ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Invertebrate Pathology 83 (2003) 185–195

Journal of INVERTEBRATE PATHOLOGY

www.elsevier.com/locate/yjipa

Special feature

Description and analysis of two internet-based databases of insect pathogens: EDWIP and VIDIL

S.M. Braxton,[a] D.W. Onstad,[b,*] D.E. Dockter,[b] R. Giordano,[c] R. Larsson,[d] and R.A. Humber[e]

[a] *8900 Milner Library, Illinois State University, Normal, IL 61790-8900, USA*
[b] *Department of Natural Resources and Environmental Sciences, University of Illinois, 1201 S. Dorner Dr., Urbana, IL 61801, USA*
[c] *Department of Biology, 321 Marsh Life Science Building, University of Vermont, Burlington, VT 05405, USA*
[d] *Department of Zoology, University of Lund, Helgonavagen 3, S-223 62 Lund, Sweden*
[e] *USDA-ARS, Tower Rd., Ithaca, NY 14853-2901, USA*

Received 7 March 2002; accepted 28 April 2003

Abstract

In 1996, two searchable databases covering insect pathogens were posted on the World Wide Web: the Ecological Database of the World's Insect Pathogens (EDWIP) and the Viral Diseases of Insects in the Literature database (VIDIL). In this paper, we describe the format and contents of EDWIP and VIDIL on the World Wide Web. EDWIP contains over 9400 pathogen–host association records, 677 negative test result or "no association" records, 4454 host species, 2285 pathogen species records, and 2057 bibliographical references. Species of Coleoptera and Lepidoptera are the best represented groups in EDWIP. Lepidopteran species account for the most associations of any host order in EDWIP, over 2500, or 27%. Of the pathogen groups, Protozoa (including microsporidia) accounted for nearly 66% of the pathogen species records and over 40% of the association records in EDWIP. Fungi account for only 18% of the pathogen species, but nearly 33% of the association records. Habitats dominated by human activities (e.g., crop, stored product, and human dwelling) account for most of the host habitats recorded in EDWIP. The United States and Japan are the most common locations and the Nearctic and Palearctic are the most common biogeographic regions reported in EDWIP. There are 4801 annotated bibliographic records in VIDIL.
© 2003 Elsevier Science (USA). All rights reserved.

Keywords: World Wide Web; Internet databases; Bioinformatics; Insect pathogens; Entomopathogens; Protozoa; Microsporidia; Bacteria; Fungi; Viruses

1. Introduction

Arthropod pathogens are potentially important biological control agents in pest management (Cook et al., 1996), and are important factors that regulate natural populations of their hosts (Lacey et al., 2001). Yet they are covered infrequently in biodiversity studies. For example, a recent search of *Biological Abstracts* identified approximately 2500 titles containing "biodiversity," "biological diversity," or "species diversity." Of these, only four records also contained references to entomopathogens, insect pathogens or arthropod pathogens. Similarly, in the *CAB Abstracts*, of approximately 5800 items designated by CAB indexers to contain information on "biodiversity," only 18 items were indexed with the term "entomopathogens" as well. Both indexing terms are from the *CAB Thesaurus*, a controlled vocabulary used by professional indexers to indicate document subject matter.

Records of arthropod pathogens and their hosts are widely dispersed in the entomological, microbiological, and related literatures. That arthropod pathogens belong to disparate taxonomic groups adds to the diffuse nature of the literature. There is no single source wherein new records of arthropod pathogens are recorded as they occur. Pathogen species descriptions and

* Corresponding author. Fax: 1-217-333-4949.
E-mail address: onstad@uiuc.edu (D.W. Onstad).

0022-2011/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0022-2011(03)00089-2

host records frequently are published in journals not regularly read by entomologists. For example, many descriptions of fungal pathogens are published in botany journals. The *Zoological Record* indexes newly described protozoans and nematodes, the *Index of Fungi* covers systematic mycology. But no comparable information resources exist for bacterial or virus systematics. In addition, literature reports of associations between insects, spiders, mites, ticks, and their pathogens frequently come from work outside systematics (e.g., field surveys, preliminary research for biological control), and these are not consistently covered in any single index.

How many species of bacteria infect insects around the world? What is the typical number of different diseases in the average insect? How host-specific are viruses? How should pathologists and entomologists design research explorations to find more pathogens and potential microbial control agents? Does a specific host species have a known pathogen belonging to certain genus, or have I just made a novel discovery? These are just a few of the questions that can be answered by gathering and organizing the voluminous information concerning relationships between pathogens and insects.

Recently, use of the Internet has increased by researchers in all fields. The Internet has become an important research tool for entomologists, especially in the form of literature based databases (Cockburn, 1998). Widespread use of the Internet for information transfer has had profound and beneficial impacts on research, primarily from access to large cooperative databases (Zenger and Walker, 2000). Online information has positively affected cooperative extension programs as well (VanDyk, 2000).

Several databases covering arthropod pathogens have been developed over the past 30 years. Martignoni et al. (1973) described the viral diseases of insects that were listed in a database he developed, and he published annual catalogs (e.g., Martignoni and Iwai, 1975), but the database itself was not generally accessible. Catalogs from the ARSEF fungal germplasm database (e.g., Humber, 1992) are now accessible as PDF files on the World Wide Web, although the database itself is not yet directly searchable. Releases of Beneficial Organisms (ROBO; Coulson, 2001), covers arthropod pathogens in addition to other beneficial organisms imported into or released in the US, and is now available on the World Wide Web. The Canadian Forest Service's Bt Toxin Specificity Database is also available (Van Frankenhuyzen and Nystrom, 1999).

In 1996, two searchable databases covering insect pathogens were posted on the World Wide Web: the Ecological Database of the World's Insect Pathogens (EDWIP) and the Viral Diseases of Insects in the Literature database (VIDIL). In this paper, we describe the format and contents of EDWIP and VIDIL on the World Wide Web. We also provide a variety of summaries of the records that should help scientists justify the importance of research and microbial-control application and development to administrators, students, colleagues, and policy makers.

