An integrated database on ticks and tick-borne zoonoses in the tropics and subtropics with special reference to developing and emerging countries

Published version:

DOI:10.1007/s10493-010-9414-4

(Article begins on next page)

An integrated database on ticks and tick-borne zoonoses in the tropics and subtropics with special reference to developing and emerging countries

Umberto Vesco[1], Natasa Knap[2], Marcelo B. Labruna[3], Tatjana Avsic Zupanc[2], Agustín Estrada-Peña[4], Alberto A. Guglielmone[5], Gervasio H. Bechara[5], Arona Gueye[7], Andras Lakos[8], Anna Grindatto[1], Valeria Conte[1], Daniele De Meneghi[1*]

1 - Università di Torino, Torino (Italy)
2 - Univerza v Ljubljani, Ljubljana (Slovenia)
3 - Universidade de São Paulo, São Paulo (Brazil)
4 - Universidad de Zaragoza, Zaragoza (Spain)
5 – Instituto Nacional de Tecnología Agropecuaria, Rafaela (Argentina)
6 - Universidade Estadual Paulista, Jaboticabal (Brazil)
7 - Institut Sénégalais de Recherches Agricoles, Dakar (Senegal)
8 - Center for Tick-borne Diseases, Budapest (Hungary)

*corresponding author: Daniele De Meneghi, Via Leonardo da Vinci, 44 10050 Grugliasco (TO) Italy email: daniele.demeneghi@unito.it

ABSTRACT.

Tick-borne zoonoses (TBZ) are emerging diseases worldwide. A large amount of information (e.g. case reports, results of epidemiological surveillance in humans, vertebrate hosts and vectors, etc.) is dispersed through various reference sources (ISI and non-ISI journals, conference proceedings, technical reports etc.). As part of the activities of the ICTTD-3 project (http://www.icttd.nl), an integrated database was developed in order to gather TBZ records in the (sub-)tropics, collected both by the authors and collaborators worldwide. Data collected are then made freely available to researchers for analysis by spatial methods, integrating mapped ecological factors to be used for predicting TBZ risk. The authors present the assembly process of the TBZ Database: the compilation of an updated list of TBZ relevant for the tropics and subtropics, the design and structure of the Database, the method of bibliographic search, the assessment of spatial precision of geo-referenced records.

At the time of writing, 868 records extracted from 388 publications related to 59 countries in the (sub-)tropics, have been entered in the Database. TBZ distribution maps were also produced. Imported cases have been also accounted for. The most important datasets with geo-referenced records were those on SFG rickettsiosis in Latin-America and CCHF in Africa.

The authors stress the need for international collaboration in data collection to update and improve the Database. Supervision of data entered still remains necessary. Means to foster collaboration are discussed. The paper is also intended to describe the challenges encountered to assemble spatial dataset from various sources and to help developing similar data collections.

INTRODUCTION.

Ticks constitute a serious economic and health problem worldwide. Ticks can transmit a great number of pathogens to both animals and humans. Zoonotic tick-borne agents such as protozoa, rickettsiae, spirochaetes and viruses are having an increasing impact on public health in tropical and subtropical countries (Parola *et al.* 2005, Parola *et al.* 2005a). Furthermore ticks are responsible for both direct damage and decrease in animal production, especially in tropical and sub-tropical areas, where parasite control is more difficult due to lack of effective measures and great habitat suitability for tick growth (Jongejan and Uilenberg, 2004).

The Integrated Consortium on Ticks and Tick-borne Diseases (ICTTD-3, http://www.icttd.nl) is a coordinated action financed by the International Cooperation Program (INCO) of the European Union.
The ICTTD-3 project is jointly executed by a consortium of 43 institutions in 29 countries, and supports different research lines which include: (1) study of biosystematics and molecular phylogeny of ticks and tick-borne pathogens, (2) development of new molecular diagnostic tools, (3) design of novel integrated vaccine strategies targeting ticks and tick-borne pathogens, (4) creation of a geographical database on tick distribution; (5) development of a geographical database on the occurrence of tropical and sub-tropical tick-borne zoonoses (TBZ). It is the latter that forms the subject of this paper.

Since a great amount of information concerning TBZ has been published, bibliographic data need to be gathered and analyzed by means of integrated databases to provide usable information for future research activities, analogously to Guerra *et al.* (2006) for malaria, and Murty *et al.* (2005) and Kumar *et al.* (2005) for filariasis Data sharing is an important resource in vector-borne disease research as it enables the involvement of a large number of collaborators and the public-access databases developed are useful in providing preliminary information for further studies (Moffet *et al*., 2009 and Craig *et al*., 2007), as well as in guiding adequate investments for fighting the diseases (Guerra *et al*., 2006).
Tick-borne diseases, including TBZ, are strongly related to environmental factors limiting vector distribution. The assessment of ticks and TBZ risk is now feasible on a large scale thanks to the recent availability of free accessible digital climatic and ecological layers, the development of geographic information system (GIS) technology, increased computer power available and the advancements in ecological niche modelling (i.e. Kitron, 2000; Pearce and Boyce, 2006; Tsoar *et al.,* 2007).
The scope of this paper is to describe the development of an integrated Database (focused on ticks-hosts-pathogens interface) on the geographical distribution of TBZ in the (sub-)subtropics, with particular emphasis on developing and emerging countries, according to the definition and the list produced by the International Monetary Fund - IMF (http://www.imf.org/external/pubs/ft/weo/2010/01/weodata/groups.htm#oem).
The main aim of our Database is to gather geo-referenced data on the presence of ticks, hosts and pathogens to produce updated TBZ distribution maps (available online at http://www.icttd.nl/index.php?id=12), and eventually develop ecological niche models for vectors and zoonoses and TBZ risk maps in the (sub-)tropics. The Database management and co-ordination of this work was carried out by the Dept. of Animal Production, Epidemiology & Ecology, University of Turin, Italy.

MATERIALS AND METHODS.

Database structure.

A Microsoft® Access (2000/XP) relational database was developed for collection and management of TBZ data. Data were structured in 3 related tables: operator, bibliographic source and TBZ, respectively.

a. Operator table: allows unambiguous identification of the operator through the e-mail address entered as ID as well as the institution involved.

b. Bibliographic source table: references were recorded according to BibTeX standard field requirements. The table collects basic bibliographical information such as title, journal, author(s) and publication year. Further optional information was collected about month of publication, volume, number and pages. In case the abstract only was available online, the reference is included in a list published on the project web page and periodically updated, so that collaborators having access to the full-text publication can send it to the project team.

c. TBZ data table: this table gathers information on detected pathogens, diseases and diagnostic methods used, type of data collection (active or passive surveillance), number of positives over tested samples (active surveillance) or number of clinical cases (passive surveillance or case reports), vertebrate species involved (incl. humans) and vector ticks. Geo-referenced location and record date (month/year) were entered. A field was dedicated to include imported TBZ cases in countries outside the (sub-)tropical belt.

