**The Ecological Database of the World’s Insect Pathogens (Metadata file)**

***Author:*** David Onstad

***2015 Data Conversion and Metadata Compilation:*** Tad Dallas, Patrick R Stephens for Research (University of Georgia) Coordination Network “The Macroecology of Infectious Disease” (DEB 1316223, Stephens PI)

***History/ Description:***

The Ecological Database of the Worlds Insect Pathogens (EDWIP) was originally compiled by David Onstad and colleagues in the late 90s (last data modifications in 1999). It was originally described in:

Braxton, S. M., Onstad, D. W., Dockter, D. E., Giordano, R., Larsson, R., & Humber, R. A. (2003). Description and analysis of two internet-based databases of insect pathogens: EDWIP and VIDIL. *Journal of invertebrate pathology*, *83*(3), 185-195.

The database was originally housed at the University of Illinois. However some time during the last decade the data became unavailable when the websites allowing access to them went offline. Further, the original data files were in a proprietary format that only users of an expensive database management program would be able to access directly. In the Spring of 2015 Tad Dallas (UGA) with assistance from Brian Perkins (UGA) converted the files to csv and SQL format. The data are currently available via remote access with the freeware SQL browsers MYSQL, SSH (SSH secure shell), and Navicat Lite. Contact Stephens for connection info.

***Contents:***

EDWIP contains more than 30000 lines of data. The primary databases it includes are:

*assocref* all associations between parasites and insects/ insect allies in EDWIP (11005 lines of data)

*hosts* ecological and taxonomic information on insect hosts (4932 lines of data)

*negative* laboratory tests of bacteria, fungi, mollicutes, protozoa, and viruses with insects and insect allies in which no infection occurred (529 lines of data)

*nematode* host associations and ecological information for nematode parasites (234 lines of data)

*nvpassoc* associations of bacteria, fungi, mollicutes, and protozoa with insects and insect allies (7164 lines of data)

*viruses* associations of viruses with insects and insect allies, as well as information on the status of some viruses as insect control agents (1659 lines of data)

***Full description of databases in EDWIP***

*assocref*: Insect host-parasite associations

**Description**

Edgelist of known associations between insect host (‘Host‘) and pathogen (‘Pathogen‘), and associated reference and indexing values.

**Format**

A dataframe with 11005 observations of 5 variables

**Variables**

• ERNnvp: EDWIP record number

• refCode: reference code

• Reference: full citation

• Host: host species

• Pathogen: pathogen species

*citation*: List of citations used to build EDWIP

**Description**

A list of citations and citation codes, as well as information on when data from various citations were entered into the database. Citations were used to populate tables of Non-Viral Pathogen Associations, Viral Pathogen Associations, Nematode Associations, and Negative Test Results files.

**Format**

A dataframe with 1966 observations of 7 variables

**Variables**

• Citation\_Code: reference code

• Reference: full citation

• Creation\_Date: date entry first created (20 seems to represent 19 in years, e.g., 2096 = 1996)

• Modification\_Date: date of final modifications (20 seems to represent 19 in years)

• ReadBy: Notes entries from “smb” and flags entries that smb made based only on abstracts (sparsly populated)

• GetIt: notes citations that were difficult to obtain (obtained from alternate libraries, ILL, ect.)

• nvpCount: unknown

*hosts* taxonomic and ecological information about host insects and insect allies.

**Description**

Data on host species taxonomy, habitat, diet preferences, and distribution (in Canada)

**Format**

A dataframe with 4392 observations of 18 variables

**Variables**

• RecordNo: Sequence from 1:nrow(hosts)

• DateEntered: Date of initial data entry

• Habitat: Brief description of the habitat of the host species. Information is entered as a single text string.

• HostSpecies: Genus species and subspecies if applicable.

• HostOrder: Host order circa 1996

• HostFamily: Host family circa 1996

• Synonyms: Other scientific name(s) that have been used for this species, including common misspellings from the insect pathology literature.

• Food: Foods on which the host feeds. Scientific or common names of host plants and animals may be used. Information is entered as a single text string.