2. EDWIP: Ecological Database of the Worlds Insect Pathogens (http://insectweb.inhs.uiuc.edu/Pathogens/EDWIP)

EDWIP provides information on fungi, viruses, protozoa, mollicutes, nematodes, and bacteria that are infectious in insects and closely related arthropods. The Ecological Database of the World's Insect Pathogens (EDWIP) was developed to fill the need for a centralized source of information on organisms that are pathogenic to insects and related arthropods. The focus is on information that can help the user deal with or solve ecological problems, but EDWIP fills a variety of information needs, both practical and academic. Its primary purposes are:

- to facilitate research on pathogens for biological control (including potential effectiveness against given hosts, and potential impacts on non-target hosts);
- to aid in research on the basic biology of pathogens and their hosts and patterns of relationship between them; and
- to serve as an indicator of biodiversity of arthropod pathogens.

A primary rationale behind the development of EDWIP was evaluation of pathogen–host ranges. A host range is the set of species that allow survival and reproduction of a pathogen. The ecological host range is the current set of known species with which a parasite naturally forms associations resulting in viable parasite offspring (Onstad and McManus, 1996). Physiological host range is based solely on laboratory observations of infection and propagule production. Species identified as hosts in the laboratory may not be hosts in the field (Federici and Maddox, 1996). In nature, a potential host and pathogen may not form a relationship because they do not occur together in time and/or space or because natural behaviors prevent contact. Changes in a pathogen's spatial distribution, such as emigration to a new continent or its transportation by humans to a new habitat do not guarantee a successful relationship or expansion of its host range. Host shifts are also possible at sites where the pathogen is already established (Secord and Kareiva, 1996). These scenarios represent some of the dynamics of host range evolution.

The foundation of EDWIP is the record of infection of a single host species by a single pathogen species. We call this an association. An association between a

pathogen and an insect exists when the host is infected in the field or in the laboratory by the pathogen and infectious propagules are produced. When infection has been attempted but not observed, then no association exists, and EDWIP maintains records of this occurrence as well. It is necessary to clarify that most of the virus associations in EDWIP are actually records of viral diseases and not virus species. We have separated viruses from other pathogens in EDWIP's data structure to make this distinction.

Some taxa and experiments are purposefully excluded from EDWIP. We exclude *Bacillus thuringiensis* because the data are overwhelming, and because toxicity rather than infectivity is the practical emphasis with this species. Polydnaviruses are also excluded. In addition to taxon-based exclusions, we also omit vectored pathogens that do not infect the vector, cell-culture data, and records in which pathogens are injected into hosts.

Data in EDWIP are taken from reports in the worldwide literature, including books, journals, dissertations, collection catalogs, and government publications from various sources. We have accepted contributions of published and unpublished information sent to us by colleagues and investigated any leads encountered in the literature, including current and historical work. Many of the sources cited in EDWIP are catalogs or databases in their own right. For example, many nematode associations in EDWIP were obtained from a journal article (and computerized spreadsheet) by Peters (1996). Similarly, we obtained many fungus association records from the Agricultural Research Service Collection of Entomopathogenic Fungi (Humber, 1998–1999). Sources of unpublished information in EDWIP include collection records of Joe Maddox (Illinois Natural History Survey, Champaign), and the catalog of Microsporidia maintained by Ronny Larsson (Lund University, Sweden).

Bibliographic records are part of the database, and all association records are supported by one or more citations; however, EDWIP is not intended to be an exhaustive bibliography of insect pathogen associations. When citations are entered into EDWIP, considerable value is added, far beyond what a user could obtain from a traditional abstracting and indexing service such as the *Biological Abstracts*. EDWIP brings together the scattered, individual reports of associations from the literature and captures details of the association in a searchable form, including the stages and tissues infected, host habitats, and food sources, whether the association has been observed in field populations, and localities where the association has been found. The data elements of EDWIP are listed below, along with brief descriptions of how data are compiled and entered. Detailed data entry protocols are available on the EDWIP web site.

2.1. Scientific name (to species or subspecies) and classification (order, family) of the host insect, mite, tick, or spider

We make every attempt to use current, valid scientific names, and authority lists and other sources used are cited on the EDWIP web site. Synonyms or common misspellings we have encountered are recorded in a separate field. Both database fields should be searched to ensure that records for a species have been found. The names for orders and families of the arthropods follow the scheme of Bosik (1997).

2.2. Infected life stage(s) of the host

We use egg, larva, nymph, pupa, and adult. We do not distinguish among larval or nymphal instars, nor do we distinguish between male and female adults.

2.3. Infected tissue(s) of the host

We report the tissues as named in the references cited for the association, although sometimes with less detail. Additional information may be entered in this field to indicate the progression of an infection or whether a certain tissue is infected only in hosts of a certain stage or sex. An index of infected tissues reported in EDWIP is available on the EDWIP web site.

2.4. Food of the host

This element adds a third trophic level to the database. We use scientific names of food plants and animals in most instances. Food lists may be summarized for polyphagous insects, (e.g., most deciduous trees, all Solanaceae). In general, we do not specify which part of the food the host consumes (e.g., leaves, stems, fruit, etc.). Ecological information may or may not be available in EDWIP for a given host; we enter associations as they are found, and make association data available even if ecological information on the host is incomplete or absent. Thus, one should not rely on searches of EDWIP for exhaustive lists of all pathogens infecting, for example, hosts that feed on apple. This caveat applies to other host ecological data in EDWIP.

2.5. Habitat of the host

This indicates any habitats known to be occupied by the host, not merely the habitat where the pathogen association was observed. It is possible that the pathogen does not occur in one or more of the habitats. The habitat may consist of a general descriptor (e.g., soil-dwelling, aquatic), a specific type of habitat (e.g., forest, crop), or any combination of such terms.

2.6. Number of host generations per year

May be less than one (cicadas requiring many years), one year, more than one year, or a combination if the number of generations varies across the geographic range of the host. This variable provides information about the availability of the host over time for infection and reproductive increase by the pathogen during a year.