The tables are connected by common identification fields (IDs) with one-to-many relationships: all publications are connected to the operator who entered them; each publication record is then connected to all TBZ records retrieved from that publication. All tables are connected to a user-friendly entry form (Figure 1) to facilitate unskilled users in entering data. As the Database is intended to be compiled by many users, also occasional ones, standardization and checking for data consistency are key issues. The forms were designed with the aim to standardize data format by means of automated data check, implemented through VBA programming (2,785 lines of code). This was intended to reduce the risk of record duplication, and to obtain "clean" data, ready to be directly analysed.

A dedicated website (http://www.icttd.nl/index.php?id=12) was created to facilitate collaboration in data collection. A blank copy of the Database is available for collaborators to download, to fill in the forms with TBZ records and send it back to the co-ordinating institution. Its responsibility is to receive, gather, check and merge in the master database the various records provided by the collaborators, and to periodically update the list of bibliographic references and the TBZ distribution maps. These resources are made available for free consultation on the ICTTD project website.

Data sources.

Most data were gathered through electronic search using CAB Abstracts (www.cabi.org), PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Scopus (http://www.scopus.com) bibliographic databases, as well as Google Scholar (scholar.google.com) and general search engines. As a search strategy, we included all scientific and common names for both pathogen and diseases or syndromes it causes and the names of states and continents.

To reduce the number of queries we used Boolean operators with the following syntax:

(<pathogen/disease name variant 1 > OR <variant 2> … OR <variant N>) AND (< state 1> OR <state 2> … OR <continent 1> OR <continent 2> …).

Search results were checked by an operator looking for relevant titles. Abstracts and, when available, full texts were read in order to retrieve information to be entered in the Database. Additional data were obtained by checking and including relevant bibliographic references from the articles already obtained.

Other main data sources were official websites of national and international organizations dealing with animal and human health (i.e. ministries of health, agriculture, veterinary services, FAO, WHO, OIE).
Further data (i.e. unpublished reports from various national and international institutions, conference proceedings and personal communications from relevant experts, etc.) were also collected through collaboration -within and outside the ICTTD3 network- with field scientists, biologists, veterinarians, medical doctors, and public health officers from research institutions worldwide. Other potential collaborators outside the direct network of contacts of the authors (researchers, officials of relevant ministries, etc.) were approached using e-mails, letters to obtain data which would not have been available through other sources.
To avoid the potential duplication of data by different collaborators in case of data published more than once (e.g. research notes, conference proceedings, full paper), only the record with greater data accuracy (primary record) was used as a reference record to fill in the related records in the "TBZ data" table. The secondary publications are then entered by completing only the "bibliographic sources" table in order to point out that pathogen or TBZ data have already been entered in the Database. Data collected were finally reviewed by an operator in order to check for possible duplications.

Information on TBZ

In order to define TBZ geographical distribution, key data include the species of pathogens involved, geographical location and recording date. When available, additional information on TBZ reservoirs, vectors, type of study and diagnostic method were also collected. We focused on data referred to the tropical and sub-tropical zones, with reference to developing and emerging countries. These study areas were selected for initial inclusion in our Database because of the epidemiological importance of these diseases in those countries with generally scarce resources for TBZ prevention and control. However, although the definition of subtropical zone is not referred to latitude only but also to eco-climatic features (rare sub-0º C periods), as maps defining the geographical limits of subtropical zones are not available, we arbitrarily approximated it to the belt between 35°N and 35°S latitude. Thus only records geographically referenced to locations within these latitudes were included.

A description of the information/data collected in our Database is now presented.

Tick-borne pathogens

A list of the zoonotic tick-borne agents reported in the (sub-)tropical belt was compiled according to updated scientific literature and with the involvement of TBZ experts within the ICTTD-3 project and related networks. With regard to standard taxonomy, we referred to the List of Prokaryotic names with Standing in Nomenclature (http://www.bacterio.net/ last visited July 2010), Universal Virus Database of the International Committee on Taxonomy of Viruses (http://www.ncbi.nlm.nih.gov/ICTVdb; last visited July 2010). The tick-borne agents recognized with a zoonotic potential (Acha and Szyfres, 2003; Hugh-Jones et al., 2000; Labruna, 2009; Hubalek, personal communication) are listed in the following table (Table 1).

TBZ	**Agents**
Babesiosis	*Babesia bovis, B. divergens, B. microti*
Lyme borreliosis	*Borrelia afzelii, B. burgdorferi s.s., B. garinii, B. lusitaniae, B. valaisiana, B. spielmanii*
Relapsing fever	*Borrelia crocidurae, B. duttonii, B. hispanica, B. persica, B. turicata, B. venezuelensis*
Human ehrlichiosis	*Ehrlichia canis, E. chaffeensis, E. ewingii, E. ruminantium,*

	Anaplasma phagocytophilum
Spotted fever group rickettiosis	*Rickettsia aeschlimannii, R. africae, R. australis, R. conorii, R. felis, R. helvetica, R. honei, R. massiliae, R. monacensis, R. parkeri, R. rhipicephali, R. rickettsii, R. sibirica, R. sibirica mongolotimonae, R. slovaca*
Typhus group rickettsiosis	*Rickettsia prowazekii** * usually loose-borne.
Q Fever	*Coxiella burnetii*
Arbovirosis	*Bunyaviridae,* genus *Nairovirus:* Crimean-Congo haemorrhagic fever Virus (CCHFV), Nairobi sheep disease Virus (NSDV), Dugbe Virus (DUGV), Ganjam Virus (GANV)*; Bunyaviridae,* genus *Orthobunyavirus:* Bhanja Virus (BHAV); *Bunyaviridae*, unclassified genera: Issyk-Kul Virus (ISKV), Keterah Virus (KTRV), Tamdy Virus (TDYV), Wanowrie Virus (WANV) *Flaviviridae:* Alkhurma Virus (AHFV), Kyasanur Forest disease virus (KFDV), Karshi Virus (KSIV) *Orthomyxoviridae:* Dhori / Batken Virus (DHOV), Thogotovirus (THOV) Unclassified: Quaranfil Virus (QRFV)

Table 1: zoonotic tick-borne agents reported in the (sub-)tropical belt

<u>Species involved.</u>
We considered records of TBZ pathogens detected in human and animal hosts, and ticks (host seeking ticks). In case of tick feeding on-the-host, both host and tick species were reported. Tick species were recorded according to the comprehensive lists by Horak *et al*. (2002) and Guglielmone *et al.* (2010).

