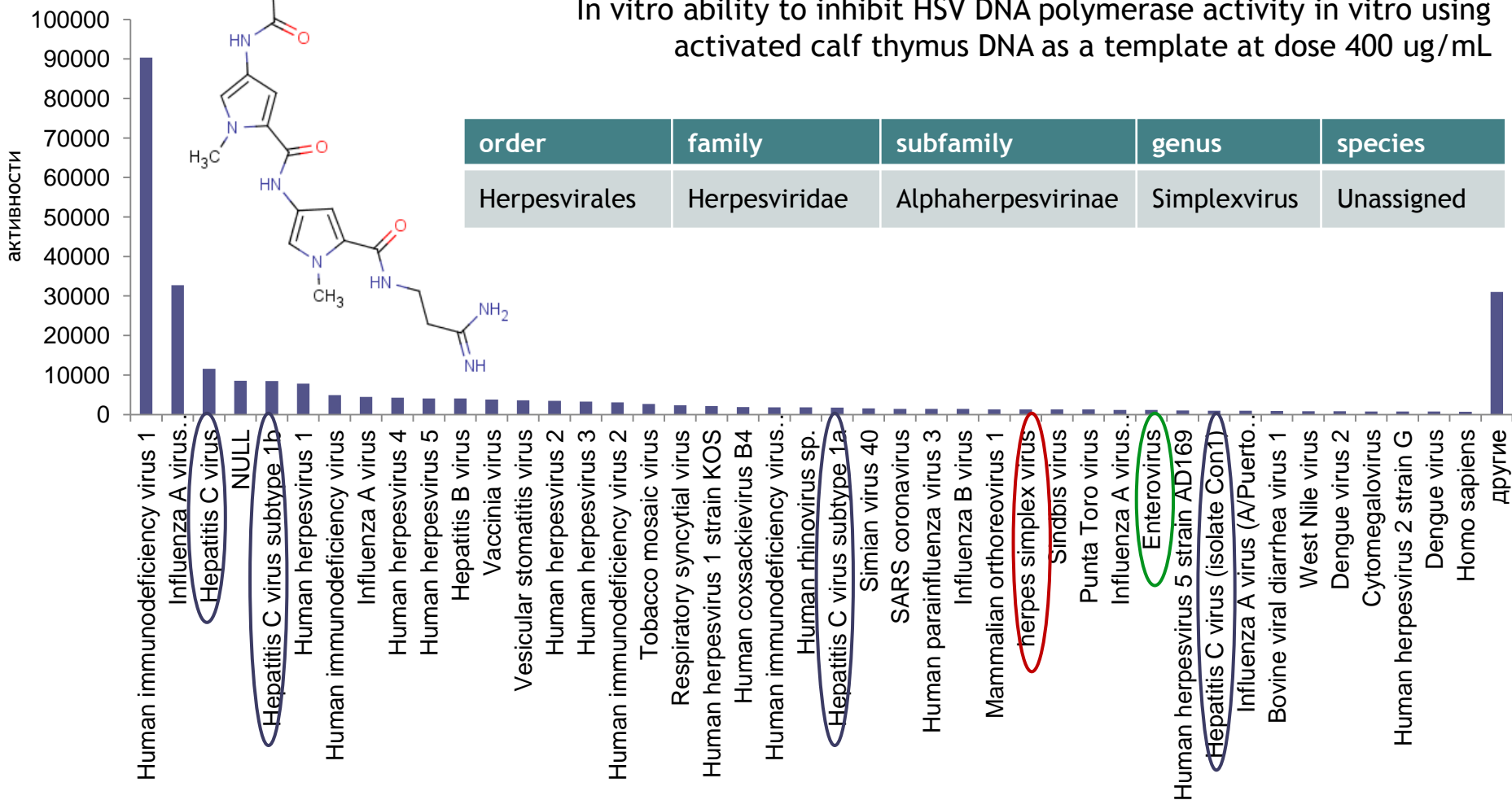
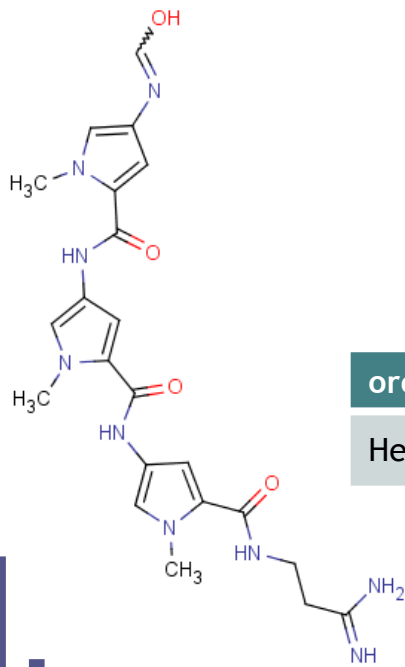
Разнообразие аннотаций. Уровни детализации