2.7. Scientific name (to species or subspecies) and classification (high taxon, low taxon) of the pathogen

We do include fungal varieties and nematode strain data, but no other designations below subspecies are recorded. We make every attempt to use current, valid names and classifications. Pathogen synonym lists are maintained for internal use, but are not searchable at this time. The taxonomic levels used for pathogen classifications differ among the pathogen groups in EDWIP, according to the conventions of researchers who work on those groups. Our choices for high and low taxa for most groups are based on the classification schemes outlined in Tanada and Kaya (1993). Our selection is intended to provide reasonable access points for those unfamiliar with particular groups; specialists may disagree with our choices. Further details on pathogen classifications are available on the EDWIP web site. For fungi, high taxa are classes and low taxa are orders. The fungi imperfecti are entered as Deuteromycetes without known orders. Harpellales (an order of Zygomycetes) and Amoebidiales (in Trichomycetes) are used in EDWIP although not listed by Tanada and Kaya (1993). We follow Humber's (1989) classification of the genera of Entomophthorales, and use *Pandora*, *Erynia*, *Furia*, and *Zoophthora* as distinct genera. For bacteria, high taxa are Actinomycetes, spore-forming bacteria (Bacillaceae), and non-spore-forming bacteria; low taxa are families. Mollicutes are separated from bacteria in EDWIP. Mollicute high taxa are Mycoplasma and Spiroplasma; low taxa are the families Mycoplasmataceae and Spiroplasmataceae. Protozoa high taxa are phyla, and low taxa are classes, after Corliss (1994). *Helicosporidium* has been tentatively placed under Protozoa; EDWIP contains only a handful of association records for this genus. Advice from Dr. Gary Blissard (Boyce Thompson Institute, Ithaca, New York) and the standard nomenclature prepared by Murphy et al. (1995) represent the state-of-the-art virus taxonomy, and we recognize that virus names in EDWIP may not agree with some current rules.

2.8. Field or laboratory observation

An infection in a laboratory colony or in a laboratory host-range test does not equal an infection under natural conditions. If the association was observed in a specimen anywhere outside of a laboratory colony, "field" is entered (greenhouse observations are considered field observations). If any report of the association qualifies as a "field" observation, "field" is the only location used in EDWIP. The distinction between laboratory and field associations distinguishes those associations that may be considered to occur naturally.

2.9. Country (or countries) where the association was observed

This is entered as reported in supporting references. Country names are not updated to reflect current geopolitical conditions. Island names are appended to country names if reported in the literature (e.g., USA Hawaii). Our choice of locality could be criticized as arbitrary, especially given the diversity of biomes within a given country; however, exact coordinates, or more specific geopolitical designations are frequently lacking in published reports. Our use of country speeds data entry, with an admitted loss of precision. Users may find more specific locality information in the references cited for an association.

2.10. Biogeographic regions for associations observed in the field

We follow the designations of the *Zoological Record Thesaurus* wherever possible. Field observations of associations may be listed as occurring in one or more of nine regions: Australasian, Ethiopian, Nearctic, Neotropical, Oceanic Islands (Atlantic), Oceanic Islands (Indian), Oceanic Islands (Pacific), Oriental, and Palearctic. This is a recent addition to EDWIP.

2.11. Scientific name of intermediate host

An intermediate host is one that is necessary for the completion of the pathogen's life cycle. Very few (23) association records indicate an intermediate host. Absence of an intermediate host in an EDWIP association record does not necessarily mean that there is none, however.

2.12. Nematode ecological data

We record bacterial symbionts (if known) for nematodes in EDWIP, as well as soil types and habitats where the nematode association occurs.

2.13. Cited references

Although EDWIP is not a comprehensive bibliographic database like VIDIL, at least one reference is cited for each association record. We enter the citation

as a text string in a single field. The citation includes authors, year of publication, title, source title and volume (if applicable), and pagination. Association records in EDWIP have been gathered from more than 2000 references, in several languages, with publication dates ranging from 1826 to 2001. Background information for classification of hosts and pathogens and ecological information on hosts has been gathered from many sources, and although these sources are cited on the EDWIP web site, they are not included in the database itself.

As of August 2001, EDWIP contained 9407 pathogen–host association records, 677 negative test result or "no association" records, 4454 host species, 2285 pathogen species records, and 2057 bibliographical references. Data entry has continued since August 2001, and current counts are slightly higher.

Table 1 shows numbers of species and associations by host order and by pathogen group. Among hosts, Coleoptera and Lepidoptera are the best represented groups in EDWIP, with more than 1000 species (approximately one quarter of the species in EDWIP) each. Lepidoptera account for the most associations of any host order in EDWIP, over 2500, or 27%. The orders Orthoptera, Homoptera, Hymenoptera, Heteroptera, Acari, Trichoptera, and Odonata are represented by between 56 and 320 species, and 21 additional host orders are represented in EDWIP by fewer than 50 species (see Table 1 notes for complete list). Of the pathogen groups, Protozoa accounted for nearly 66% of the pathogen species records and over 40% of the association records in EDWIP. Fungi account for only 18% of the pathogen species, but nearly 33% of the association records.

Do numbers of species in EDWIP reflect species diversity in nature? Probably not, although there are some parallels, at least for hosts. According to Wilson (1992) the six most diverse groups in nature, in order of decreasing numbers of known species are Coleoptera, Lepidoptera, Hymenoptera, Diptera, Hemiptera (Heteroptera plus Homoptera), and Arachnida. In EDWIP, the ranks of host groups by number of species is similar: Coleoptera, Lepidoptera, Diptera, Hemiptera (Heteroptera plus Homoptera), Orthoptera, Hymenoptera, and Arachnida (Acari, Araneae, and Opiliones combined). The major differences are that Hymenoptera in EDWIP rank sixth rather than third, and the Orthoptera are better represented in EDWIP than their diversity in nature might predict.