<u>Imported cases</u>
TBZ are an emerging problem for travel medicine. Travelling to endemic countries can result in infection with TBZ pathogens, so we also included TBZ imported cases and respective country of origin, in order to collect information on possible cases and outbreaks of (sub-)tropical TBZ in non (sub-)tropical countries.

<u>Diagnosis.</u>
We recorded the diagnostic method used, as cited in the paper, according to the following types: clinical diagnosis, serology, microscopy, PCR, isolation. Such information is very important to assess the reliability of the record, as the specificity of diagnosis varies greatly according to the method used. With regard to serological diagnosis, we included only those records where a sero-conversion was reported in the bibliographic source and/or a gold standard diagnostic test was used for confirmation.

<u>Type of data collection</u>
We included records obtained from surveys (active data collection) and from clinical case reporting (passive data collection). Active data collection can be defined as an ongoing or periodic collection of samples/data based on a specific design, which involves sampling of the whole population. Passive data collection is the reporting of disease cases occurring in a population, without the active selection of units through a sampling process (Dufour and Hendrikx, 2009).

We recorded the number of individuals sampled and the number of positive results in case of active surveillance, and the number of clinical cases in case of passive surveillance.

Time

We included records from both cross-sectional and longitudinal studies. In the latter case, we recorded both start and end dates of data collection. No time limit for data inclusion into the Database was set (the oldest record dating 1910).

Data geo-positioning.

Ideally data on TBZ cases should be published along with the respective coordinates measured with a GPS receiver. More commonly, a paper provides a geographic name referring to a village or community and/or to different administrative levels (municipalities, departments and provinces, and higher levels up) or a description suggesting a location related to a precise location (i.e. 20 km SW from the place X, along provincial road Y). Each toponym does not designate a point -in the geometric sense- over the earth surface, but a geographic extension of competence, as a municipality name does not designate a point, but an area, with specific extension and shape, called geographic footprint (Hill, 2000). It is recognised that such footprints may vary over the time (Vestavik, 2003).

Moreover, location names reported in literature can be misleading, ambiguous and/or imprecise for numerous reasons: i) several places can share the same name, so that the place name is unique only within a limited geographic area (i.e. municipality, county) but in a wider area (i.e. province, state) more locations with the same name can be present; ii) some place names used in texts are "historical" or cultural conventional names rather than official names, requiring more understanding to be able to link them to the true geographic location; iii) some official place names changed over time; iv) the same place name can be written differently in different texts, because the author has misspelled the name or because there are different legal spellings in different languages or different transliterations from a non-Latin alphabet (Larson, 1996; Vestavik, 2003).

The reliability of data for different scale of analysis varies along with the accuracy (Hengl, 2006). We chose to record all data, but we added a quality field for the geographical accuracy in order to be able to filter the appropriate data for every (each) analysis. Choices are: (1) coordinates provided by authors, (2) coordinates of a geographical place retrieved from a gazetteer, (3) centroid of an administrative division. When not provided by authors, coordinates were found through a web gazetteer, mainly GeoNet Names Server (http://earth-info.nga.mil/gns/html/index.html) which provides names, along with the administrative units in which it is found, helping to discriminate between identical or similar names. We entered the coordinates in decimal degrees format with at least two decimal places of precision (~1.2 km).

Spatial accuracy.

Data are of high quality if they are fit for their intended use (i.e. in operations, decision-making, and planning) (Chapman, 2005). In order to use data in ecological studies the accuracy of the geographical coordinates retrieved from a gazetteer needs to be assessed.

To assess the accuracy of the coordinates retrieved, we chose to apply the same georeference procedure to a tick dataset (n=108 records) provided by one of the authors (AG). Location names of tick collection sites along with the respective coordinates, taken with a GPS receiver, were given in that dataset. So, we have been able to compare the coordinates from the dataset (obtained by a GPS) with those, referred to the same place names, extracted from the selected gazetteer. This procedure made it possible to calculate the positional error (as an Euclidean distance) of each pair of coordinates from the "true" point. The distance has been geometrically calculated: $[(x_1 - x_2)^2 + (y_1 - y_2)^2]^{1/2}$, where x_1 is the GPS longitude, x_2 is the longitude extracted from gazetteer, y_1 is the GPS latitude, and y_2 is the latitude extracted from gazetteer.

RESULTS.

Data sources.

We extracted 868 records, belonging to 388 publications, referred to the emerging and developing countries of the tropical and sub-tropical belt (latitude between 35°N and 35°S).

Out of 388 publications, 350 (90%) were scientific papers, 19 conference proceedings, 6 master's theses, 4 PhD theses, 2 book chapters, 2 technical reports and the remaining 5 were from other sources (named "miscellanea", including websites).

Data collected.

TBZ pathogens

Table 2 includes data on active and passive data collection in humans and animals, and records of detection of the agents in ticks.

	human		animal		tick
PATHOGENS	**passive**	**active**	**passive**	**active**	
Babesia bigemina	0	1	0	7	0
Babesia bovis	0	1	0	14	2
Babesia canis	0	0	0	2	0
Babesia microti	0	0	0	2	0
Babesia sp.	9	5	0	8	1
Tot. Babesiosis agents	**9**	**7**	**0**	**33**	**3**
Borrelia crocidurae	5	4	0	9	**1**
Borrelia sp.	4	3	2	4	3
Tot. Relapsing fever agents	**9**	**7**	**2**	**13**	**4**
Borrelia valaisiana	1	0	0	0	0
Borrelia burgdorferi s.l.	1	3	0	0	0
Tot. Lyme borreliosis agents	**2**	**3**	**0**	**0**	**0**
Anaplasma marginale	0	1	0	3	1
Anaplasma phagocytophilum	1	3	1	4	2
Anaplasma sp.	0	2	0	0	0
Ehrlichia canis	1	1	6	19	5
Ehrlichia chaffeensis	8	8	0	5	7
Ehrlichia ewingii	0	0	0	1	1
Ehrlichia ruminantium	2	0	2	24	0
Ehrlichia sp.	2	3	0	1	9
Tot. Ehrlichiosis agents	**14**	**18**	**9**	**57**	**25**
Coxiella burnetii	**4**	**18**	**0**	**14**	**4**
Rickettsia prowazekii	**0**	**1**	**0**	**0**	**1**
Rickettsia aeschlimannii	2	0	0	0	5
Rickettsia africae	20	1	0	0	16
Rickettsia bellii	0	0	0	0	3
Rickettsia conorii	47	22	7	16	9
Rickettsia felis	3	1	0	0	4
Rickettsia helvetica	0	1	0	0	0
Rickettsia honei	2	0	0	0	1
Rickettsia massiliae	0	0	0	0	2
Rickettsia monacensis	0	0	0	0	1
Rickettsia parkeri	0	1	0	2	4

