Open data on fungi and bacterial plant pathogens in New Zealand

P. R. Johnston, B. S. Weir and J. A. Cooper
Fungal and Bacterial Systematics, Landcare Research, Auckland, New Zealand

ABSTRACT
The Landcare Research collections of biological specimens and their associated databases provide an essential resource for local and international researchers, administrators of New Zealand’s plant-disease-related biosecurity system, border control agencies and non-professional citizen scientists. This article focuses on data provided for New Zealand’s non-lichenised fungi and plant pathogenic bacteria, through the NZFungi nomenclatural and bibliographic database and the associated specimen databases for the PDD fungarium and International Collection of Microorganisms from Plants culture collection. A brief history of the development of the databases is provided, along with the resources used to deliver the data to users. The data is managed through an enterprise data management system rather than an off-the-shelf product, using international standards to allow exchange of data with other providers such as GBIF, Index Fungorum, StrainInfo and Species 2000.

ARTICLE HISTORY
Received 20 November 2016
Accepted 20 December 2016

KEYWORDS
NZFungi; PDD fungarium; ICMP culture collection; data standards; citizen scientist; molecular data

Introduction
New Zealand has arguably the most complete record of fungi, chromista, plant pathogenic bacteria and plant viruses, for any country in the world, through the resources provided by the Landcare Research NZFungi databases. For fungi (excluding lichenised forms), these databases include all taxonomic names that have been used in a New Zealand context, the higher classification and synonymy for each name, the literature, specimens and GenBank records that support the use of that name in New Zealand, alongside an opinion about the current taxonomic name for each fungus, according to cited literature. The database provides an opinion about the biostatus of each fungus – whether it is present or absent from New Zealand, and when present if it is indigenous or exotic. Links between fungi, bacteria and their hosts are also provided, based on literature and specimen records. Images and descriptions are provided for many species (Table 1). Most descriptions have been extracted from the literature (e.g. protologues of fungi described from a New Zealand type specimen), but some aimed at a less scientific audience have been prepared especially for the database. The ancillary data provided for bacterial and virus names is not as extensive as that for fungi.

The NZFungi (http://nzfungi2.landcareresearch.co.nz/) web interface provides public access to the taxonomic data through searches using scientific fungal/bacterial or host names. The name search provides access to web pages with taxonomic information, including a bibliography for protologues and recombining articles, articles recording the use of that name or any of its synonyms in New Zealand, descriptions and images, and a list of the articles and specimens recording the hosts associated with that fungus, bacterium or virus.

Landcare Research also provides more detailed data on the fungal and bacterial specimens in the New Zealand Fungarium (PDD, http://www.landcareresearch.co.nz/resources/collections/pdd) and the International Collection of Microorganisms from Plants (ICMP, http://www.landcareresearch.co.nz/resources/collections/icmp) culture collection through the Systematics Collections Database (SCD, https://scd.landcareresearch.co.nz/). The SCD web pages provide collecting event data and a determination history for each specimen. Some specimens have images and descriptive notes and hyperlinks to external resources relevant to that specimen, including DOIs for literature citations, GenBank records and
UNITE Species Hypothesis DOIs for specimens that cannot be identified to species level (Figure 1).

**Mycology collections in New Zealand and development of the databases**

Professional mycology was established in New Zealand in the 1920s, with the appointment of G.H. Cunningham as New Zealand’s first resident mycologist (McKenzie 2004). Cunningham established what is now the New Zealand Fungarium (PDD) and PDD remains the only large fungarium in New Zealand. PDD is maintained by Landcare Research and is used by local and international researchers to deposit voucher specimens of importance to New Zealand. PDD has the type specimens of about 70% of the species described from a New Zealand specimen (Johnston 2006). Landcare Research also maintains ICMP. Established in 1952 by Dr Doug Dye, ICMP now holds the largest collection of living cultures of fungi and plant pathogenic bacteria in New Zealand.