Of the pathogen groups in EDWIP, Wilson (1992) estimates that fungi are the most diverse in nature, followed by protozoans, nematodes, bacteria, and viruses. EDWIP species counts for pathogen groups produce very different ranks, as indicated in Table 1. It is possible that the low number of species of fungi and bacteria in EDWIP relative to the known numbers species of fungi and bacteria from Wilson (1992) may be explained by the large numbers of species in both groups that are not pathogens. Membership in the Society of Invertebrate Pathology (Becnel, 2001) by division (i.e., by the pathogen group) also does not match the relative numbers of species or associations in EDWIP.

The representation of various groups in EDWIP could easily be an artifact of research focus. For example, the five best-represented host orders in EDWIP (Coleoptera, Lepidoptera, Diptera, Orthoptera, and Homoptera, Table 1) are all very important economically, containing both crop pests and medically important pests (e.g., mosquitoes). In addition to the pest status of hosts in EDWIP, other factors influence the representation of host and pathogen groups in EDWIP. Extensive historical reviews and culture catalogs of some

Table 1
Host and pathogen species in EDWIP, categorized by major taxonomic groups

Taxonomic group	No. of species[a] (%)	No. of associations[b] (%)
Host order		
Coleoptera	1097 (26.1)	2015 (21.4)
Lepidoptera	1068 (25.4)	2536 (27.0)
Diptera	656 (15.6)	1954 (20.8)
Orthoptera	320 (7.6)	629 (6.7)
Homoptera	269 (6.4)	826 (8.8)
Hymenoptera	190 (4.5)	358 (3.8)
Heteroptera	173 (4.1)	264 (2.8)
Acari	73 (1.7)	114 (1.2)
Trichoptera	66 (1.6)	114 (1.2)
Odonata	56 (1.3)	76 (0.8)
Other[c]	233 (5.5)	479 (5.1)
Total host species and associations	4201	9407
Pathogen group		
Protozoa	1504 (65.8)	3910 (41.6)
Fungi	411 (18.0)	3075 (32.7)
Viruses[d]	168 (7.4)	1663 (17.7)
Nematodes	146 (6.4)	475 (5.1)
Bacteria	51 (2.2)	265 (2.8)
Mollicutes	5 (0.3)	6 (<0.1)
Total no. of pathogen species and associations	2285	9407

Number and percent of species, and number and percent of associations in EDWIP are indicated.

[a] Species counts exclude records for which organism was not identified beyond the genus level.

[b] From all association records in EDWIP, including those in which host and/or pathogen were not identified beyond the genus level. For host orders, all pathogen types are included, and for pathogen groups, all host orders are included.

[c] The following host orders are represented by fewer than 50 species and are not shown: Blattodea, Siphonaptera, Thysanoptera, Ephemeroptera, Isoptera, Collembola, Opiliones, Neuroptera, Dermaptera, Thysanura, Diplura, Plecoptera, Araneae, Anoplura, Embiidina, Mallophaga, Mantodea, Mecoptera, Phasmatodea, and Psocoptera.

[d] Virus species count does not include records for viral diseases. Virus association count does include both virus species and viral disease association records.

pathogen groups are readily available, and we have used these. Thus, the representation of host and pathogen groups, and the number of associations between groups in EDWIP clearly could be influenced by selective availability of information, or by selective inclusion of available information, however unintentional.

Table 2 shows host habitats in EDWIP, along with the number of species in EDWIP recorded for each habitat. Habitats dominated by human activities (e.g., crop, stored product, and human dwelling) account for most of the host habitats recorded in EDWIP. Even forests are frequently managed. Species occupying habitats important to humans are likely to be considered pests, and pests tend to be well studied, and their life histories well documented, relative to non-pest species. Species occupying aquatic habitats are also well represented in EDWIP; this is primarily a factor of data from blood-sucking dipterans. Again, the data in EDWIP probably reflect differential research effort rather than natural phenomena. It is important to point out that only 1707 host species records in EDWIP contain habitat data.

Table 3 indicates whether associations were field collections or laboratory infections for the major insect orders and six groups represented in EDWIP. Associations are reported as field associations in EDWIP if observed in a specimen anywhere outside of a laboratory colony. The distinction between lab and field associations distinguishes those associations that may be considered to occur naturally. The data in Table 3 indicate that the pathogen associations with Homoptera and the host associations of Fungi and Nematodes are known primarily from naturally occurring associations (77.2% for Homoptera, 73.3% for Fungi, and 83% for Nematodes) as opposed to laboratory tests. Because the nature of the associations are unknown for so many associations involving other host orders and pathogen groups, little can be said about the relative proportion of natural to experimental associations, other than that the documentation of associations could be improved.

Table 2
Host habitats represented in EDWIP

Habitat	No. of host species (%)[a]
Crop/garden/orchard/vineyard/plantation/nursery	660 (38.7)
Forest	488 (28.6)
Aquatic	460 (26.9)
Grassland/savannah/prairie/meadow	133 (7.8)
Pasture/rangeland	78 (4.6)
City/urban	76 (4.5)
Soil	68 (4.0)
Stored product	47 (2.8)
Human dwelling/buildings	47 (2.8)
Greenhouse	38 (2.2)
Wetland/riparian/coastal/swamp	26 (1.5)
Animal shelter/feedlot	19 (1.1)
Lawn/turf	8 (0.5)
Animal body	5 (0.2)
Animal burrow	3 (0.1)
Tundra	2 (<0.1)
Desert	1 (<0.1)
Scrub	1 (<0.1)
Total no. of species in EDWIP with habitat data	1707
Species for which habitat data are unavailable	2747

[a] Percentages calculated based on the species for which habitat data are available in EDWIP. Note that there may be multiple habitats for one species, so these percentages should not add up to 100%.