Rickettsia rhipicephali	0	0	0	0	1
Rickettsia rickettsii	103	22	3	14	15
Rickettsia sibirica	0	1	0	0	1
Rickettsia sibirica mongolotimonae	2	0	0	0	1
Rickettsia slovaca	0	0	0	0	1
Rickettsia sp.	41	27	8	15	33
Tot. Spotted fever group (SFG) rickettsiosis	**220**	**76**	**18**	**47**	**97**
Alkhurma Virus (AHFV)	4	0	0	0	1
Bhanja Virus (BHAV)	0	1	0	2	0
Crimean-Congo Haemorrhagic Fever Virus (CCHFV)	36	17	0	37	17
Dhori/Batken Virus (DHOV)	0	1	1	1	1
Ganjam Virus (GANV)	0	0	0	0	5
Kyasanur Forest Disease virus (KFDV)	5	4	4	2	4
Wanowrie Virus (WANV)	0	8	0	2	0
Tot. Arboviruses	**45**	**31**	**5**	**44**	**28**

Table 2: Number of records about active (act.) and passive (pass.) data collection in human and vertebrate hosts (animals) and in ticks

Table 3 shows the number of human cases (passive data collection) retrieved about each pathogen and the number of positives over humans tested (active data collection).

	Active data collection		**Passive data collection**
Pathogens	**positives**	**tested**	**cases**
Babesia bigemina	17	47	
Babesia bovis	21	47	
Babesia sp.	125	7528	15
Tot. Human babesiosis agents	**163**	**7622**	**15**
Borrelia burgdorferi s.l.	78	424	2
Borrelia valaisiana			1
Tot. Lyme disease agents	**78**	**424**	**3**
Borrelia crocidurae	94	9560	27
Borrelia sp.	97	1608	57
Tot. Relapsing Fever agents	**191**	**11 168**	**84**
Anaplasma marginale	12	47	
Anaplasma phagocytophilum	20	100	2
Anaplasma sp.	68	7107	
Ehrlichia canis	6	20	1
Ehrlichia chaffeensis	101	2960	8
Ehrlichia ruminantium			7
Ehrlichia sp.	50	292	5
Tot. Human Ehrlichiosis agents	**257**	**10 526**	**23**
Coxiella burnetii	**197**	**2057**	**51**
Rickettsia aeschlimannii			2
Rickettsia africae	7	118	62
Rickettsia conorii	540	2223	99
Rickettsia felis	1	46	6
Rickettsia honei			1
Rickettsia helvetica	5	46	
Rickettsia parkeri	36	349	

Rickettsia rickettsii	393	3179	532
Rickettsia sibirica			2
Rickettsia sp.	2 507	6 6378	235
Tot. Spotted fever group rickettsiosis	**3 489**	**72 339**	**939**
Alkhurma Virus (AHFV)			72
Bhanja Virus (BHAV)	4	43	
Crimean-Congo haemorrhagic fever Virus (CCHF)	769	5 068	420
Dhori / Batken Virus (DHOV)	1	43	
Kyasanur Forest Disease Virus (KFDV)	876	41 508	747
Wanowrie Virus (WANV)	1	43	
Tot. Arbovirosis	**1 645**	**46 705**	**1 239**

Table 3: number of records about active (act.) and passive (pass.) data collection in human and vertebrate hosts (animal) and in ticks

Diagnosis

As reported in Table 4 serology was the diagnostic method most used for confirmation of cases and/or for surveillance (55%, n=868), followed by PCR (19%), isolation (11%) and microscopy (4.1%), respectively. TBZ records diagnosed on clinical basis only were 3.2%, but for 8% of records the method of diagnosis was not clearly reported (i.e. a national report on a given TBZ, that summarizes annual incidence, not specifying the diagnostic method(s) used).

Method of diagnosis	Number of records
clinical	28
serology	479
microscopy	36
PCR	161
isolation	93
not unambiguously reported	71

Table 4: diagnostic methods reported in the bibliographic sources

Imported TBZ

We found that 80 out of 868 (9.2%) records were TBZs cases imported from (sub)tropical areas (Table 5). Most of them were due to Spotted Fever Group (SFG) rickettsiosis agents (50), mainly *Rickettsia africae* and *Rickettsia conorii*. Besides, 23 records of imported Crimean-Congo Haemorrhagic Fever (CCHF) were also collected.

TBZ	Records

Relapsing fever	4
Lyme borreliosis	1
Crimean-Congo haemorrhagic fever	23
Q fever	1
Human ehrlichiosis	1
Spotted fever (incl. Africa Tick bite fever)	50

Table5: Records of TBZ imported cases.

<u>Time</u>

Publications retrieved were within the time span 1911 to 2009 (Figure 2). Until the ’80s the number of publications on TBZ was distributed homogeneously through the time line (~10 papers/decade), while the percentage of papers published during the last 2 decades had grown two fold each decade.

Data reliability has changed over time with the development of new more sensitive and/or specific diagnostic techniques. Figure 3 shows the trend of the diagnostic methods employed during the various decades according to the literature reviewed. Isolation was the most commonly used method until the ‘60s, while serology became the most important diagnostic method throughout the ‘60s-80s. Most recently -during the last 10 years- PCR emerged as the most common diagnostic tool for its specificity and sensitivity, and the number of publication on TBZ adopting this diagnostic method has respectively increased.

<u>Spatial information.</u>

Within the tropical and subtropical belt (latitude between 35°N and 35°S), we obtained records from 59 countries (Figure 4), of which 368 records were referred to the African continent, 307 to Latin America and 193 to Asia. The geo-referenced records were 574, out of which 43 had coordinates provided by the authors, 326 records were geo-referenced by means of gazetteers at the level of a village or other geographical place, and 205 records as the centroid of a first/second level administrative division. The remaining records referred to the country only.

The most reported TBZ in the (sub-)tropics were SFG rickettsiosis (Figure 5), ehrlichiosis (Figure 6) and CCHF (Figure 7).

<u>Spatial accuracy.</u>

It was not possible to retrieve the coordinates of 17 out of 108 locations of tick collection in Argentina from the gazetteer, because of ambiguous names or names not present in the gazetteer.

Comparing the coordinates of the Argentina tick dataset -as extracted from the gazetteer- with the “true coordinates“ taken by a GPS receiver, the maximum error found was 0.49°. The plot of georeferencing errors is given in Figure 8.

Assuming a Gaussian distribution of the gazetteer-retrieved coordinates around the “true” coordinates (GPS), the standard deviation is 0.148° (~18 km).

DISCUSSION.

