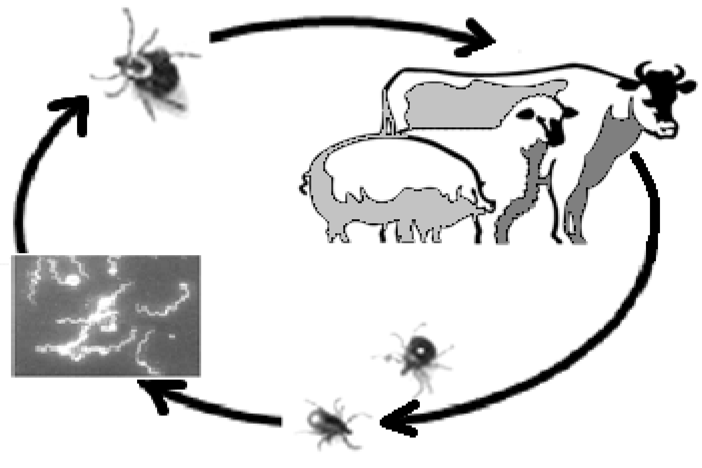
What, where and weather? Integrating open-source taxonomic, spatial and climatologic information into a comprehensive database of livestock infections

K.M. McIntyre[1]*, C. Setzkorn[1], M. Wardeh[1], P. J. Hepworth[1], A.D. Radford[2] & M. Baylis[1]

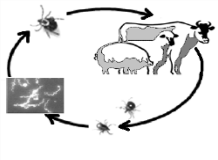
[1]Department of Epidemiology & Population Health, Institute of Infection & Global Health (IGH), University of Liverpool (UoL), UK
[2]Department of Infection Biology, IGH, UoL, UK

*k.m.mcintyre@liv.ac.uk



ENHanCEd Infectious Diseases (EID2) database

UNIVERSITY OF LIVERPOOL



THE ROYAL SOCIETY

doi 10.1098/rstb.2001.0889

Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence

S. Cleaveland*, **M. K. Laurenson and L. H. Taylor**

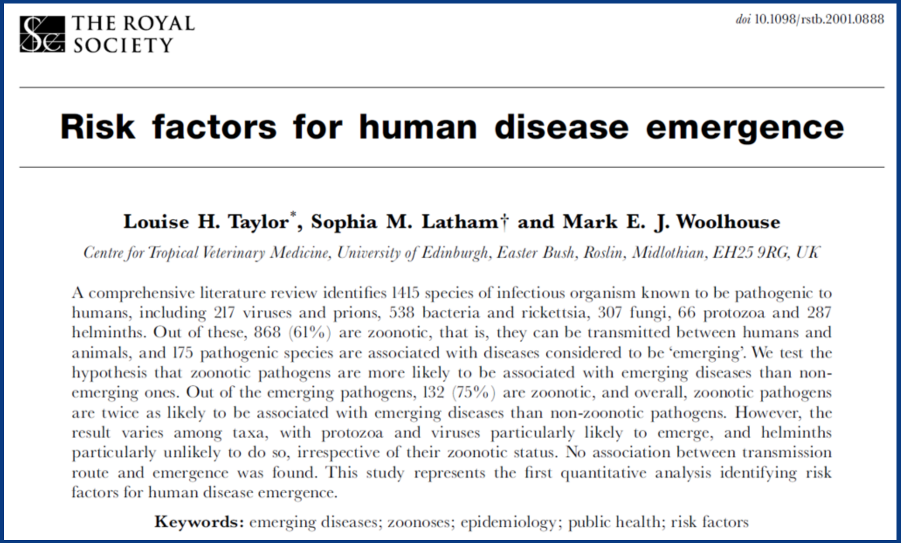
Centre for Tropical Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian EH25 9RG, UK

Pathogens that can be transmitted between different host species are of fundamental interest and importance from public health, conservation and economic perspectives, yet systematic quantification of these pathogens is lacking. Here, pathogen characteristics, host range and risk factors determining disease emergence were analysed by constructing a database of disease-causing pathogens of humans and domestic mammals. The database consisted of 1415 pathogens causing disease in humans, 616 in livestock and 374 in domestic carnivores. Multihost pathogens were very prevalent among human pathogens (61.6%) and even more so among domestic mammal pathogens (livestock 77.3%, carnivores 90.0%). Pathogens able to infect human, domestic and wildlife hosts contained a similar proportion of disease-causing pathogens for all three host groups. One hundred and ninety-six pathogens were associated with emerging diseases, 175 in humans, 29 in livestock and 12 in domestic carnivores. Across all these groups, helminths and fungi were relatively unlikely to emerge whereas viruses, particularly RNA viruses, were highly likely to emerge. The ability of a pathogen to infect multiple hosts, particularly hosts in other taxonomic orders or wildlife, were also risk factors for emergence in human and livestock pathogens. There is clearly a need to understand the dynamics of infectious diseases in complex multihost communities in order to mitigate disease threats to public health, livestock economies and wildlife.

Keywords: pathogen; epidemiology; emerging diseases; zoonoses; wildlife; multihost pathogen

Time cited = 248

Human, livestock & domestic carnivore pathogens



THE ROYAL SOCIETY

doi 10.1098/rstb.2001.0888

Risk factors for human disease emergence

Louise H. Taylor*, **Sophia M. Latham† and Mark E. J. Woolhouse**

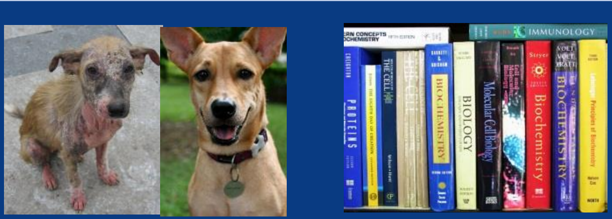
Centre for Tropical Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian, EH25 9RG, UK

A comprehensive literature review identifies 1415 species of infectious organism known to be pathogenic to humans, including 217 viruses and prions, 538 bacteria and rickettsia, 307 fungi, 66 protozoa and 287 helminths. Out of these, 868 (61%) are zoonotic, that is, they can be transmitted between humans and animals, and 175 pathogenic species are associated with diseases considered to be 'emerging'. We test the hypothesis that zoonotic pathogens are more likely to be associated with emerging diseases than non-emerging ones. Out of the emerging pathogens, 132 (75%) are zoonotic, and overall, zoonotic pathogens are twice as likely to be associated with emerging diseases than non-zoonotic pathogens. However, the result varies among taxa, with protozoa and viruses particularly likely to emerge, and helminths particularly unlikely to do so, irrespective of their zoonotic status. No association between transmission route and emergence was found. This study represents the first quantitative analysis identifying risk factors for human disease emergence.

Keywords: emerging diseases; zoonoses; epidemiology; public health; risk factors

Time cited = 446

Limitations



~~Ehrlichia equi~~ -> Anaplasma phagocytophilum

- Clinical vs non-clinical infection & source biases
- Updating phylogenetic information

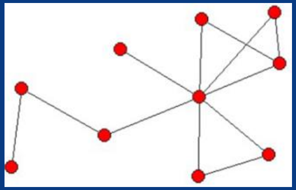
- Updating for emerging pathogens

- Surveillance/health-care/host biases

- Domestication of animals

~~0~~/1 ~~2,3,5,8,15~~

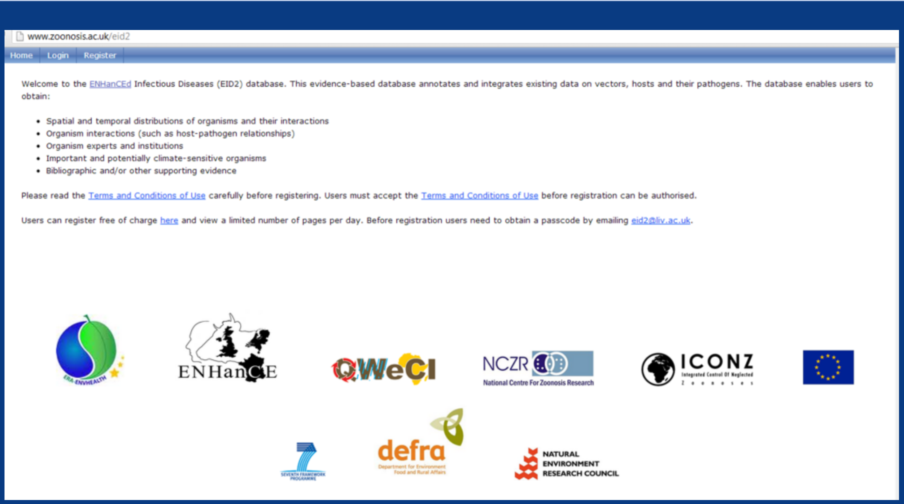
- Give no quantitative estimate of pathogens; presence-only lists



