Preventive Veterinary Medicine xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed

Using open-access taxonomic and spatial information to create a comprehensive database for the study of Mammalian and avian livestock and pet infections☆

K.M. McIntyre[a,*], C. Setzkorn[a], M. Wardeh[a], P.J. Hepworth[a], A.D. Radford[b], M. Baylis[a]

[a] Department of Epidemiology and Population Health, Institute of Infection and Global Health (IGH), University of Liverpool (UoL), Leahurst Campus, Neston, Cheshire CH64 7TE, UK
[b] Department of Infection Biology, IGH, UoL, UK

ARTICLE INFO

Article history:
Received 5 March 2013
Received in revised form 21 June 2013
Accepted 3 July 2013

Keywords:
Database
Disease
Pathogen
Zoonosis
Emerging
Surveillance

ABSTRACT

What are all the species of pathogen that affect our livestock? As 6 out of every 10 human pathogens came from animals, with a good number from livestock and pets, it seems likely that the majority that emerge in the future, and which could threaten or devastate human health, will come from animals. Only 10 years ago, the first comprehensive pathogen list was compiled for humans; we still have no equivalent for animals. Here we describe the creation of a novel pathogen database, and present outputs from the database that demonstrate its value.

The ENHanCEd Infectious Diseases database (EID2) is open-access and evidence-based, and it describes the pathogens of humans and animals, their host and vector species, and also their global occurrence. The EID2 systematically collates information on pathogens into a single resource using evidence from the NCBI Taxonomy database, the NCBI Nucleotide database, the NCBI MeSH (Medical Subject Headings) library and PubMed. Information about pathogens is assigned using data-mining of meta-data and semi-automated literature searches.

Here we focus on 47 mammalian and avian hosts, including humans and animals commonly used in Europe as food or kept as pets. Currently, the EID2 evidence suggests that:

- Within these host species, 793 (30.5%) pathogens were bacteria species, 395 (15.2%) fungi, 705 (27.1%) helminths, 372 (14.3%) protozoa and 332 (12.8%) viruses.
- The odds of pathogens being emerging compared to not emerging differed by taxonomic division, and increased when pathogens had greater numbers of host species associated with them, and were zoonotic rather than non-zoonotic.
- The odds of pathogens being zoonotic compared to non-zoonotic differed by taxonomic division and also increased when associated with greater host numbers.
- The pathogens affecting the greatest number of hosts included: *Escherichia coli*, *Giardia intestinalis*, *Toxoplasma gondii*, *Anaplasma phagocytophilum*, *Cryptosporidium parvum*, Rabies virus, *Staphylococcus aureus*, *Neospora caninum* and *Echinococcus granulosus*.

* Corresponding author. Tel.: +44 151 794 6079; fax: +44 151 794 6005.
E-mail address: k.m.mcintyre@liv.ac.uk (K.M. McIntyre).

0167-5877/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.prevetmed.2013.07.002

- The pathogens of humans and domestic animal hosts are characterised by 4223 interactions between pathogen and host species, with the greatest number found in: humans, sheep/goats, cattle, small mammals, pigs, dogs and equids.
- The number of pathogen species varied by European country. The odds of a pathogen being found in Europe compared to the rest of the world differed by taxonomic division, and increased if they were emerging compared to not emerging, or had a larger number of host species associated with them.

1. Introduction

A great deal of time and resources have been put into studying individual pathogens or groups of pathogens affecting the animals with which humans have the most contact, because they potentially affect food security including socio-economic impacts and zoonotic disease transmission. For instance, it is estimated that one in four people in the UK annually suffer from diarrhoeal disease (Tam et al., 2012) and as previous work has suggested that around 60% of infectious organisms known to be pathogenic to humans are zoonotic (Taylor et al., 2001; Woolhouse and Gowtage-Sequeria, 2005), it is likely that many of these episodes of illness are a result of preceding transmission from animals, causing emerging infections in humans.

Just over 10 years ago, the first attempt to produce an inventory of the pathogens of humans was undertaken (Taylor et al., 2001). At the same time, this list was combined with further information for livestock, dogs, cats and wildlife, in a study which aimed to quantify both pathogen characteristics and their interactions with key features of pathogen–host epidemiology including host range, zoonotic or emerging disease status and socio-economic importance (Cleaveland et al., 2001). The main material for both studies came from gathering specific information from textbooks; however, scientific literature was also examined to ascertain emerging infections. Such sources are potentially biased towards clinical infections and the data gathering would have been a lengthy process. There may be gaps in knowledge and therefore biases in animal and particularly companion animal disease data due to a lack of joined-up surveillance, which several recent projects aim to rectify (Moore et al., 2004a,b; Paiba et al., 2007; Radford et al., 2010). In addition, the job of identifying publications on pathogens and their hosts may be made more difficult due to our domestication of animals. For example, pathogens are found in domestic hosts in which they would not, naturally, be expected to occur, and husbandry practices can introduce host or pathogen vector species into new areas, change host susceptibility or behaviour, exposing hosts to new pathogens via modified transmission routes. Previous control programmes using, for instance antibiotics, can also promote further evolution of pathogens, changing their pathogenesis in host populations.

Within the present study, we describe the creation of an open-access pathogen database which was constructed as a part of the ENHanCE project (McIntyre et al., 2010), and present a number of outputs from the database to demonstrate its value as a tool. Within the results section, we focus on 47 mammalian and avian hosts, including humans and animals commonly used in Europe as food or kept as pets.

The open-access evidence-based ENHanCEd Infectious Diseases (EID2) database (University of Liverpool, 2011), provides a large, automated and systematically generated method of studying the main pathogens and hosts involved in disease transmission. It describes the pathogens of humans and animals, their host and vector species, and also their global occurrence, and the information contained within it is likely to reflect biases in the research undertaken on pathogens and their hosts.

The main way in which the results of our study differ from that of Cleaveland et al. (2001) is in the use of the EID2 source for pathogen information. All evidence within the EID2 comes from, and is linked to, previously published sources; the database extracts and analyses material contained in the meta-data of millions of nucleotide sequences and in publications, storing it in a hierarchical phylogenetic tree structure. As a result, where Cleaveland et al. (2001) used textbooks to find pathogens of specific hosts, in the EID2 approach the evidence comes from individual reports. The semi-automated nature of information gathering has also meant: a much larger quantity of proof of host–pathogen interactions has been used as evidence; this proof comes from primary, usually peer-reviewed literature; it has been possible to study many more pathogens if information has been published on them; and it has been possible to use a more exhaustive list of domestic animal hosts. In addition, spatial information for pathogens assigned at the country-level has been built into the EID2. Within this study we have examined differences in the evidence for the occurrence of pathogens in Europe compared to the rest of the world.

Our main aims were to: (1) Provide a description of the structure of the EID2 database including the data-sources used to create it. (2) Carry out an analysis demonstrating the usefulness of the EID2. This was achieved by comparing some of the content of the EID2 with the results of several earlier seminal papers. We show that the EID2 can be used to recreate these results potentially in a quicker, less biased, more easily repeatable and updateable way. The information it contains is a reflection of biases in the scientific research which is undertaken on pathogens and their hosts, however the database can be quickly updated when new information become available. Further, we emphasise that the EID2 is a much bigger resource which could be adapted to answer questions on other species, vector and pathogen assemblages, and on other drivers of disease.

Please cite this article in press as: McIntyre, K.M., et al., Using open-access taxonomic and spatial information to create a comprehensive database for the study of Mammalian and avian livestock and pet infections. PREVET (2013), http://dx.doi.org/10.1016/j.prevetmed.2013.07.002