Data collected/TBZ pathogens

Almost every pathogen group was reported in our bibliographic review, both in vertebrate and invertebrate hosts. However, only those pathogens that are able to cause serious and deadly human diseases in developing/emerging countries in the (sub-)tropics have been thoroughly studied and the relevant data published in the scientific literature. Amongst those zoonoses, the most important reported in the (sub-)tropics were SFG rickettsiosis, ehrlichiosis and CCHF. Other TBZs, such as tick-borne relapsing fever, characterized by less severe clinical signs and/or clinically indistinguishable from other endemic human diseases (i.e. malaria), have not been adequately studied, nor received enough attention from the relevant health institutions and researchers. Few records derived from epidemiological surveys and sporadic imported cases have been reported.

Zoonotic babesiosis has recently gained increased attention, chiefly due to the interest in TBZ generated by the emergence of Lyme borreliosis and Tick-borne Encephalitis, and to increased awareness of diagnostic and treatment difficulties associated with co-infection cases. Most cases have been caused by *Babesia divergens* in splenectomised patients, and although rare, this disease is very dangerous. Most human babesiosis caused by *B. microti* have occurred in the temperate belt (Kjemtrup and Conrad, 2000), and can affect –with mild chronic course- spleen-intact as well as asplenic patients, but dangerous acute infections can occur in immuno-compromised patients. Few cases of zoonotic babesiosis have been reported from (sub-)tropical countries. We collected 9 records reporting 15 cases and 163 positivities from 7622 individuals surveyed.

Q fever has been described in almost every country, with the exception of New Zealand (Maurin and Raoult 1999). As regards *Coxiella burnetii*, tick-bite transmission of the pathogen still remains episodic and not yet rigorously demonstrated, while the major cause of human infection is by inhalation of infected aerosols and skin/mucosal contact (Cutler 2007). However the few cases of Q fever we reported from (sub-)tropical countries are almost certainly attributable to tick bites.

The richest and most complete data-set obtained was that on SFG rickettsiosis: information gathered included infection in humans, animals and ticks. Clinical cases refer predominantly to humans, nevertheless we also collected 20 clinical case records in animals. With regards to humans, we found more records from passive (clinical cases reports: 229) than from active surveillance (epidemiological surveys: 81), probably because SFG rickettsiosis is characterized by typical and evident symptoms even in endemic areas. The availability of large geographical datasets on vector ticks, like the ones being developed by another ICTTD-3 working group, may allow the development of habitat suitability models for tick vectors (i.e. Austin, 2002; Estrada-Peña, 2006; Zeman and Lynen, 2006) that can be employed to explore the ecology and epidemiology of tick-borne pathogens, such as SFG rickettsiosis in Latin America.

Finally, we gathered also clinical case reporting records (n:26) of CCHF, a dramatic and deadly disease. CCHF cases outside the 35°N 35°S belt (e.g. in Turkey) have been excluded from our data collection and analysis. However, as other researchers within the ICTTD-3 project and related networks are dealing solely with CCHF, a great deal of updated information on CCHF epidemiology is already available and published (Ergonul and Whitehouse, 2007).

Diagnosis

As regards diagnosis, the identification of pathogen species during the first decades of the past century was less certain, due to the lack of effective diagnostic methods. As pathogen detection was often limited to *genus* level, several problems were encountered in the case of old records (before the 1970s). Furthermore, along with the development of new diagnostic tools, pathogens taxonomy is being continuously updated, and genus or species names may have been changed, hence making it necessary to counter-check and compare old vs. new names.

Imported TBZ

TBZs are an emerging problem in travel medicine. According to Jensenius *et al*. (2006) 14 tick-borne diseases have been reported in international travellers worldwide. The majority of cases are caused by Lyme borreliosis agents (in North America and Eurasia) and by *Rickettsia africae* (sub-Saharan Africa and East Caribbean), causative agent of African tick bite fever (ATBF). If compared with the number of endemic cases and the distribution of pathogens, imported TBZ cases aren't so frequent, but *R. africae.* In fact, ATBF reports in indigenous population are quite scarce, while the number of reported cases in travellers/tourists has recently increased significantly, thus suggesting that the disease may be somewhat under-reported in endemic zones. Nevertheless, the number of reports are continuing to rise likely due to the development of more effective diagnostic methods and increased tourism. Therefore a suspected case of TBZ should always be considered in a febrile traveller returning from rural areas during the tick season.

Time

As we managed to find many old bibliographical sources with information about TBZ distribution and their epidemiological history, we did not impose a time limit (time frame) for the inclusion of data in the Database. However the accuracy of the oldest records (before 1970s) might have been affected by the lack of effective diagnostic methods for the identification of pathogen species.

Spatial information / geographical distribution

We obtained an extremely variable geographical distribution of data. Data from countries with higher resources or GDB such as Brazil or Republic of South Africa were the most numerous, probably due to the abundance of research institutions, laboratories, universities and hospitals specifically dealing with the TBZ of importance in these countries. On the other hand, there are wide areas, especially in the African continent, totally lacking information on any TBZ. This is the situation of many developing countries affected by famine, wars and with very limited health and research infrastructures. These aspects may cause some bias on the distribution of pathogens/diseases since the presence of the pathogen cannot be excluded in such areas, as we don't know whether it has not been recorded due to lack of surveillance or whether the pathogen/disease is truly absent (Moffett et al., 2009).

Spatial accuracy

For such applications, the assessment of the accuracy of spatial references is a key aspect to correctly correlate the pathogen position with environmental variables. A classification of the precision of the spatial attribution (georeference) is needed to include or exclude the records during analysis. We managed to carry out the assessment of the accuracy of coordinates extracted from gazetteer, and this is a fundamental aspect because it defines the maximum definition usable of the spatial variables (Hengl, 2006).

Future developments

The Database is intended as a tool to gather a great amount of information dispersed throughout various sources (peer-reviewed journals, governmental reports and unpublished sources, generally speaking "grey literature"). So far the data collected are mainly from scientific literature, accessible via web, but the collaboration by local experts from (sub-)tropical countries –mostly obtained through personal relationships- has proven essential to obtain data and information, such as bibliography stored as hard copy in local libraries, which otherwise would have not been accessible through the normal channels.

Despite such collaboration, it was sometimes difficult to access and review data published in national journals written in languages not known to the authors, so it is likely that valuable data might have been missed. As anticipated above, personal relationship is a key factor to foster the collaboration of local experts, yet another way to improve collaboration at general level is to warrant a good return of the information stored in the Database. To confirm the importance of the personal communication in such extensive collaborative data collection, Guerra (2007) -who dealt with world distribution of malaria- also faced a problem of poor feedback when he tried to obtain information by sending e-mail circulars to locate sources of data not readily accessible by traditional search strategies, as we did: we received few answers (and data) from the recipients of the e-mails/letters sent out to seek for collaboration and most data were collected by the authors themselves.