- No taxonomic information to pathogen or host nodes

Introduction, EID2 database & Methods, Results, Discussion, Conclusions, Future

ENHanCEd
Infectious Diseases database



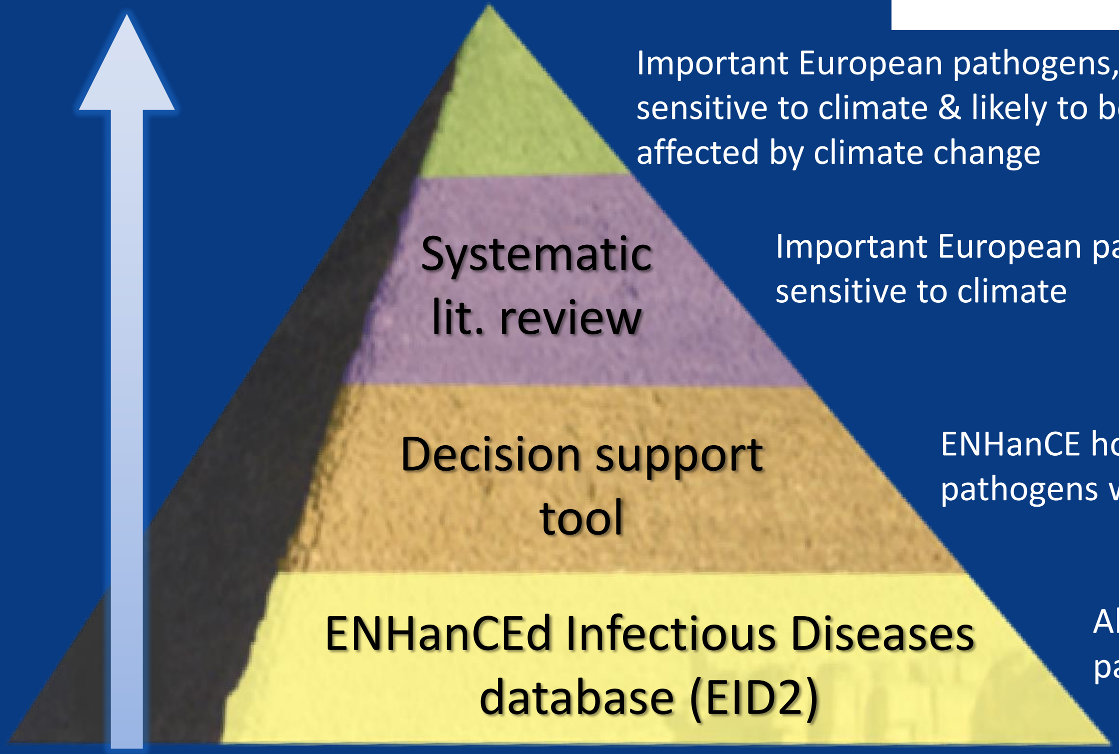
www.zoonosis.ac.uk/eid2

Welcome to the ENHanCEd Infectious Diseases (EID2) database. This evidence-based database annotates and integrates existing data on vectors, hosts and their pathogens. The database enables users to obtain:

- Spatial and temporal distributions of organisms and their interactions
- Organism interactions (such as host-pathogen relationships)
- Organism experts and institutions
- Important and potentially climate-sensitive organisms
- Bibliographic and/or other supporting evidence

Please read the Terms and Conditions of Use carefully before registering. Users must accept the Terms and Conditions of Use before registration can be authorised.

Users can register free of charge here and view a limited number of pages per day. Before registration users need to obtain a passcode by emailing eid2@liv.ac.uk.

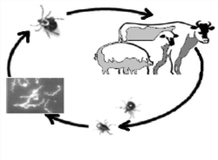


Important European pathogens, sensitive to climate & likely to be affected by climate change

Important European pathogens sensitive to climate

ENHanCE host European pathogens with high impact

All pathogens - > All European pathogens in 'ENHanCE hosts'



BBSRC TRDF - What, where and weather?

> "..finalise the development of EID2 into a tool and resource for researchers of pathogens of livestock and domestic pets."

Aims

(1) Describe the EID2 database

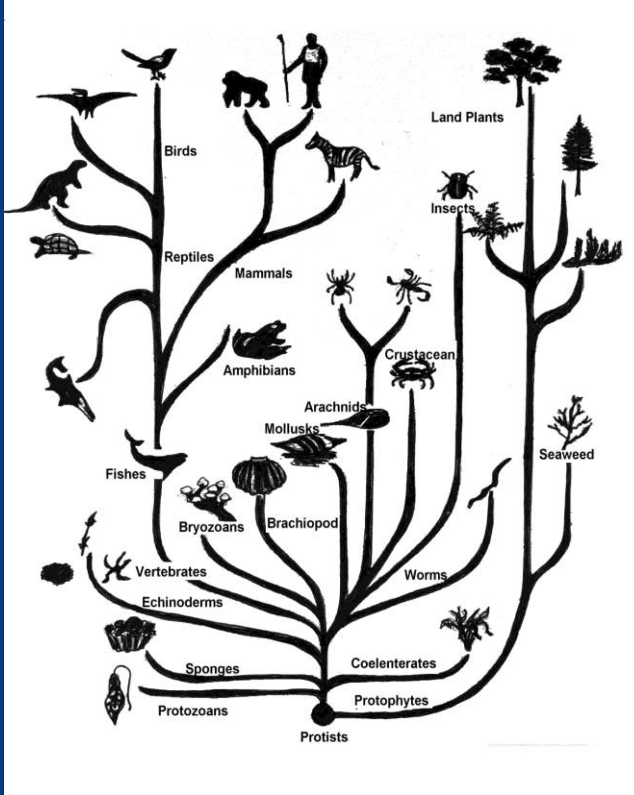
(2) Investigate the validity of using semi-automated literature gathering techniques

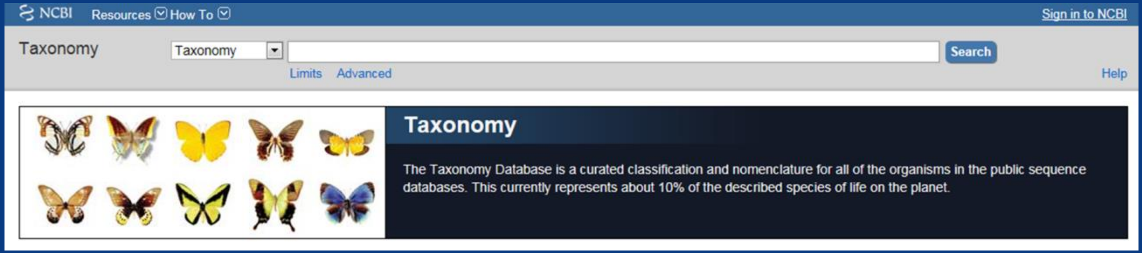
(3) Identify the pathogens and hosts at the root of human and domestic animal disease networks

(4) Examine the characteristics of these hosts and pathogens to distinguish potential drivers of disease transmission

(5) Compare EID2 results with those of previous work

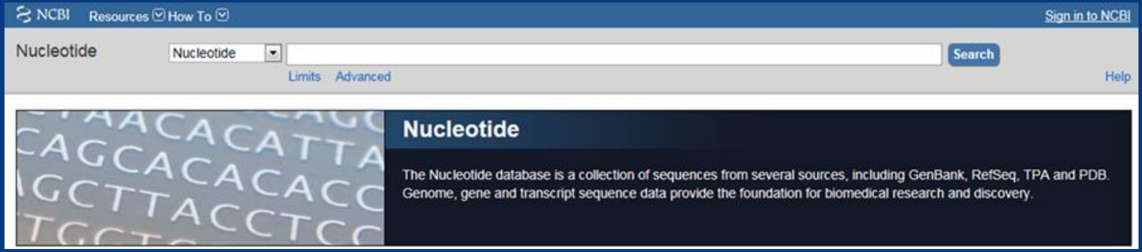
EID2 sources





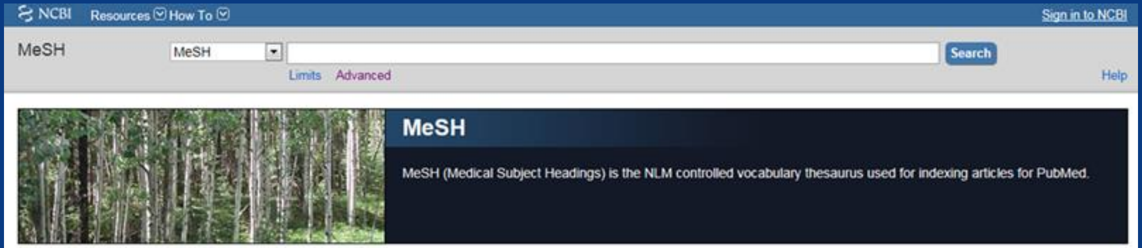
Taxonomy

The Taxonomy Database is a curated classification and nomenclature for all of the organisms in the public sequence databases. This currently represents about 10% of the described species of life on the planet.