Table 3
Natural and experimental pathogen–host associations in EDWIP, by major host order and by pathogen group, presented in order of decreasing total number of associations

	Field (%)[a]	Laboratory (%)[a]	Total[b]
Host order			
Lepidoptera	1123 (44.3)	508 (20.0)	2536
Coleoptera	869 (43.1)	132 (6.6)	2015
Diptera	1057 (54.1)	258 (13.2)	1954
Homoptera	638 (77.2)	35 (4.2)	826
Orthoptera	245 (39.0)	129 (20.5)	629
Hymenoptera	155 (43.3)	59 (16.5)	358
Heteroptera	93 (35.2)	16 (6.1)	264
Acari	66 (57.9)	16 (14.1)	114
Trichoptera	16 (14.0)	0	114
Odonata	21 (27.6)	0	76
Total			9407
Pathogen group			
Protozoa	1221 (31.2)	532 (13.6)	3910
Fungi	2253 (73.3)	265 (8.6)	3075
Viruses	543 (32.6)	239 (14.4)	1663
Nematodes	394 (83.0)	73 (15.4)	475
Bacteria	79 (29.8)	83 (31.3)	265
Mollicutes	1 (14.3)	3 (42.9)	7
Total			9407

[a] Percent of the total associations for the host order or pathogen group.
[b] Host order and pathogen group totals include records for which it is unreported whether the association occurred in the field or in the laboratory.

Table 4 shows the biogeographic regions reported in EDWIP for field observations of associations for the major pathogen groups and for the four host orders with the most field associations in EDWIP, Lepidoptera, Diptera, Coleoptera, and Homoptera. For both the total number of associations across all groups and within each group of organisms in Table 4, observations from the Palearctic region are the most common. The Nearctic region is the second most common region for field observations in EDWIP. Table 5 shows the distribution of association records according to country of observation. The US (Nearctic) and Japan (Palearctic) are the two most common locations reported in EDWIP.

The geographic distribution of observations recorded in EDWIP presents another pattern with a complex or

Table 4
Biogeographic distributions of field associations in EDWIP, by pathogen group and by host order

	Palearctic	Nearctic	Oriental	Neotropical	Australasian	Ethiopian	Oceanic[a]
Total[b]	2213	1155	308	282	279	171	38
Pathogen groups[c]							
Fungi	1034	570	241	188	194	77	23
Protozoa	595	377	48	71	35	81	5
Viruses	391	84	11	8	14	13	9
Nematodes	162	98	6	15	36	0	1
Bacteria	29	24	2	0	0	0	0
Host orders[d]							
Lepidoptera	688	254	47	37	65	22	9
Diptera	460	323	99	74	42	70	14
Coleoptera	438	204	20	57	66	40	2
Homoptera	267	125	98	58	74	17	9

Biogeographic regions follow those used in the *Zoological Record*. Associations may be reported from multiple biogeographic regions.
[a] Includes Atlantic, Indian, and Pacific Ocean Islands.
[b] Total number of field associations, by biogeographic region, regardless of organisms involved.
[c] Mollicutes are not shown as they comprise so few records.
[d] The host orders shown here are those with the most records of associations observed in the field (see Table 3).

Table 5
Countries in EDWIP from which host–pathogen associations have been reported

Country	All associations	Field observations
United States	1637	962
Japan	807	763
France	374	222
USSR[a]	345	274
Canada	256	202
New Zealand	237	168
United Kingdom	216	102
China	214	195
India	203	126
Germany	198	141
Brazil	158	128
Czechoslovakia	153	71
Australia	133	106
Israel	126	122
Denmark	113	104
Italy	94	63
Poland	91	77
Philippines	65	61
Argentina	62	48

Countries recorded in fewer than 60 records are not shown.
[a] USSR includes Russia (total of 73, 67 in field) and other parts of the former Union of Soviet Socialist Republics.

obscure origin. Data in Tables 4 and 5 indicate greater research effort (or publication effort) in certain geographic areas, bias in our literature collection, or both. The Palearctic region is by far the best represented in terms of natural field observations of host-pathogen associations, with nearly twice the number of association records of any other region (Table 4). This may reflect a high concentration of pathology expertise in Europe and Japan, intensive foreign exploration efforts by US government agencies for potential biological control agents (and the accompanying documentation, see for example Coulson 2001), size of landmass, degree of human habitation, or a combination of these factors. Poor representation in EDWIP of all but the Palearctic and Nearctic regions suggests that arthropod pathogens in most areas of the world are poorly known. Given frequent estimates of high biodiversity in some tropical ecosystems, one might expect a correspondingly high diversity of host-pathogen associations from the Neotropical and Ethiopian regions, although EDWIP's data does not conform to this expectation. The most probable explanation for this is a weakness in either reporting of associations or in our gathering of literature from these regions, or both. The geographic distribution of SIP membership over the last several years supports the idea of a strong tradition of invertebrate pathology in North America and western Europe (Becnel, 2001), and this could very well account for the patterns of geographic coverage observed in EDWIP.

There are 677 records in EDWIP of negative test results from laboratory pathogen bioassays. Of these results, 228 records are for viruses, 193 are for fungi, and 114 are for protozoans. Species of Lepidoptera account for 421 (62%) of the negative test result records, whereas members of the Diptera and Coleoptera account for about 12% each, and Hymenoptera account for 6.3%. Other host orders representing $\leqslant 2\%$ of the negative test records each include Acari, Araneae, Blattodea, Heteroptera, Homoptera, Neuroptera, Odonata, Orthoptera, Siphonaptera, and Thysanoptera.

In both associations and negative test results, pathogens had higher maximum records per species than did hosts (425 associations recorded for the fungal pathogen *Beauveria bassiana* versus 66 records for the lepidopteran pest *Lymantria dispar*, 55 negative tests recorded for the fungal pathogen *Entomophaga maimaiga* versus 16 for *L. dispar*). Frequencies of both host