A Web-Based Database may allow redistribution of the data to all collaborators and to the general public, as attempted by Moffett et al (2009). Additionally, the use of an integrated WebGIS platform could permit visualization of the interactive TBZ distribution maps. This could improve the collaboration, assuring a real-time update and redistribution of data, also avoiding the risk of double-entering the same data. However a supervision of the data entered still remains necessary in order to assure a constant standard for data quality.

Since TBZ represent an emerging global health problem, a continued bibliographic search needs to be carried out to keep the Database updated. The long-term goal of this Database is to include more countries and ecological zones, and the success of the project will depend on the active collaboration of researchers all over the world. For instance, the (sub-)tropical regions of developed countries (i.e. Northern Australia and Southern USA) and temperate regions worldwide could also be included in the future. However, it should be taken into account that TBZ priority categorisation -as for other diseases- is locally influenced, hence the availability of TBZ data may vary according to geographical areas.

Besides, further refinements should be made to improve the Database, and the contribution of more collaborators worldwide could contribute to expand and update the Database itself. As pointed out by Moffet et al. (2009), the fact that our Database was constructed already indicates the extent to which datasharing in epidemiology and ecology has become/is becoming a common feature rather than an exception. Once the data system has reached a critical mass, more international collaboration will be expected as the value of the information that can be retrieved will be understood. Future success in extending the Database will depend on a continuation of this trend, and the authors hope that this paper could encourage further collaboration, also with the involvement of further international agencies and institutions.

Acknowledgements

Authors would like to thanks all collaborators who contributed at different level by providing information and data, making useful comments and suggestions, in particular Maria Allsopp, Ali Bouattour, Ivan Horak, Zdenek Hubalek, Katja Strasek, AbdelRahim El Hussein. Proof-reading and useful comments to the manuscript by Sara Hannam are also acknowledged here. This research was facilitated through the Integrated Consortium on Ticks and Tick-borne Diseases (ICTTD-3), financed by the International Cooperation Program of the European Union, coordination action project No. 510561, coordinated by University of Utrecht, Centre for Ticks and Tick-borne diseases (Prof. F. Jongejan). The contents of this paper are solely the responsibility of the authors, and can under no circumstance be regarded as reflecting the position of the European Union.

BIBLIOGRAPHY.

Acha PN, Szyfres B (2003) Zoonoses and communicable diseases common to man and animals. Pan American Health Organization, Washington, DC

Austin MP (2002) Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. Ecol Modell 157 (2-3):101-118

Chapman AD (2005) Principles of data quality. Report for the Global Biodiversity Information Facility. Copenaghen

Craig MH, Sharp BL, Mabaso MLH, Kleinschmidt I (2007) Developing a spatial-statistical model and map of historical malaria prevalence in botswana using a staged variable selection procedure. Int J Health Geogr 6:44

Cutler SJ, Bouzid M, Cutler RR (2007) Q fever. J Infect 54 (4):313 - 318

Dufour B, Hendrikx P (2009) Epidemiological surveillance in animal health. vol Ed.2. Centre de Coop�ration Internationale en Recherche Agronomique pour le D�veloppement (CIRAD), Paris

Ergonul O, Whitehouse CA (2007) Crimean-congo hemorrhagic fever. A global perspective. Springer Netherlands,

Estrada-Pe�a An (2006) Prediction of habitat suitability for ticks. Ann N Y Acad Sci 1078 (Century of Rickettsiology; Emerging, Reemerging Rickettsioses, Molecular Diagnostics, and Emerging Veterinary Rickettsioses):275-284

Guerra CA, Hay SI, Lucioparedes LS, Gikandi PW, Tatem AJ, Noor AM, Snow RW (2007) Assembling a global database of malaria parasite prevalence for the malaria atlas project. Malar J 6:17

Guglielmone AA, Robbins RG, Apanaskevich DA, Petney TN, Estrada-Peña A, Horak IG (2010) The argasidae, ixodidae and nuttalliellidae (acari: Ixodida) of the world: A list of valid species names. Zootaxa 2528:1-28

Hengl T (2006) Finding the right pixel size. Comput Geosci 32 (9):1283-1298

Hill L (2000) Core elements of digital gazetteers: Placenames, categories, and footprints. Research and Advanced Technology for Digital Libraries, pp 280-290

Horak IG, Camicas J-L, Keirans JE (2002) The argasidae, ixodidae and nuttalliellidae (acari: Ixodida): A world list of valid tick names. Exp Appl Acarol 28 (1-4):27-54

Hugh-Jones M, Hubbert W, Hagstad H (2000) Zoonoses: Recognition, control, and prevention. Iowa State University Press,

Jensenius M, Parola P, Raoult D (2006) Threats to international travellers posed by tick-borne diseases. Travel Med Infect Dis 4 (1):4 - 13

Jongejan F, Uilenberg G (2004) The global importance of ticks. Parasitology 129 Suppl:S3-14

Kitron U (2000) Risk maps: Transmission and burden of vector-borne diseases. Parasitology Today 16 (8):324 - 325

Kjemtrup AM, Conrad PA (2000) Human babesiosis: An emerging tick-borne disease. Int J Parasitol 30 (12-13):1323 - 1337

Kumar DVRS, Sriram K, Rao KM, Murty US (2005) Management of filariasis using prediction rules derived from data mining. Bioinformation 1 (1):8-11

Labruna MB (2009) Ecology of rickettsia in south america. Ann N Y Acad Sci 1166 (Rickettsiology and Rickettsial Diseases Fifth International Conference):156-166

Larson RR (1996) Geographic information retrieval and spatial browsing. Geographic information systems and libraries: patrons, maps, and spatial information. Clinic on Library Applications of Data Processing

Maurin M, Raoult D (1999) Q fever. Clin Microbiol Rev 12 (4):518-553

Moffett A, Strutz S, Guda N, Gonz�lez C, Ferro MC, S�nchez-Cordero Vc, Sarkar S (2009) A global public database of disease vector and reservoir distributions. PLoS Negl Trop Dis 3 (3):e378

Murty US, Kumar DVRS, Sriram K, Rao KM, Bhattacharyulu CHNV, Praveen B, Krishna AR

(2005) A web based relational database management system for filariasis control. Bioinformation 1 (1):19-20

Parola P, Davoust B, Raoult D (2005) Tick- and flea-borne rickettsial emerging zoonoses. Vet Res 36 (3):469-492

Parola P, Paddock CD, Raoult D (2005) Tick-borne rickettsioses around the world: Emerging diseases challenging old concepts. Clin Microbiol Rev 18 (4):719-756

Pearce JL, Boyce MS (2006) Modelling distribution and abundance with presence-only data. J Appl Ecol 43 (3):405-412

Tsoar A, Allouche O, Steinitz O, Rotem D, Kadmon R (2007) A comparative evaluation of presence-only methods for modelling species distribution. Diversity and Distributions 13 (4):397-405

Vestavik O (2003) Geographic information retrieval: An overview. Dept. Computer and Information Science, Norwegian University of Technology and Science, 2004.