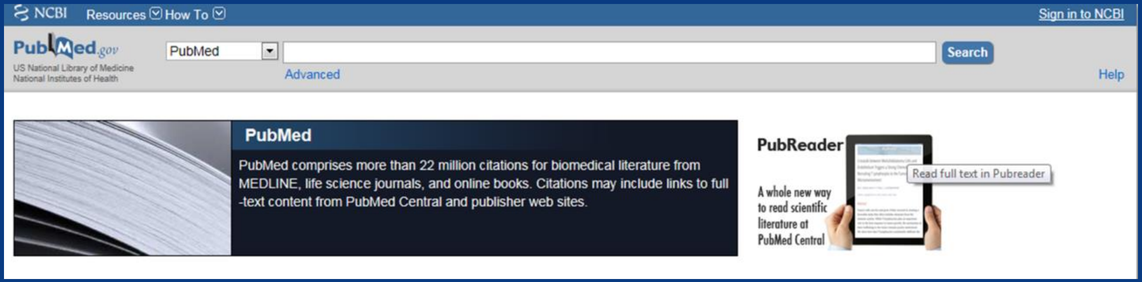
Nucleotide

The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.



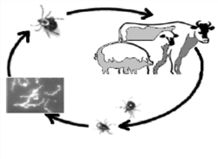
MeSH

MeSH (Medical Subject Headings) is the NLM controlled vocabulary thesaurus used for indexing articles for PubMed.

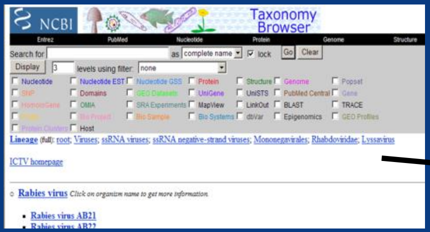


PubMed

PubMed comprises more than 22 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full -text content from PubMed Central and publisher web sites.



Organism page



Search for [ ] as complete name ☑ lock

Display 3 levels using filter: none

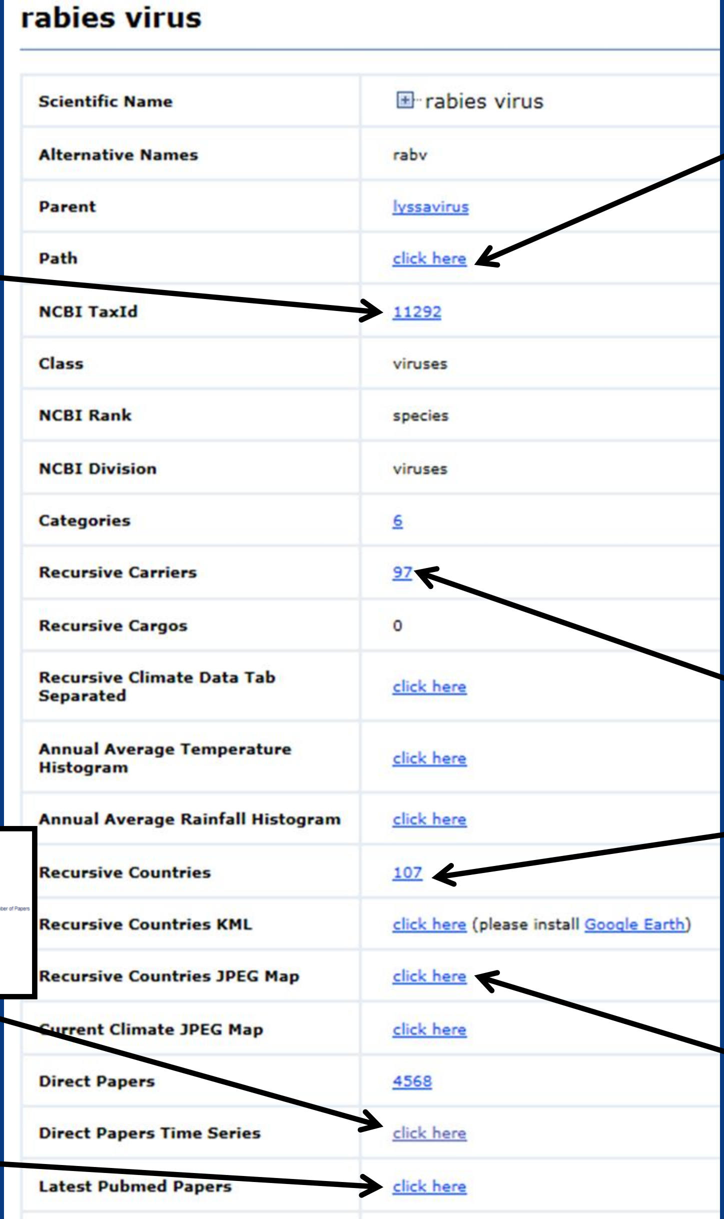
☐ Nucleotide ☐ Nucleotide EST ☐ Nucleotide GSS ☐ Protein ☐ Structure ☐ Genome ☐ Popset
☐ SNP ☐ Domains ☐ GEO Datasets ☐ UniGene ☐ UniSTS ☐ PubMed Central ☐ Gene
☐ HomoloGene ☐ OMIA ☐ SRA Experiments ☐ MapView ☐ LinkOut ☐ BLAST ☐ TRACE
☐ Probe ☐ Bio Project ☐ Bio Sample ☐ Bio Systems ☐ dbVar ☐ Epigenomics ☐ GEO Profiles
☐ Protein Clusters ☐ Host

Lineage (full): root; Viruses; ssRNA viruses; ssRNA negative-strand viruses; Mononegavirales; Rhabdoviridae; Lyssavirus

ICTV homepage

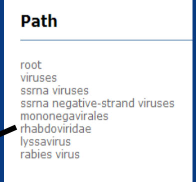
○ **Rabies virus** *Click on organism name to get more information.*

- Rabies virus AB21
- Rabies virus AB22



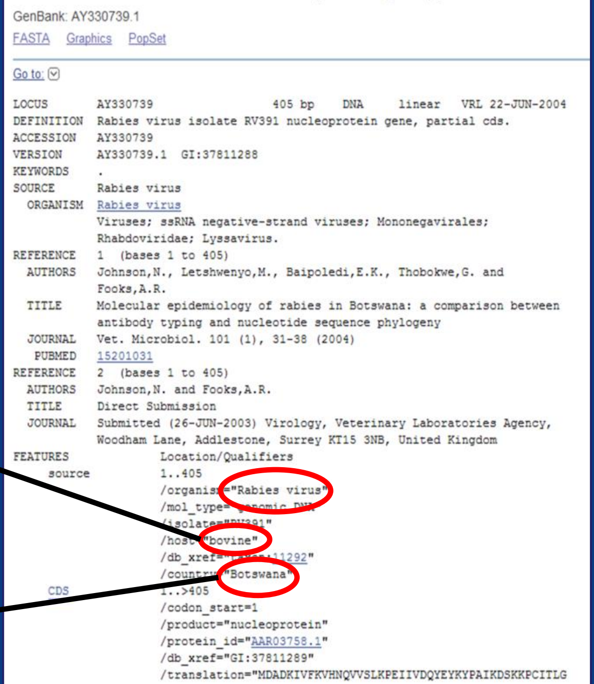
rabies virus

Scientific Name	⊞ rabies virus
Alternative Names	rabv
Parent	lyssavirus
Path	click here
NCBI TaxId	11292
Class	viruses
NCBI Rank	species
NCBI Division	viruses
Categories	6
Recursive Carriers	97
Recursive Cargos	0
Recursive Climate Data Tab Separated	click here
Annual Average Temperature Histogram	click here
Annual Average Rainfall Histogram	click here
Recursive Countries	107
Recursive Countries KML	click here (please install Google Earth)
Recursive Countries JPEG Map	click here
Current Climate JPEG Map	click here
Direct Papers	4568
Direct Papers Time Series	click here
Latest Pubmed Papers	click here