• genYr: Number of generations of host per year. Can be <1, =1, >1, or some combination of these.

• CommonName: Host common name

• ProvinceI: Canadian provinces where host has been found.

• InsectStatus: classifies hosts as pest, wildlife, beneficial, or unknown. Roughly 50% populated.

• ModificationDate: Final modification date of entry

• InCanada: Citations for presence/absence of host in Canada. Numeric indices can be related to the citations in the ‘citations.rda‘ data file. ‘Y‘ and ‘N‘ relate to presence and absence, respectively.

• AdditionalReferences: Additional reference indices.

*negative* laboratory tests of bacteria, fungi, mollicutes, protozoa, and viruses with insects and insect allies in which no infection occurred

**Description**

Hosts were challenged with pathogen, and did not become infected. These are data on what

pathogens cannot infect certain hosts.

**Format**

A dataframe with 529 observations of 12 variables

**Variables**

• ERNntr: EDWIP record number (unique within this file)

• Pathogen: Pathogen species

• DateEntered: Date of initial data entry

• DateModified: Data of final entry modification

• LogMaxDose: log of max does in pibs / unit (units include per larvae, per insect, and unist of area). ?pibs = potentially infectious bodies? (Need to ask David about this)

• HostStageTested. Host stage exposed to pathogen, Egg, Larva, Nymph, Pupa, or Adult

• Host: Host species tested

• Group: Pathogen group (e.g. viruses, bacteria)

• HighTaxon: Highest classification of pathogen (e.g. DNA virus)

• LowTaxon: Most specific classification of pathogen (e.g. Baculoviridae)

• Order: Host taxonomic order.

• Family: Host taxonomic family.

*nemaref* associations of nematodes with insects and insect allies

**Description**

Edgelist of known associations between insect hosts and nematodes, associated reference and indexing values.

**Format**

A dataframe with 338 observations of 5 variables

**Variables**

• ERNnvp: EDWIP record number

• refCode: reference code

• Reference: full citation

• Host: host species

• Pathogen: nematode pathogen species

*nematode* ecological information on nematode species

**Description**

Host associations and ecological information for nematode parasites.

**Format**

A dataframe with 234 observations of 15 variables

**Variables**

• ERNnem: EDWIP record number

• Host: Genus and species, subspecies if applicable

• Nematode: Genus and species, subspecies if applicable

• NematodeOrder: Nematode order

• NematodeFamily: Nematode family

• NematodeStrain: Strain of nematode, if any, involved in association.

• StageInfected: Host stages in which infection observed. Egg, Larva, Nymph, Pupa, or Adult

• TissueInfected: Host tissues in which infection observed. Authors' choices of terms from citations are followed.

• FieldOrLab: Indicates whether association was observed in the field or the laboratory.

• Country: Where associations observed; country only. Some Island names may be specified to distinguish from mainland areas of a country (e.g., USA-Hawaii). Country names are entered as reported by authors in citations, and may not be geopolitically current.

• SoilType: Soil type and habitat in which association was observed.

• AssociatedBacterium: Genus, species, subspecies (optional) of bacteria identified as symbiotic with the nematode species.

• IntermediateHost: An intermediate host is a species that is necessary for the completion of the pathogen's life cycle in addition to the host named at the top of the record. If an insect or insect ally is an intermediate host, a separate record will also list it as the main host. Very few records name intermediate hosts, either because none exists or because it is not known.

• CreationDate: Date of initial data entry

• ModificationDate: Date of final entry modification.

*new\_assoc* a testfile

*newnema* duplicates *nematode*

*nemaref* negative associations with insects and insect allies

**Description**

Edgelist of known negative associations between insect hosts and pathogens, associated reference and indexing values.