Zeman P, Lynen G (2006) Evaluation of four modelling techniques to predict the potential distribution of ticks using indigenous cattle infestations as calibration data. Exp Appl Acarol 39 (2):163-176

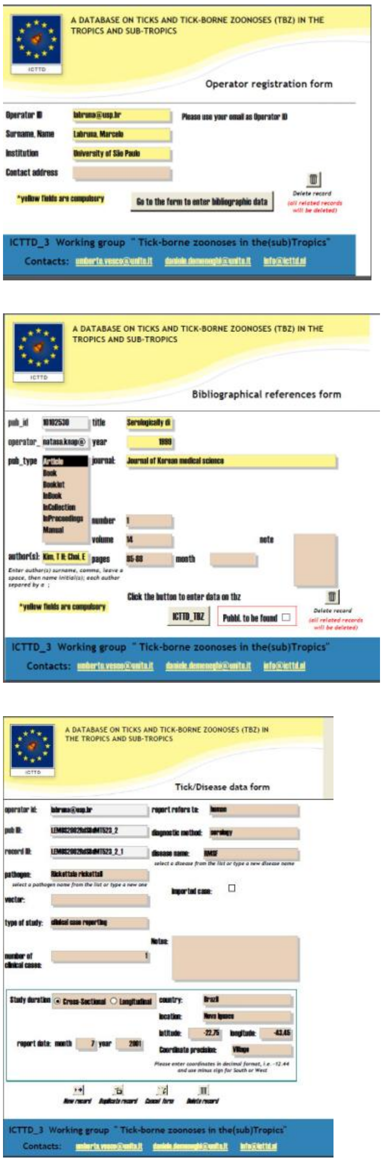

Figure 1a,b,c: Data entry forms of the Database

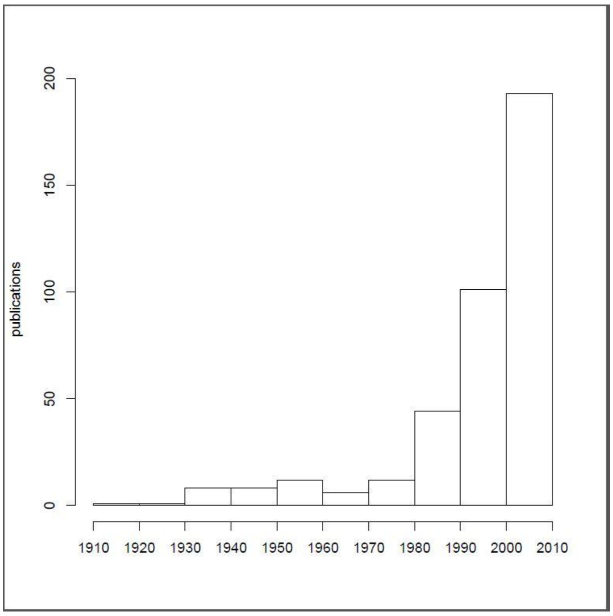

Figure 2: Number of publications per decade

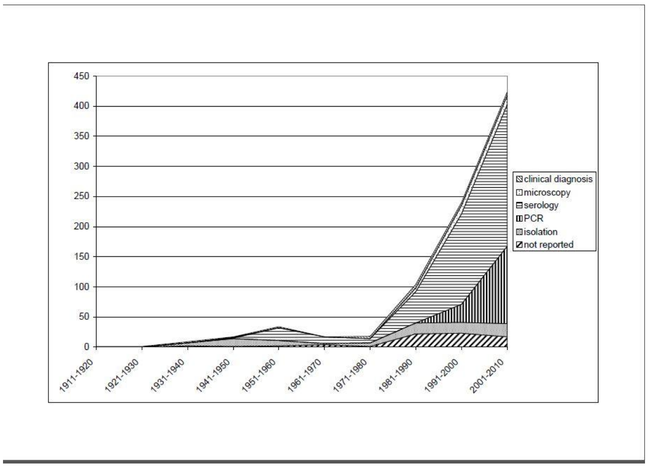

Figure 3: Diagnostic methods used along the time (number of records per decade).

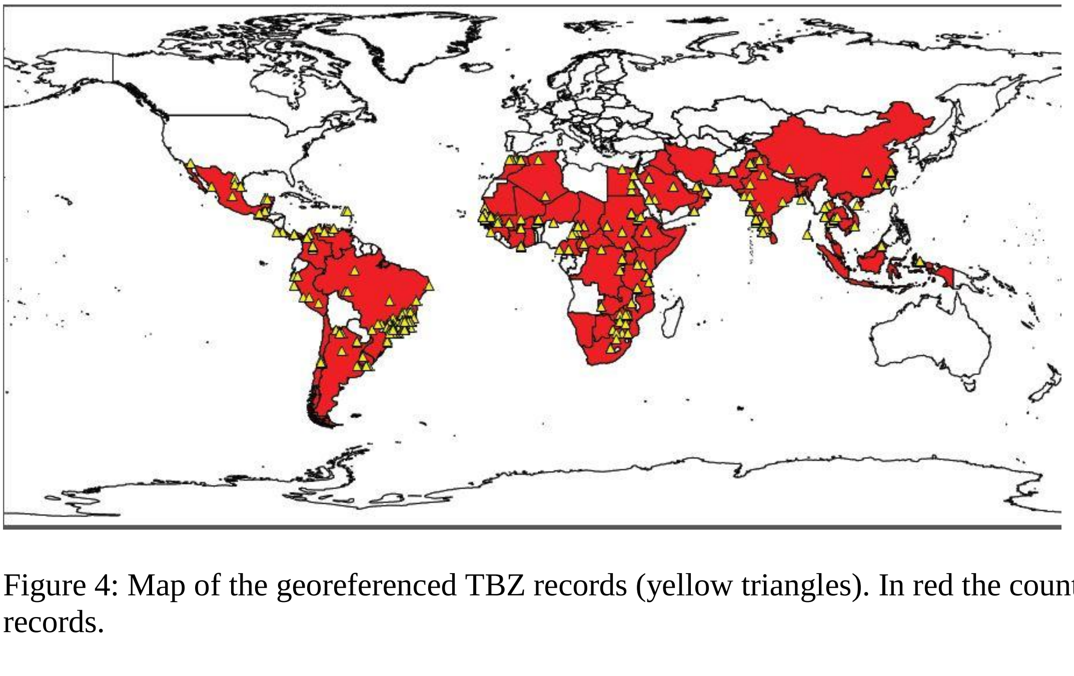

Figure 4: Map of the georeferenced TBZ records (yellow triangles). In red the countries with records.