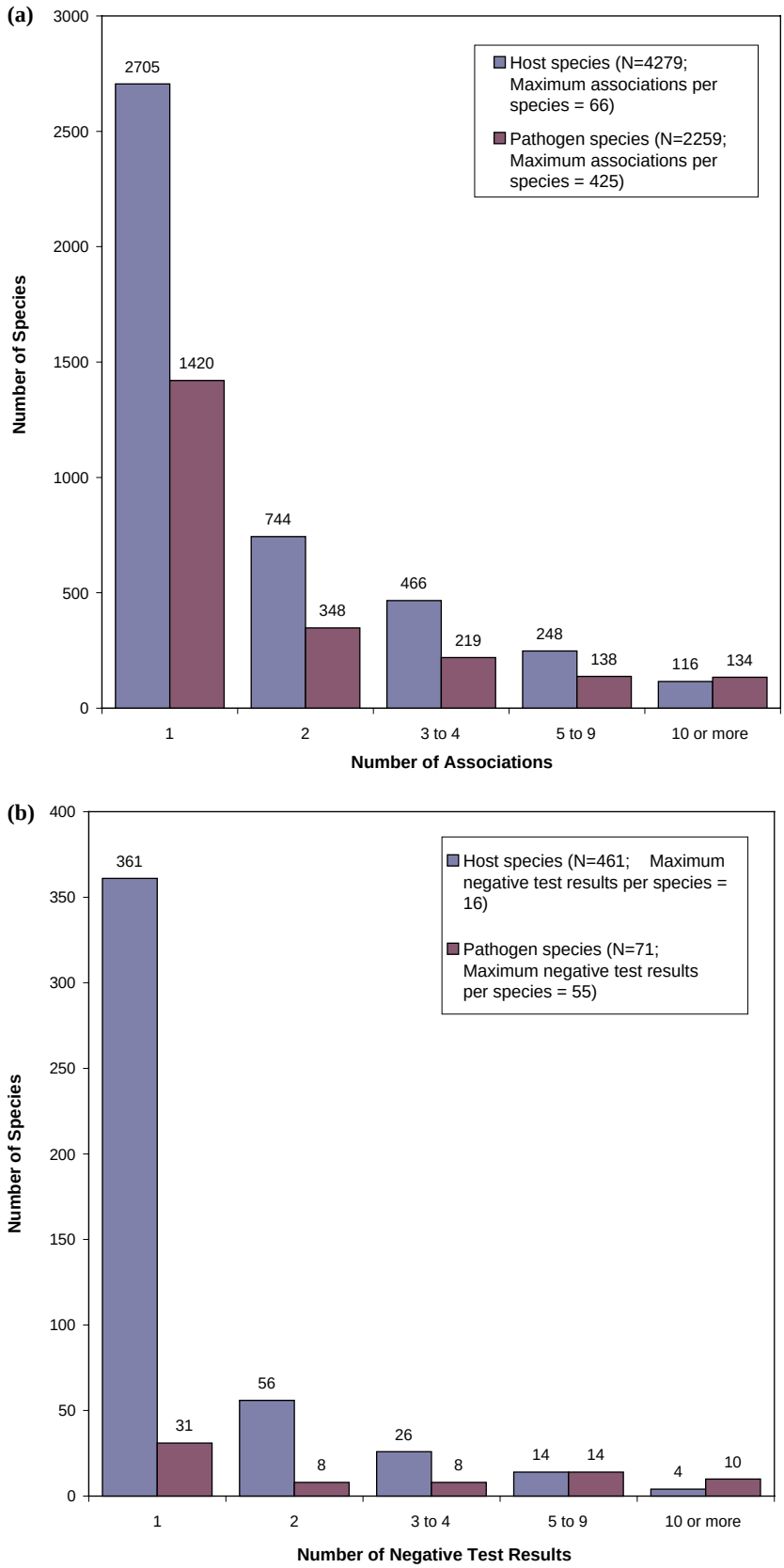

Fig. 1. Frequency of host and pathogen species by number of: (a) associations and (b) negative test results per species in EDWIP.

and pathogen species in EDWIP decrease with increasing numbers of associations per species recorded in EDWIP (Fig. 1a). More than 60% of both host and pathogen species in EDWIP associations are represented by only one association record. Approximately 80% of the host species in EDWIP negative test results are represented by only one negative test result record (Fig. 1b). In contrast, the number of pathogen species changes less dramatically with respect to number of negative test results per pathogen species (Fig. 1b). This latter pattern is consistent with intensive bioassays of a single pathogen against a broad array of hosts either to evaluate its potential as a biocontrol agent, or to test its effect on non-target organisms prior to use in biocontrol. Again we see that patterns in EDWIP data are influenced by research approaches.

3. VIDIL: Viral Diseases of Insects in the Literature (http://insectweb.inhs.uiuc.edu/Pathogens/VIDIL/)

VIDIL is an annotated, searchable bibliographic database with insect species, disease, and subject as data. The database was begun by Mauro Martignoni and published as print catalogs (Martignoni and Iwai, 1981, 1986). Much of the literature that Martignoni collected during development of his database was deposited in a special collection at the Centennial Library of the University of New Mexico in Albuquerque. The 733 references contained in the review articles by Hughes (1957) and Martignoni and Langston (1960) were added by Onstad to VIDIL in 1999. The web site summarizes the database description published by Martignoni et al. (1973). The advantage of the VIDIL database over the published catalogs is that it allows a user to: (1) search by year, author, publication name, insect name, disease, and/or subject matter; (2) find references; and (3) in some cases, relate the disease to an actual virus species by reading the literature.

There are 4801 annotated bibliographic records in VIDIL. The number of records by year of publication (prior to 1985) is shown in Table 6. No literature published after 1985 has been added, and the 1984–1985 literature is incomplete in VIDIL. Table 7 shows the viral diseases and their frequency of occurrence in VIDIL. As expected, the NPV, granulosis virus, and CPV diseases dominate the records in VIDIL.

Table 6
Bibliographic records in VIDIL, by year of publication

Year	Number of records
1945 and earlier	424
1946–1955	467
1956–1965	660
1966–1975	1658
1976–1985	1592

Table 7
Number of records for viral diseases in the VIDIL database

Disease	Number of records
Nuclear polyhedrosis or nucleopolyhedrosis	2552
Granulosis	656
Cytoplasmic polyhedrosis	508
Iridescent virosis	227
Other non-occluded-virus diseases[a]	218
Spheroidosis, insect pox	151
Presumed virosis	129
Polyhedrosis, not further identified as cytoplasmic or nuclear	117
Densonucleosis	98
Flacherie, excluding Gattine	91
CO_2 sensitivity	76
Sacbrood	60
Malaya disease	43
Bee acute paralysis	32
Gattine	32
Bee chronic paralysis	30
Paralysis, other than bee acute and bee chronic paralyses	24
Hairless-black syndrome	11
Filamentous-virus disease	7
Watery disintegration	5
Crystalline-array virosis	4
Other occluded-virus diseases	4

[a] Includes virus-like particles.