Path

root
viruses
ssrna viruses
ssrna negative-strand viruses
mononegavirales
rhabdoviridae
lyssavirus
rabies virus



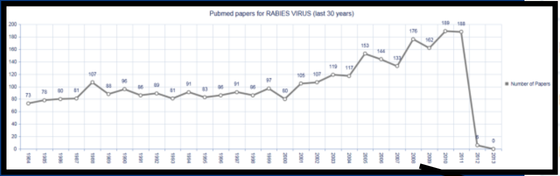
Rabies virus isolate RV391 nucleoprotein gene, partial cds

GenBank: AY330739.1

FASTA Graphics PopSet

Go to:

```
LOCUS       AY330739                 405 bp    DNA     linear   VRL 22-JUN-2004
DEFINITION  Rabies virus isolate RV391 nucleoprotein gene, partial cds.
ACCESSION   AY330739
VERSION     AY330739.1  GI:37811288
KEYWORDS    .
SOURCE      Rabies virus
  ORGANISM  Rabies virus
            Viruses; ssRNA negative-strand viruses; Mononegavirales;
            Rhabdoviridae; Lyssavirus.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Johnson,N., Letshwenyo,M., Baipoledi,E.K., Thobokwe,G. and
            Fooks,A.R.
  TITLE     Molecular epidemiology of rabies in Botswana: a comparison between
            antibody typing and nucleotide sequence phylogeny
  JOURNAL   Vet. Microbiol. 101 (1), 31-38 (2004)
   PUBMED   15201031
REFERENCE   2  (bases 1 to 405)
  AUTHORS   Johnson,N. and Fooks,A.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2003) Virology, Veterinary Laboratories Agency,
            Woodham Lane, Addlestone, Surrey KT15 3NB, United Kingdom
FEATURES             Location/Qualifiers
     source          1..405
                     /organism="Rabies virus"
                     /mol_type="genomic DNA"
                     /isolate="RV391"
                     /host="bovine"
                     /db_xref="taxon:11292"
                     /country="Botswana"
     CDS             1..>405
                     /codon_start=1
                     /product="nucleoprotein"
                     /protein_id="AAR03758.1"
                     /db_xref="GI:37811289"
                     /translation="MDADKIVFKVHNQVVSLKPEIIVDQYEYKYPAIKDSKKPCITLG
```



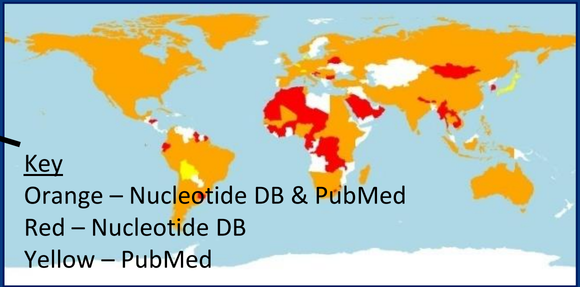
PubMed: "rabies virus"[Title] OR "rabies"[Title] AND "chile"[MeSH Terms]

Display Settings: Summary, 20 per page, Sorted by Recently Added

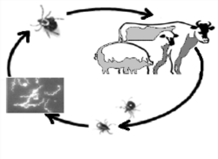
Results: 9

☐ [Epidemiological description of **rabies** reservoir in bats in the Metropolitan Region. **Chile**. 2000-2009].
1. Favi C M, Bassaletti C A, López D J, Rodríguez A L, Yung P V.
Rev Chilena Infectol. 2011 Jun;28(3):223-8. Epub 2011 Jul 14. Spanish.
PMID: 21879147 [PubMed - indexed for MEDLINE] Free Article
Related citations

☐ [**Rabies** in **Chile**. 1989-2005].
2. Favi C M, Rodríguez A L, Espinosa M C, Yung P V.
Rev Chilena Infectol. 2008 Apr;25(2):S8-13. Epub 2008 Apr 16. Spanish.
PMID: 18425219 [PubMed - indexed for MEDLINE] Free Article



Key
Orange – Nucleotide DB & PubMed
Red – Nucleotide DB
Yellow – PubMed



Scientific name	Common name	Scientific name	Common name
Agapornis personata	Masked lovebird	*Lama glama*	Lama
Agapornis roseicollis	Rosy-faced lovebird	*Lama pacos*	Alpaca
Anas platyrhynchos	Domestic duck	*Meleagris gallopavo*	Turkey
Anser anser	Domestic goose	*Melopsittacus undulatus*	Budgerigar
Bison bison	American bison	*Meriones unguiculatus*	Mongolian gerbil
Bison bonasus	European bison	*Mesocricetus auratus*	Syrian golden hamster
Bos indicus	Zebu	*Mus musculus*	House mouse
Bos taurus	Cow	*Mustela putorius furo*	Domestic ferret
Camelus dromedarius	Dromedary	*Numida meleagris*	Helmeted guineafowl
Canis lupus familiaris	Domestic dog	*Nymphicus hollandicus*	Cockatiel
Capra hircus	Domestic goat	*Oryctolagus cuniculus*	Domestic rabbit
Capreolus capreolus	Roe deer	*Ovis aries*	Sheep
Cavia porcellus	Domestic guinea pig	*Ovis aries musimon*	Mouflon
Cervus elaphus	Red deer	*Pavo cristatus*	Blue peafowl
Chinchilla lanigera	Chinchilla	*Phasianus colchicus*	Ring-necked pheasant
Columba livia	Domestic pigeon	*Rangifer tarandus*	Reindeer
Cricetus cricetus	Common hamster	*Rattus norvegicus*	Brown rat
Dama dama	Fallow deer	*Rattus rattus*	Black rat
Equus asinus	Domestic donkey	*Rhombomys opimus*	Great Gerbil
Equus caballus	Domestic horse	*Serinus canaria*	Canary
Felis catus	Domestic cat	*Struthio camelus*	Ostrich
Gallus gallus	Chicken	*Sus scrofa*	Wild boar
Homo sapiens	Humans	*Sus scrofa domesticus*	Domestic pig
Lagopus lagopus scotica	Red grouse		

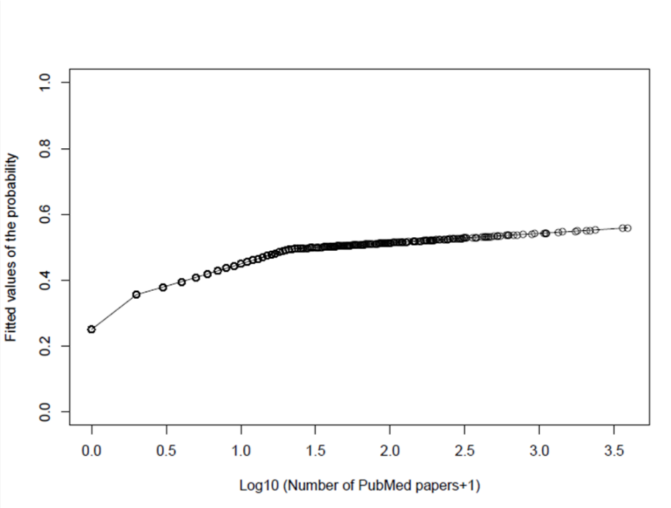
Using semi-automated literature gathering techniques to ascertain pathogen-location interactions

Positive predictive value (PPV) approach

- Stratify papers by pathogen and continent
- Sub-sample papers to check for ‘true’ evidence of pathogens occurring in countries
- PPV = 0.95 (SE=0.05)
- Threshold of 5 papers (0.95^5) = 99.9% certainty