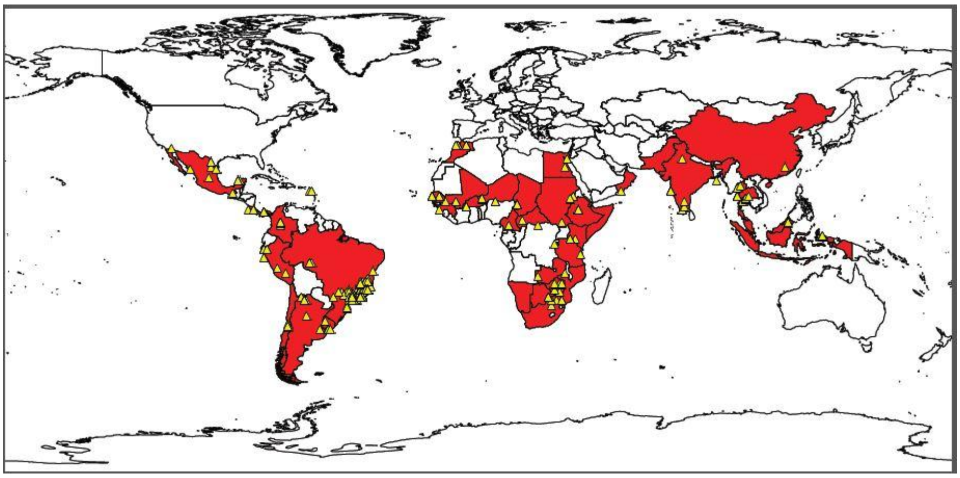

Figure 5: Map of the georeferenced SFG *Rickettsiae* records (yellow triangles). In red the countries with records.

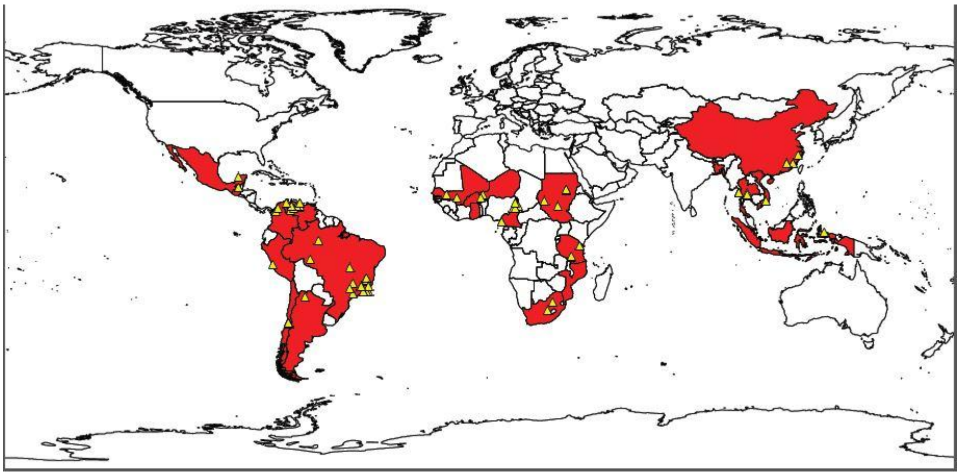

Figure 6: Map of the georeferenced human ehrlichiosis agents records (yellow triangles). In red the countries with records.

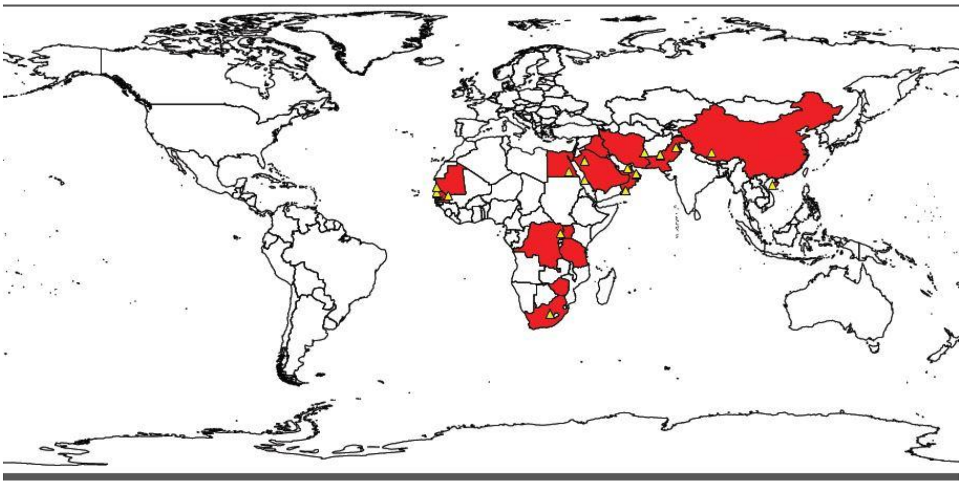

Figure 7: Map of the georeferenced CCHF virus records (yellow triangles). In red the countries with records (except Turkey).

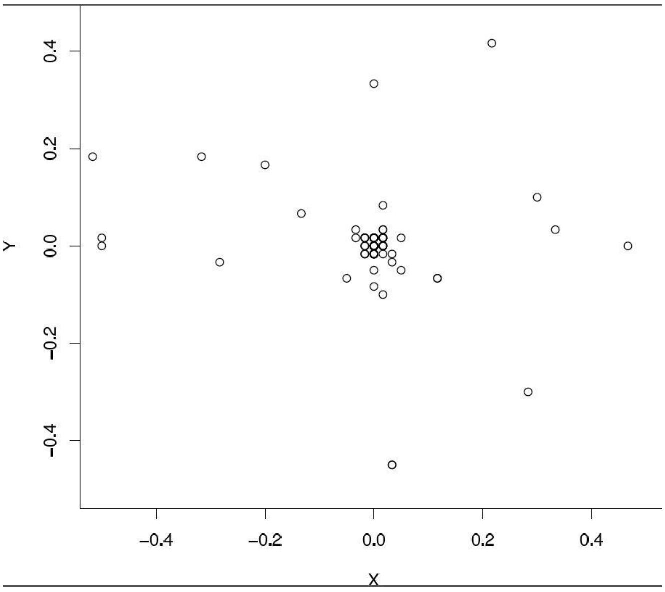

Figure 8: Plot of the error of the coordinates retrieved with the gazetteer compared with true coordinates.