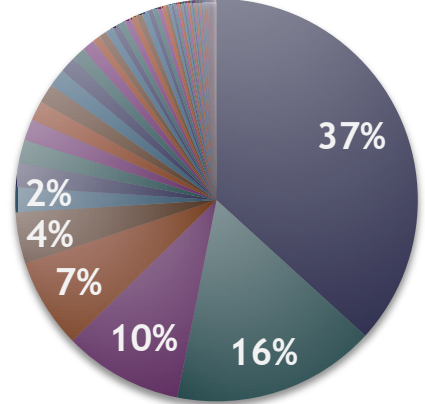
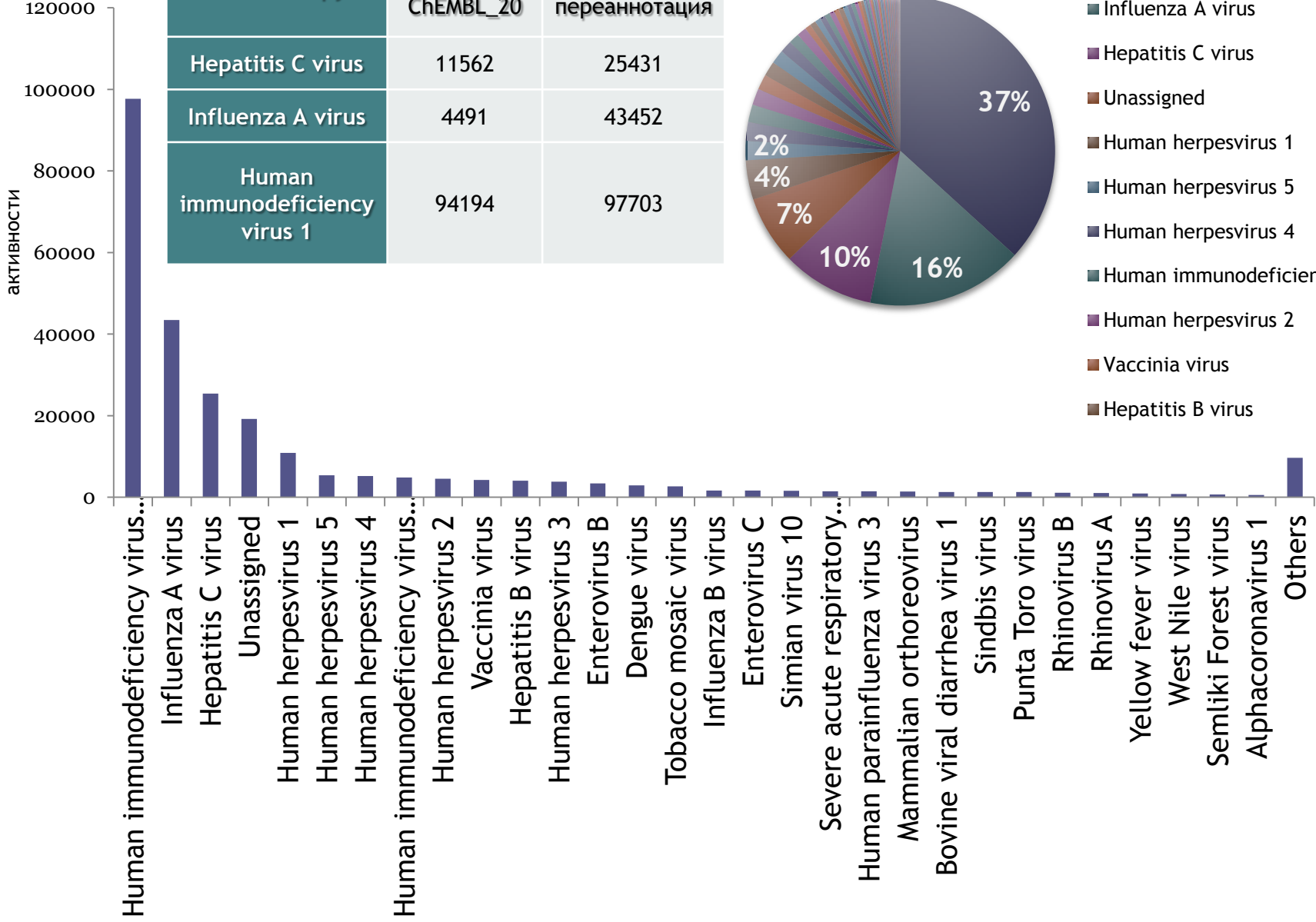
Разнообразие аннотаций. Уровни таксономии



Разнообразие аннотаций. Исторические названия



Название вируса	КОЛИЧЕСТВО АКТИВНОСТЕЙ	
	ChEMBL_20	переаннотация
Hepatitis C virus	11562	25431
Influenza A virus	4491	43452
Human immunodeficiency virus 1	94194	97703



- Human immunodeficiency virus 1
- Influenza A virus
- Hepatitis C virus
- Unassigned
- Human herpesvirus 1
- Human herpesvirus 5
- Human herpesvirus 4
- Human immunodeficiency virus 2
- Human herpesvirus 2
- Vaccinia virus
- Hepatitis B virus

Выводы

- Проведен анализ данных о противовирусных соединениях;
 - Разработана методика поиска и переаннотации данных о противовирусной активности в соответствии с современной таксономией вирусов ICTV;
 - Переаннотированы данные с корректно заполненными полями *assays.organism* и *target_dictionary.organism* (270 018).
-

Направления будущей работы

- Переаннотация оставшихся данных с использованием словарей подстрок;
 - Анализ химического пространства противовирусных соединений на основе полученных данных.
-

Работа поддержана грантом РФФ №16-15-10307