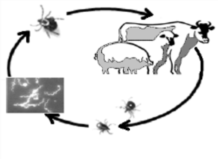
Binomial regression modelling approach

0/1+ NCBI Nucleotide sequence ~ no. of PubMed papers in which pathogen name & country MeSH term co-occurred

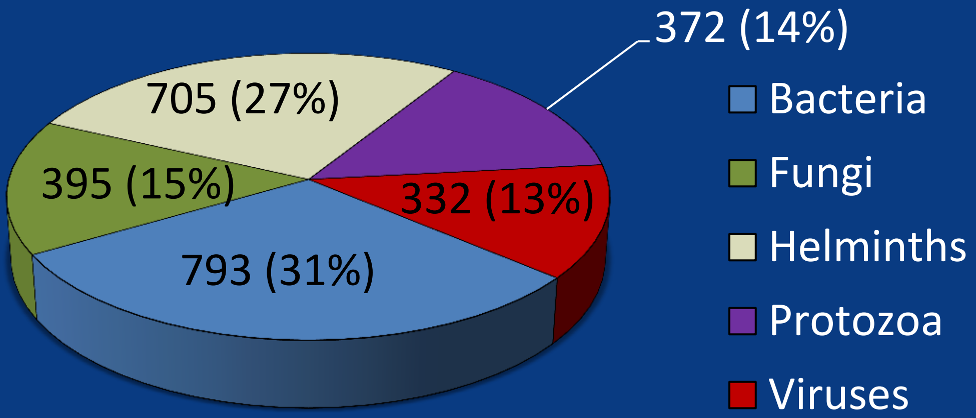


With No. papers coded as a factor:

No. of papers	Odds Ratio
0 -> 1	Baseline
2 -> 12	1.86
13+	2.89



Pathogen range within human & domestic animal hosts



No. pathogen species in hosts, n=2597

Host group	% pathogens
Humans only	49.8
Domestic animals only	32.5
Both	17.8

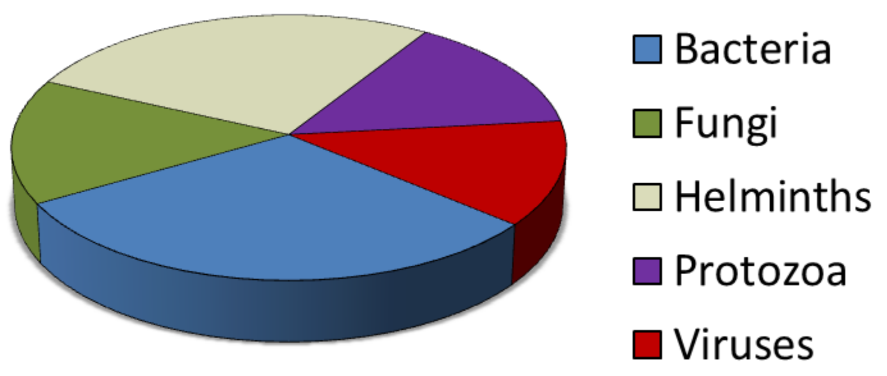
Human pathogens	% pathogens
Humans only	74.9
Humans & d. animals	25.1

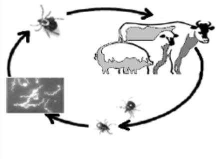
No. of hosts	% pathogens
1	70.9
Up to 2	86.0
Up to 6	98.6

Domestic animal pathogens	% pathogens
Domestic animals only	65.7
Humans & d. animals	34.3

Greater proportion of pathogens affect human-only than animal-only (χ^2 = *P*<0.001)

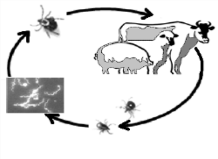
Pathogen name	Pathogen type	Number of hosts	Pathogen name	Pathogen type	Number of hosts
Escherichia coli $_{\text{E,Z}}$	Bacteria	21	Gongylonema pulchrum $_{\text{NE,Z}}$	Helminth	9
Giardia intestinalis $_{\text{E,Z}}$	Protozoa	20	Leptospira interrogans $_{\text{E,Z}}$	Bacteria	9
Toxoplasma gondii $_{\text{E,Z}}$	Protozoa	18	Ovine Herpesvirus 2 $_{\text{NA}}$	Virus	9
Anaplasma phagocytophilum $_{\text{E,Z}}$	Bacteria	15	Rotavirus A $_{\text{E,Z}}$	Virus	9
Cryptosporidium parvum $_{\text{E,Z}}$	Protozoa	14	Clostridium perfringens $_{\text{NE, Z}}$	Bacteria	8
Rabies virus $_{\text{E,Z}}$	Virus	13	Cowpox virus $_{\text{NE, Z}}$	Virus	8
Staphylococcus aureus $_{\text{E,Z}}$	Bacteria	13	Enterococcus faecalis $_{\text{E,Z}}$	Bacteria	8
Neospora caninum $_{\text{NA}}$	Protozoa	12	Enterococcus faecium $_{\text{E,Z}}$	Bacteria	8
Echinococcus granulosus $_{\text{E,Z}}$	Helminth	11	Enterocytozoon bieneusi $_{\text{E, NZ}}$	Fungi	8
Borna Disease virus $_{\text{NE,Z}}$	Virus	10	Hepatitis E virus $_{\text{E,Z}}$	Virus	8
Newcastle Disease virus $_{\text{NE,Z}}$	Virus	10	Malassezia sympodialis $_{\text{NE, Z}}$	Fungi	8
Pasteurella multocida $_{\text{NE,Z}}$	Bacteria	10	Brachyspira pilosicoli $_{\text{NA}}$	Bacteria	7
Trypanosoma cruzi $_{\text{E,Z}}$	Protozoa	10	Influenza A virus $_{\text{E, Z}}$	Virus	7
Babesia divergens $_{\text{NE,Z}}$	Protozoa	9	Mecistocirrus digitatus $_{\text{NE, Z}}$	Helminth	7
Chlamydophila psittaci $_{\text{NE,Z}}$	Bacteria	9	Pneumocystis carinii $_{\text{E, Z}}$	Fungi	7
Cryptosporidium muris $_{\text{NA}}$	Protozoa	9	Saccharomyces cerevisiae $_{\text{NE, NZ}}$	Fungi	7
Echinococcus canadensis $_{\text{NA}}$	Helminth	9	Trichostrongylus colubriformis $_{\text{NE, Z}}$	Helminth	7
Encephalitozoon cuniculi $_{\text{E,Z}}$	Fungi	9	West Nile virus $_{\text{E, Z}}$	Virus	7
Fasciola hepatica $_{\text{NE, Z}}$	Helminth	9			





Emerging versus not emerging

	AOR	95% CI	
Number of hosts species			
1	baseline	-	-
2	1.41	0.75	2.63
>2	**4.82**	3.00	7.74
Taxonomic division			
Bacteria	baseline	-	-
Fungi	1.02	0.59	1.75
Helminths	**0.18**	0.09	0.37
Protozoa	**2.06**	1.00	4.21
Viruses	**6.16**	3.94	9.63
Zoonotic status			
Non-zoonotic	baseline	-	-
Zoonotic	**1.64**	1.07	2.53
Hosmer-Lemeshow goodness-of-fit test		*P*=0.47	