Table 8 lists the subject indexing vocabulary developed by Martignoni, and the frequencies of those terms in VIDIL records. Host specificity is the most commonly applied subject term (702 records, or nearly 15%), and microbial control is the second most common subject (609 records, or nearly 13%). The prevalence of these subjects in VIDIL, which represents an exhaustive literature collection for the publication years covered, highlights the relative importance of host range and microbial control in literature reports of insect viral diseases. Other well-represented subjects in VIDIL include biochemistry (600 records, 12%), ultrastructure (540 records, 11%), cytology (486 records, 10%), and epizootiology (438 records, 9%). Subjects with at least 400 records, in order of increasing numbers, include epizootiology, cytology, ultrastructure, biochemistry, microbial control, and host specificity.

Users should be aware that both the subject vocabulary used in VIDIL and the literature indexed in VIDIL were produced prior to recent trends in genomic analysis and genetic engineering. For virus studies in which molecular identification techniques are not used, it may be difficult to determine whether cross-infection or activation of latent virus is occurring in a host range test (McKinley et al., 1981). Thus, users should be cautious in interpreting associations recorded from old laboratory studies of viruses, and this caveat applies to both VIDIL and EDWIP.

Table 8
Number of records in VIDIL for each of the subject terms in Martignoni's subject vocabulary

VIDIL subject	Number of records
Host specificity	702
Microbial control	609
Biochemistry	600
Ultrastructure	540
Cytology	486
Epizootiology	438
Procedure	355
Bioassay	349
Replication	316
Tissue culture	316
Histology	307
Serology	302
Inactivation	298
Hereditary transmission	196
Resistance	191
Survey	190
Insect virus review	180
Biophysics	170
Virulence	160
Safety evaluation	157
Bibliography	154
Primary characterization	150
Hemolymph	149
Genetics	141
Diagnosis	140
Interaction	137
Attenuated infection	124
Large-scale process	122
Induction	109
Horizontal transmission	103
Host list	95
Stressors	93
Environmental monitoring	92
History	90
Classification	89
Sanitation	82
Inhibition	68
Nomenclature	63
Attachment-penetration	55
Economics	39
Quality control	27
Immunization	21
Translation	13
Biography	11
Culture collection	5
Recombination	5
Transcription	5
Transfection	5
Plasmid	0

4. Conclusions

Incomplete though EDWIP and VIDIL may be, they are the most complete catalogs of their kind in existence, and as such are the best available source for the information they cover. EDWIP, especially, offers considerable value over what one might get from a typical indexing source, and it is free for public use, unlike many such sources. As with any collection, the weakness of each of these databases lies in the records they do not contain, and what can and cannot be inferred from what is present or absent. We have shown through our very coarse analysis that areas of the world outside the Palearctic and Nearctic region and that diseases in certain host groups (Hymenoptera, for example) are poorly known. Finer grades of analysis are possible, the results of which may indicate patterns of ecological, evolutionary, or practical interest.

There is a need for insect pathologists, ecologists, and government regulators to understand the ecology and evolutionary biology of pathogens and their hosts. Can we predict the host-specificity or host range of a pathogen? EDWIP and VIDIL may assist regulators and decision makers in evaluating potential risks of introduced pathogens to beneficial or endangered insect species, or in identifying appropriate bioassays for evaluating such risks. EDWIP indicates species that are known to be infectious to a given pest, and is, thus, a good source for candidate biological control agents. While neither database can be considered exhaustive, and both could perpetuate erroneous identifications of pathogens or hosts present in the literature, both do serve as an initial review of current knowledge and point to areas for further study. EDWIP and VIDIL are potentially valuable, time saving tools in the study and use of insect pathogens.

Ellis and Kalumbi (1999) reported that short term funding (1–5 years) was uncertain for over two-thirds of public biological databases surveyed, despite the fact that such enterprises tend to have low administrative overhead. EDWIP is no exception, and EDWIP's value will diminish if data entry and maintenance is not supported on a continuing basis. Data collection and entry in EDWIP continues, and the backlog of data to be entered is a constant companion to the EDWIP staff. A forthcoming edition of EDWIP includes a web-based form for entering data for inclusion in the database. We look forward to community participation in keeping EDWIP up to date. If adequately supported monitarily, and by submission of data, EDWIP can be expected to represent the state of knowledge of insect–pathogen associations. It can serve the function for which it was intended, and serve as a catalyst for future research. It is our hope that new research programs can be developed by using EDWIP to identify patterns in pathogen–host associations and, perhaps more importantly, to identify gaps in the knowledge of host–pathogen relationships.

5. Use policy

All are welcome to use data from EDWIP and VIDIL for non-commercial research and educational purposes. If data from the two databases are used, acknowledgement is requested, using the citation formats given here.

Onstad, D.W. EDWIP: Ecological Database of the World's Insect Pathogens. Illinois Natural History

Survey, Champaign, Illinois [day/month/year of use]. http://insectweb.inhs.uiuc.edu/Pathogens/EDWIP.

Onstad, D.W. VIDIL: Viral Diseases of Insects in the Literature. Illinois Natural History Survey, Champaign, Illinois [day/month/year of use].