**Format**

A dataframe with 569 observations of 5 variables

**Variables**

• refCode: reference code

• ERNntr: EDWIP record number

• Host: host species

• Pathogen: pathogen specie

• Reference: full citation

*Nvpassoc* non-nematode and non-viral host parasite associations

**Description**

Associations of bacteria, fungi, mollicutes, and protozoa with insects and insect allies

**Format**

A dataframe with 7164 observations of 16 variables

**Variables**

• ERNnvp: EDWIP record number (unique within this file)

• Pathogen: Genus and species, subspecies if applicable

• Group: Pathogen group (e.g. Protozoa, Fungi)

• HighTaxon: Most general taxonomic classification of pathogen (e.g. Microspora)

• LowTaxon: Most specific taxonomic classification of pathogen (e.g. Microsporea)

• Host: Genus and species, subspecies if applicable.

• Order: Host Order.

• Family. Host Family.

• HostStageTested: Host stages in which infection observed; Egg, Larva, Nymph, Pupa, or Adult

• HostTissueInfected: Host tissues in which infection observed. Authors' choices of terms from citations are followed.

• FieldOrLab: Indicates whether association was observed in the field or the laboratory. Associations observed in greenhouses are entered as "field."

• Country: Where associations observed; country only. Some Island names may be specified to distinguish from mainland areas of a country (e.g., USA-Hawaii). Country names are entered as reported by authors in citations, and may not be geopolitically current.

• IntermediateHost: An intermediate host is a species necessary for completion of the pathogen's life cycle, in addition to the primary host named in a record. If the intermediate host is an insect or insect ally, a separate record will also list it as the primary host. Very few records name intermediate hosts, either because none exists or because such relationships are not known.

• DateEntered: Date of initial data entry

• DateModified: Data of final entry modification

• BiogeographicRegion: Biogeographic region or regions ?in which host occurs? (Need to clarify with David)

*pathogen* taxonomy of non-viral and non-nematode pathogens in EDWIP

**Description**

Taxonomy of bacterial, fungal, mollicute, and protozoan parasites in EDWIP

**Format**

A data.frame with 3041 observations of 5 variables

**Variables**

• PathogenSpecies: Genus and species, subspecies if applicable

• Group: Pathogen group (e.g. Protozoa, Fungi)

• HighTaxon: Most general taxonomic classification of pathogen (e.g. Microspora)

• LowTaxon: Most specific taxonomic classification of pathogen (e.g. Microsporea)

• Comments: noted ambiguities

*viraref* associations of viruses with insects and insect allies

**Description**

Edgelist of known associations between insect hosts and viruses, associated reference and indexing values.

**Format**

A dataframe with 2124 observations of 5 variables

**Variables**

• RefCode: reference code

• Citation: full citation

• ERNv: EDWIP record number

• Host: host species

• Pathogen: viral pathogen species

*viruses* viral pathogen host associations

**Description**

Associations of viruses with insects and insect allies

**Format**

A data.frame with 1659 observations of 14 variables

**Variables**

• ERNv: EDWIP record number (unique within this file)

• Host: Genus and species, subspecies if applicable

• HighTaxon: Most general taxonomic classification of pathogen

• LowTaxon: Most specific taxonomic classification of pathogen

• Virus: May be a virus species or unidentified viral disease. This field may be empty in some records

• HostStageTested: Host stages in which infection observed; Egg, Larva, Nymph, Pupa, or Adult

• HostTissueInfected: Host tissues in which infection observed. Authors' choices of terms from citations are followed.

• FieldOrLab: Indicates whether association was observed in the field or the laboratory.

• Country: Where associations observed; country only. Some Island names may be specified to distinguish from mainland areas of a country (e.g., USA-Hawaii). Country names are entered as reported by authors in citations, and may not be geopolitically current.

• IntermediateHost: An intermediate host is a species that is necessary for the completion of the pathogen's life cycle in addition to the host named at the top of the record. If an insect or insect ally is an intermediate host, a separate record will also list it as the main host. Very few records name intermediate hosts, either because none exists or because it is not known.

• DateEntered: Date of initial data entry

• DateModified: Data of final entry modification

• ProvinceA: Canadian provinces where host-virus interaction occurs

• PathogenValue: Notes viruses that have tested for use as control agents )double check with David)