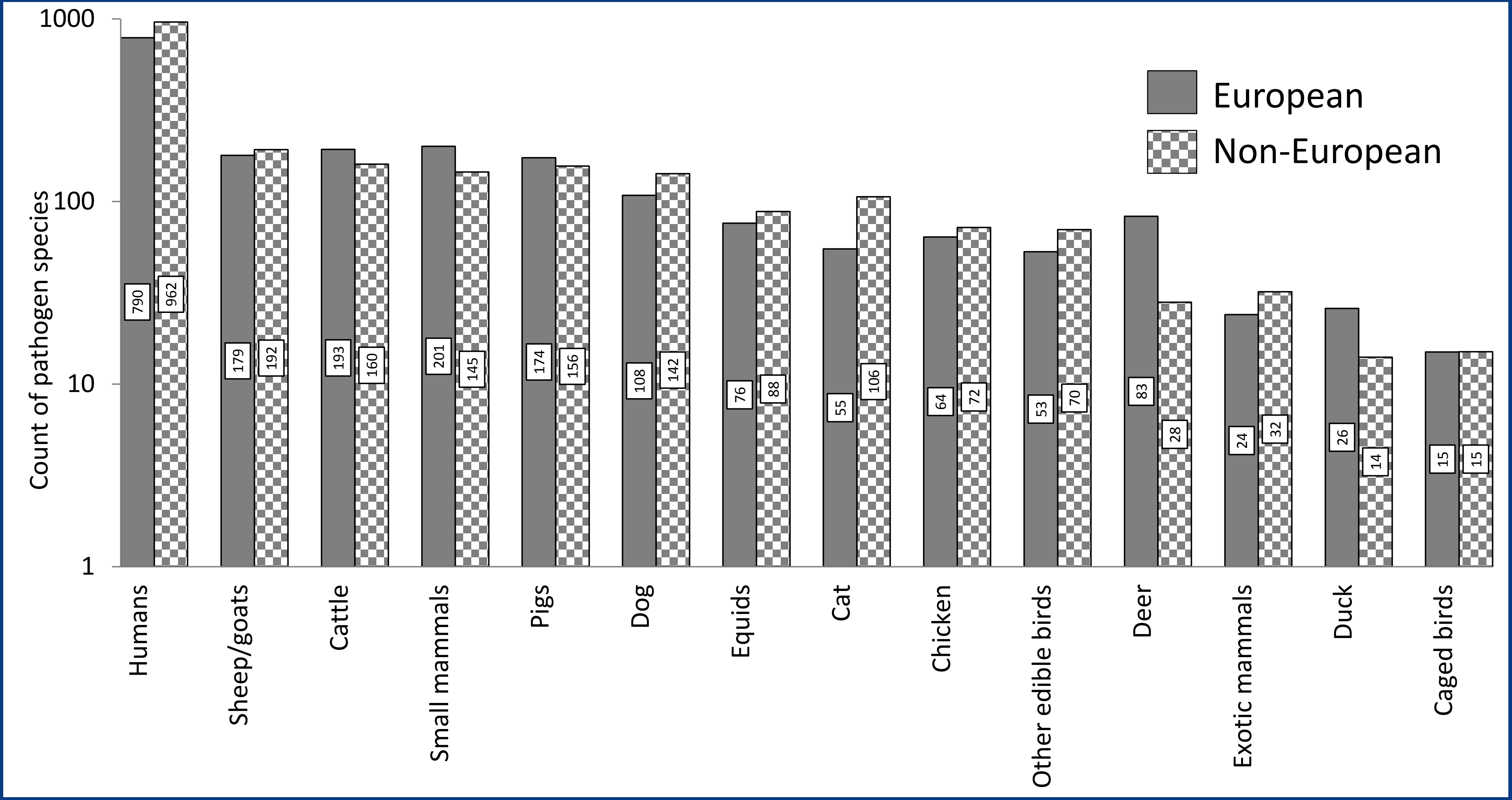


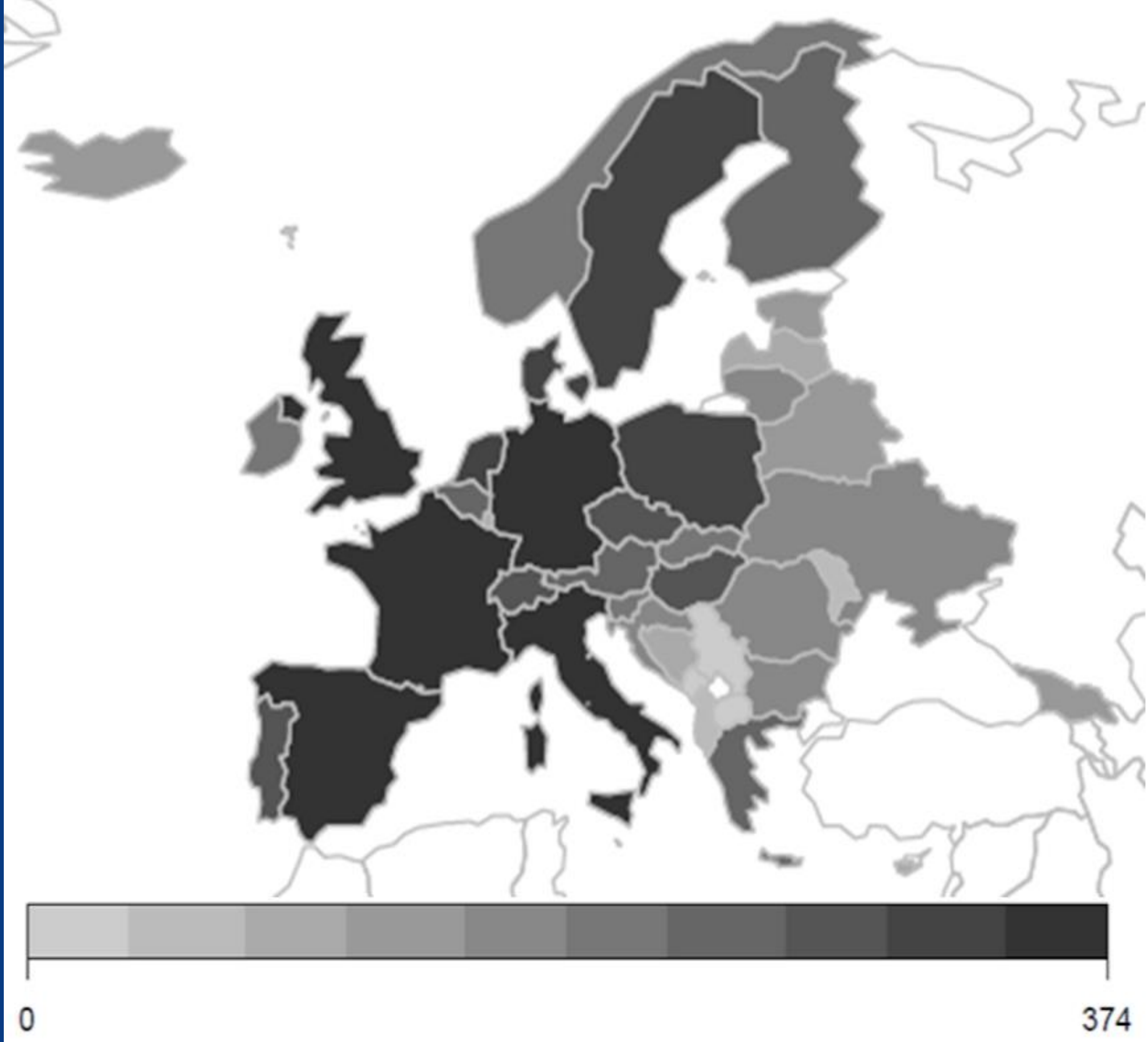
Zoonotic versus non-zoonotic

	AOR	95% CI	
Number of hosts species			
1	baseline	-	-
2	**3.53**	2.18	5.72
>2	**10.92**	5.89	20.23
Taxonomic division			
Bacteria	baseline	-	-
Fungi	0.75	0.55	1.02
Helminths	**16.35**	9.04	29.59
Protozoa	**2.29**	1.17	4.50
Viruses	**3.77**	2.56	5.56
Hosmer-Lemeshow goodness-of-fit test		*P*=0.54	

Host range for human & domestic animal hosts

4223 host-pathogen interactions for humans & domestic animals

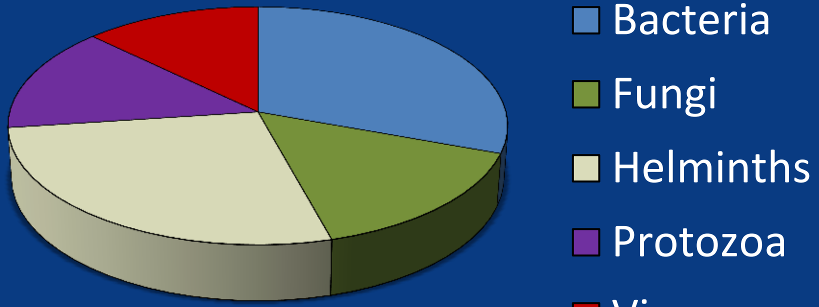



0
374

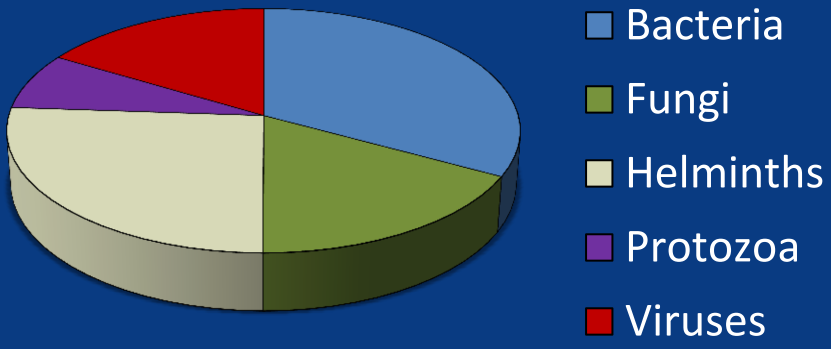
European versus non-European pathogens

	AOR	95% CI	
Number of hosts species			
1	baseline	-	-
2	**2.97**	1.97	4.46
>2	**6.66**	4.44	10.01
Taxonomic division			
Bacteria	baseline	-	-
Fungi	0.94	0.70	1.26
Helminths	**0.17**	0.11	0.25
Protozoa	**0.48**	0.26	0.91
Viruses	**0.39**	0.27	0.59
Emerging status			
Not emerging	baseline	-	-
Emerging	**3.68**	2.43	5.58
Hosmer-Lemeshow goodness-of-fit test		*P*=0.22	