Acknowledgments

Many people and organizations contributed to this project. Ellen Brewer maintains the Internet server for both databases. We thank Joe Maddox of the Illinois Natural History Survey (INHS) for his inspiration and guidance. Dan Guyot, Sara Lanka, and Dave Levin gathered and entered data to EDWEP. Gary Blissard (Boyce Thompson Institute, Ithaca, NY), Brian Federici (University of California, Riverside), Yasuhisa Kunimi (Tokyo University of Agriculture and Technology, Japan) and Joe Maddox helped add or correct pathogen names and classification. Chris Dietrich, Wally LaBerge, Kathy Zeiders, Dave Voegtlin, and Don Webb all of the INHS and Steve Passoa (USDA-APHIS, Columbus, Ohio) helped correct insect names and classification. The Society for Invertebrate Pathology, especially Presidents Bob Granados and Brian Federici as well as newsletter editor Mark Goettel, helped with philosophical support and publicity. The following people helped us put VIDIL on the World Wide Web. Jack Coulson of the USDA in Beltsville, Maryland, kept an archive of Martignoni's files since 1986, and gave the data to David Onstad. Gary Blissard of the Boyce Thompson Institute and Bruce Black of American Cyanamid converted the software to Excel and Filemaker Pro. EDWIP was made possible, in part, by the National Biological Control Institute of the United States Department of Agriculture, Animal and Plant Health Inspection Service, which provided financial support for more than two years. The text and data may not express the views of USDA-APHIS. The Illinois Council on Food and Agricultural Research Strategic Research Initiative on Information Technology also provided funds. The work of Dr. Mauro Martignoni is an inspiration and a precedent for this project.

References

Becnel, J., 2001. Membership report. Journal of Invertebrate Pathology 34 (3), 24–25.

Bosik, J.J., 1997. Common Names of Insects and Related Organisms. Entomological Society of America, Lanham, MD.

Cockburn, A.F., 1998. Insect germplasm and genome databases. American Entomologist 44 (1), 16–19.

Cook, J., Bruckart, W.L., Coulson, J.R., Goettel, M.S., Humber, R.A., Lumsden, R.D., Maddox, J.V., Mcmanus, M.L., Moore, L., Meyer, S.F., Quimby, P.C., Stack, J.P., Vaughn, J.L., 1996. Safety of microorganisms intended for pest and plant disease control—a framework for scientific evaluation. Biological Control 7 (3), 333–351.

Corliss, J.O., 1994. An interim utilitarian ("user-friendly") hierarchical classification and characterization of the protists. Acta Protozoologica 33 (1), 1–51.

Coulson, J.R., 2001. Releases of Beneficial Organisms, ROBO. Available from: <http://www.ars-grin.gov/nigrp/robo.html> (accessed 11/2001).

Ellis, L.B., Kalumbi, D., 1999. Financing a future for public biological data. Bioinformatics 15 (9), 717–722.

Federici, B.A., Maddox, J.V., 1996. Host specificity in microbe–insect interactions. BioScience 46, 410–421.

Hughes, K.M., 1957. An annotated list and bibliography of insects reported to have virus diseases. Hilgardia 26, 597–629.

Humber, R.A., 1989. Synopsis of a revised classification for the Entomophthorales (Zygomycotina). Mycotaxon 34, 441–460.

Humber, R.A., 1998–1999. USDA-ARS Collection of Entomopathogenic Fungal Cultures: ARSEF Catalogs. Cornell University, Ithaca, NY, 1999. Available from: http://www.ppru.cornell.edu/Mycology/ARSEFCultureCollection.htm 1998–2001.

Humber, R.A., 1992. Collection of Entomopathogenic Fungal Cultures: Catalog of Strains, 1992. US Department of Agriculture, Agricultural Research Service, ARS-110, 177 pp.

Lacey, L.A., Frutos, R., Kaya, H.K., Vail, P., 2001. Insect pathogens as biological control agents: do they have a future? Biological Control 21 (3), 230–248.

Martignoni, M.E., Iwai, P.J., 1975. A catalogue of viral diseases of insects and mites. USDA Forest Service General Technical Report PNW 40.

Martignoni, M.E., Iwai, P.J., 1981. A catalogue of viral diseases of insects, mites and ticks. In: Burges, H.D. (Ed.), Microbial Control of Pests and Plant Diseases, 1970–1980. Academic Press, London, pp. 897–991.

Martignoni, M.E., Iwai, P.J., 1986. A Catalogue of Viral Diseases of Insects, Mites, and Ticks. USDA Forest Service, Pacific Northwest Research Station, Portland, OR, General Technical Report PNW-195, 4th Ed., 51 pp.

Martignoni, M.E., Langston, R.L., 1960. Supplement to an annotated list and bibliography of insects reported to have virus diseases. Hilgardia 30, 1–40.

Martignoni, M.E., Williams, P., Reineke, D.E., 1973. Computer-based catalog of viral diseases of insects: a FAMULUS application. Journal of Invertebrate Pathology 22 (1), 100–107.

McKinley, D.J., Brown, D.A., Payne, C.C., Harrap, K.A., 1981. Cross-infectivity and activation studies with four baculoviruses. Entomophaga 26, 79–90.

Murphy, F.A., Fauquet, C.M., Bishop, D.H.L., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A., Summers, M.D., 1995. Virus Taxonomy: Sixth Report of the International Committee on Taxonomy of Viruses. Springer, Wien.

Onstad, D.W., McManus, M.L., 1996. Risks of host-range expansion by insect–parasitic biocontrol agents. BioScience 46, 430–435.

Peters, A., 1996. The natural host range of *Steinernema* and *Heterorhabditis* spp. and their impact on insect populations. Biocontrol Science and Technology 6, 389–402.

Secord, D., Kareiva, P., 1996. Perils and pitfalls in the host-specificity paradigm. BioScience 46, 448–453.

Tanada, Y., Kaya, H.K., 1993. Insect Pathology. Academic Press, San Diego, CA.

VanDyk, J.K., 2000. Impact of the Internet on extension entomology. Annual Review of Entomology 45, 795–802.

Van Frankenhuyzen, K., Nystrom, C., 1999. The Bacillus toxin specificity database <http://www.glfc.forestry.ca/Bacillus/Bt_HomePage/netintro99.html> (accessed 11/2001).

Wilson, E.O., 1992. The Diversity of Life. Belknap Press of Harvard University Press, Cambridge, MA.

Zenger, J.T., Walker, T.J., 2000. Impact of the Internet on entomology teaching and research. Annual Review of Entomology 45, 747–767.