EID2



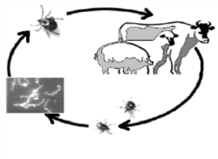
Cleaveland et al. (2001)



% pathogens affecting >1 host	
EID2	29.1
Cleaveland et al. (2001)	62.7

% human pathogens also affecting d. animals	
EID2	25.1
Cleaveland et al. (2001)	39.1

% d. animal pathogens also affecting humans	
EID2	34.3
Cleaveland et al. (2001)	39.4



Conclusions

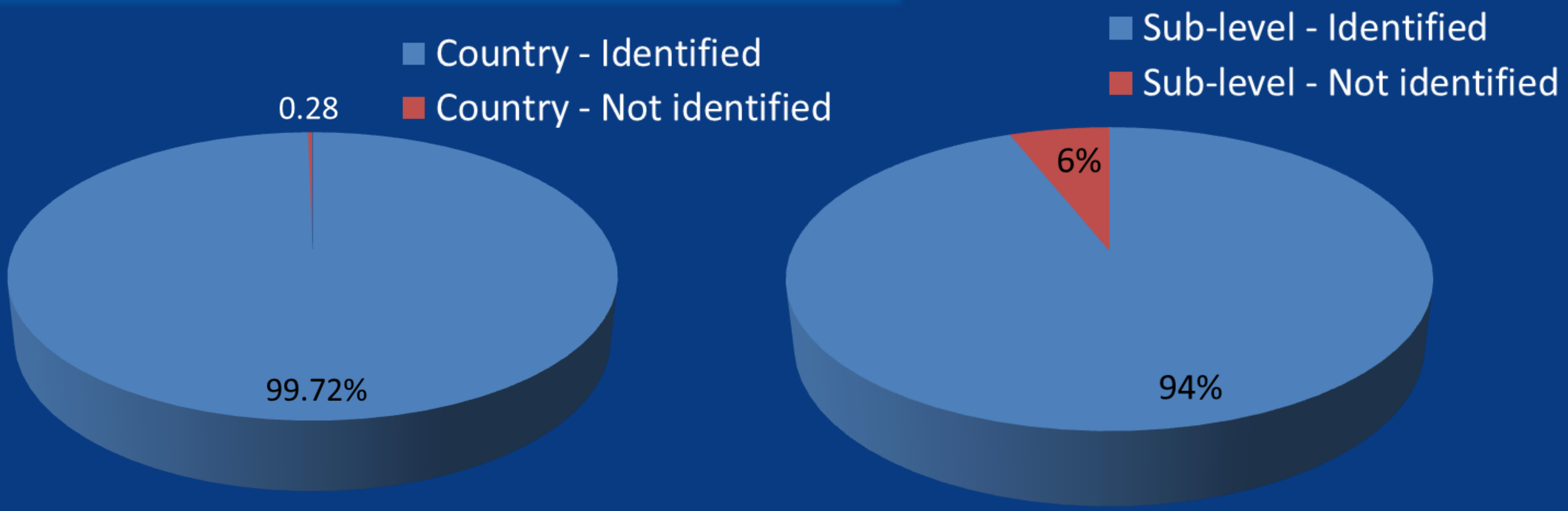
- EID2 objectively utilises individual reports of host-pathogen interactions from open sources
- Given threshold numbers of papers, semi-automated literature gathering works
- EID2 information can be used to establish:
  - Host species of pathogens
  - Pathogen species affecting hosts
  - Quantify connectedness within host-pathogen networks
  - Estimate where (and when?) pathogens & their vectors occur & predict their future distributions given future climate
  - The drivers of pathogens, given available data

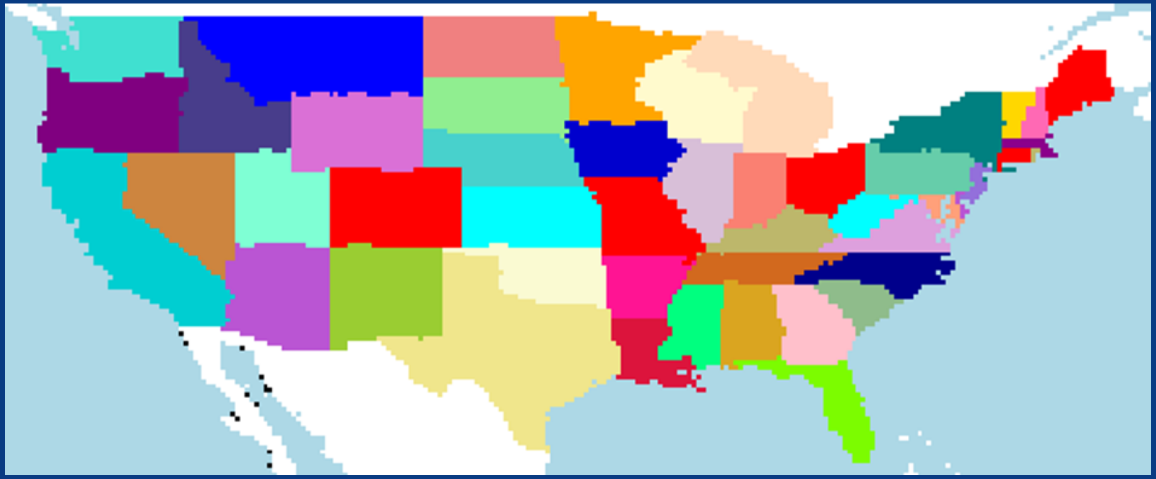
- However, EID2 has limitations due to biases in data sources

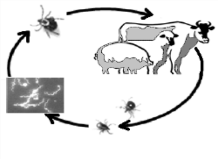
⇨ Please submit your DNA/RNA sequences with detailed meta-data!

http://www.zoonosis.ac.uk/eid2

1) Improved spatial resolution of pathogens

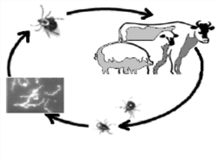






2) Improved host species identification from meta-data

Type of name rule	Example	Extra matched host species
Common	domestic dog	4
Scientific	canis lupus familiaris	4
reverse	aries ovis	1
nospacescientific	ovisaries	1
Hyphen	guinea-pig	2
male	peacock	0
female	bitch	0
young	puppy	0
young plural	puppies	0
ine	canine	8
id	canid	2
dairy / egg	dairy cow	2
meat	beef cow	1
pet	pet dog	2
domestic	domestic chicken	2



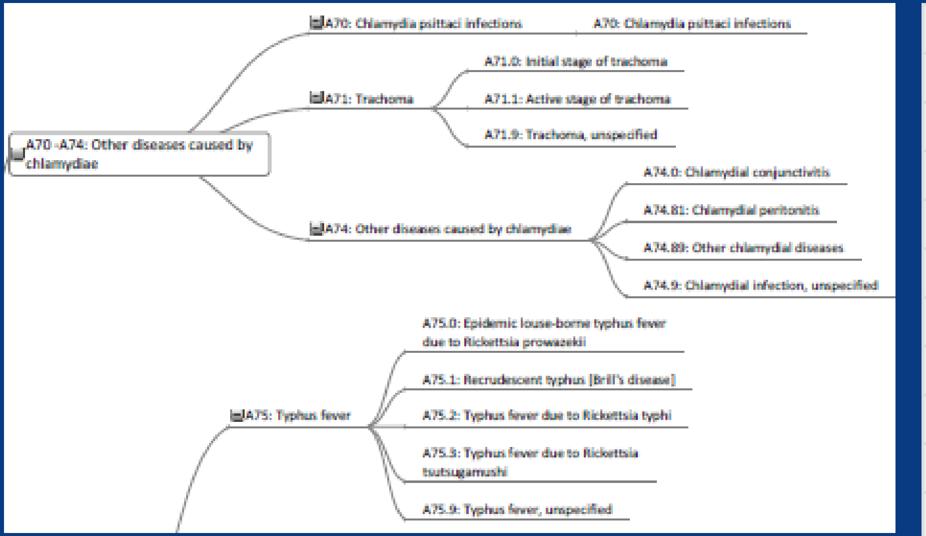
3) Improved removal of false positive hosts

Include everything containing ‘dog’ except:

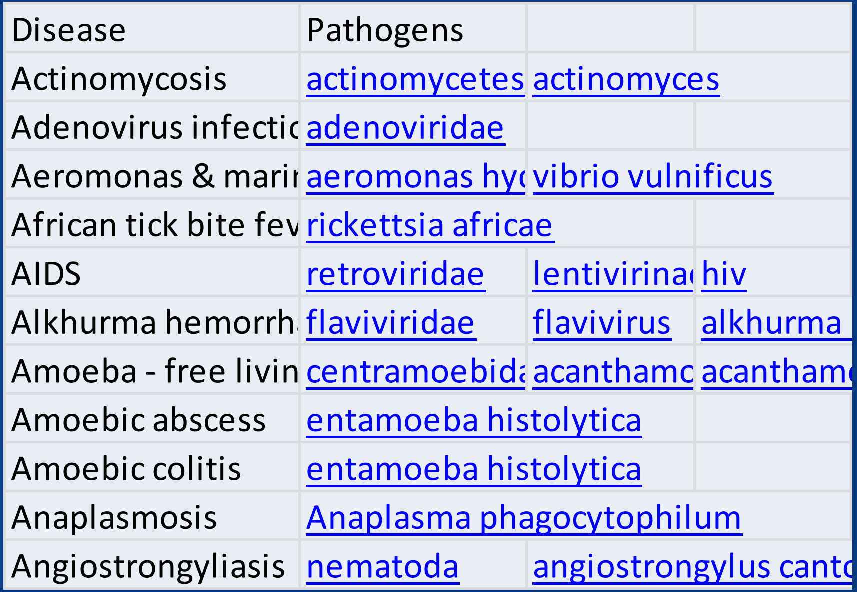
Prairie dog	Dog wood
Racoon dog	Dog fish
Raccoon-dog	Dogs mercury
Dog tick	Dog heartworm
Dogwood	meloidogyne
Dogfish	doguera
Fed on a dog	dog flea
New Guinea wild dog	Dog-faced
Aplysina gerardogreeni	New Guinea singing dog
Squalus melanurus (blacktailed spurdog)	….Many more
Wild dog	

4) Allow users to work at the level of diseases, rather than pathogens

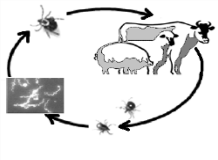
Importing of existing disease classification scheme (ICD)



Recognition & verification of associated pathogens

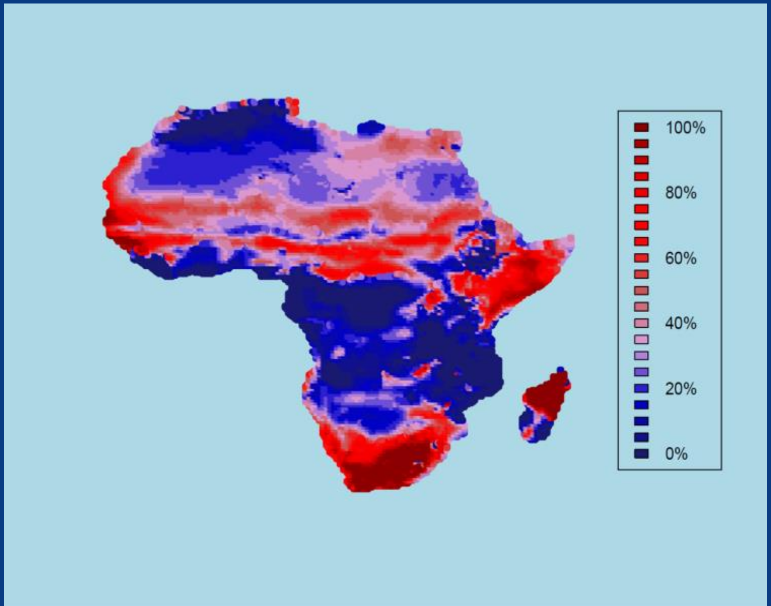


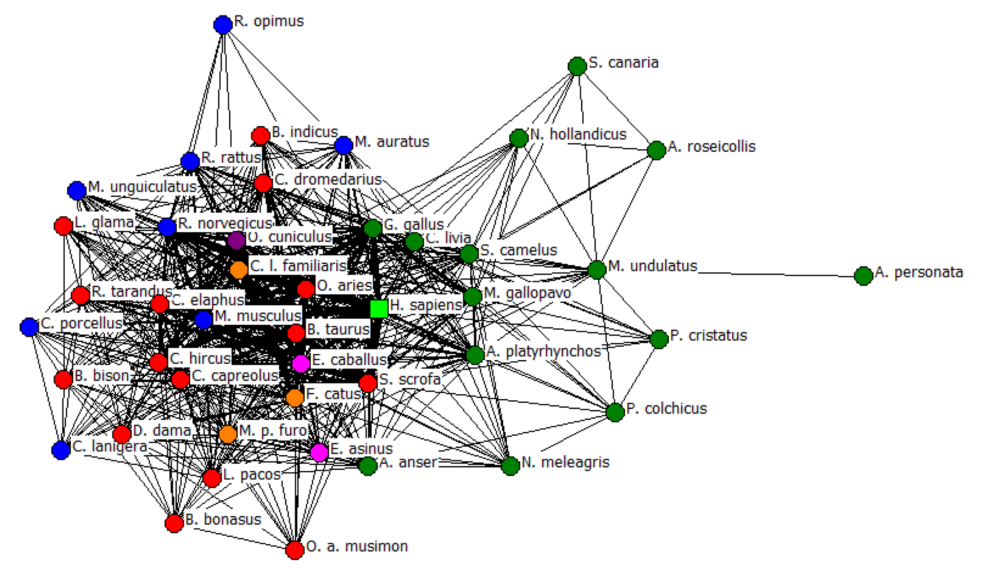
Disease	Pathogens		
Actinomycosis	actinomycetes	actinomyces	
Adenovirus infectio	adenoviridae		
Aeromonas & marin	aeromonas hyd	vibrio vulnificus	
African tick bite fev	rickettsia africae		
AIDS	retroviridae	lentivirinae	hiv
Alkhurma hemorrha	flaviviridae	flavivirus	alkhurma
Amoeba - free livin	centramoebida	acanthamo	acanthamo
Amoebic abscess	entamoeba histolytica		
Amoebic colitis	entamoeba histolytica		
Anaplasmosis	Anaplasma phagocytophilum		
Angiostrongyliasis	nematoda	angiostrongylus canto	



5) Improved Expectation Maximisation likelihood modelling using further climate, demographic & landscape layers

6) Give users the ability to add certain information of their own





- Analysis of host-pathogen relationships amongst the felidae family
- Characterising pathogen networks of primates & mankind – Chester Zoo

Acknowledgements

LUCINDA

UNIVERSITY OF
LIVERPOOL

Prof. Matthew Baylis
Dr. K. Marie McIntyre
Dr. Claire Risley
Dr. Maya Wardeh
Sally Eagle
(Dr. Christian Setzkorn)

Infection Biology

Dr. Alan Radford

defra
Department for Environment
Food and Rural Affairs

Dr. Helen Roberts

Health
Protection
Agency

(Prof. John Stephenson)

NATURAL
ENVIRONMENT
RESEARCH COUNCIL

BBSRC
bioscience for the future

http://www.zoonosis.ac.uk/eid2