Today the PDD fungarium holds almost 100,000 dried specimens, including approximately 2900 type specimens. Approximately two-thirds of the specimens are from New Zealand. PDD also holds a large collection of plant pathogen voucher specimens from the Pacific Islands (Johnston 2006). The ICMP culture collection contains about 20,000 living cultures, half fungi and half plant pathogenic bacteria. The ICMP collection includes a complete set of the type cultures of bacterial pathovars that have been described globally. The accumulation of this set of cultures was initiated by early curators Doug

---

**Table 1.** The number of names, preferred names, species present in New Zealand (number of those which are exotic species), references, taxon concepts (links between names and articles including support for synonymy and higher classification), images, descriptions and host taxa in the NZFungi database, and the numbers of PDD and ICMP specimens and links to GenBank records in the SCD database (November 2016).

| Category                                      | Number   |
|-----------------------------------------------|----------|
| Names (preferred and synonyms)                | 41,478   |
| Preferred names                               | 25,882   |
| Present in New Zealand (exotic)               | 6551 (3,771) |
| References                                    | 15,377   |
| Taxon concepts                                 | 115,306  |
| Images                                        | 44,857   |
| Descriptions                                  | 17,358   |
| Host taxa                                     | 65,000   |
| PDD fungarium specimens (type specimens)      | 99,618 (2,958) |
| ICMP cultures (type specimens)                | 20,665 (870) |
| PDD GenBank links                             | 2021     |
| ICMP GenBank links                            | 3714     |

Note that the figures include only non-lichenized fungi.

---

Figure 1. An example of some of the data available through the SCD database. A screenshot showing the images and links to GenBank records and literature DOI associated with Specimen PDD 96472, the holotype of *Clitocybe brunneocaperata*. For full details, see [https://scd.landcareresearch.co.nz/Specimen/PDD_96472](https://scd.landcareresearch.co.nz/Specimen/PDD_96472).
Dye and John Young as part of their development of the pathovar system for the nomenclature of plant pathogenic bacteria (Young et al. 1978). All cultures are stored either freeze-dried in sealed glass ampoules or under liquid nitrogen.

Electronic databasing of the PDD and ICMP specimens started in the mid-1990s. A project to deliver this data publicly over the web, along with the literature records of plant pathogens and their hosts from Pennycook (1989), was initiated in 2001. At this time, five separate flat file databases for (i) nomenclature and taxonomy, including synonymy; (ii) bibliography; (iii) host occurrences in the literature; (iv) ICMP culture collection specimens and (v) PDD fungarium specimens were integrated into a single cross-linked relational database in Microsoft Access. A “biostatus” component was added, indicating the presence/absence and exotic/introduced status of fungi in New Zealand. Coverage of records of New Zealand fungi from the literature was expanded from the plant pathogens treated by Pennycook (1989) to include all New Zealand fungi, plant pathogenic bacteria and plant viruses.

Starting in 2012, a transition was made to a new “enterprise data management” system based on SQL-Server, Visual Studio desktop applications for data management (names/literature, collections, images) and web delivery, including increasing use of Solr-based indexed retrieval (see https://en.wikipedia.org/wiki/ Apache_Solr). Although this required a large investment of time, off-the-shelf tools like Specify (http://specifysoftware.org/) or Biolomics (https://www.bioaware.com/) were unable to deliver integrated systems and services across other Landcare Research collections and data sets, which include the Allen Herbarium (CHR) for plant names and specimens, the New Zealand Arthropod Collection for arthropod names and specimens, the National Soils Database and the National Vegetation Survey.

International standards allowing the exchange of data and data checking is integral to the structure of the databases. Landcare Research staff have helped inform decisions on data standards and design through hands-on engagement in the Biodiversity Information Standards Taxonomic Databases Working Group. The role of names, and resolvable globally unique identifiers for names, was recognised as fundamental to data integration and data sharing (Patterson et al. 2010). The Index Fungorum (IF, http://www.indexfungorum.org/) database from CABI was a key tool that enabled the population of some fields and checking of critical name-name linkages and basionym/replaced name and nomenclatural status. At the same time, data linked to the Dictionary of Fungi higher classification database, itself cross-linked with IF, and CABI’s Bibliography of Systematic Mycology database provided a tool for standardising and checking references and higher classification. Periodic re-queries between IF, NZFungi and Species Fungorum (http://www.speciesfungorum.org/) improve the quality and completeness of these resources. The current database allows IF data on new names to be retrieved and cross-checked via web services, the unique key provided by the IF/Mycobank registration number linking instances of the same name across disparate data sets.

Current Landcare databases maintain taxonomic, nomenclatural and bibliographic data separately to data on specimens in the collections. The taxonomic databases provide bibliographic links to indicate sources of opinion, for example in relation to biostatus, current preferred name and parent taxon. Alternative taxonomic opinions can be maintained. As far as possible, data relating to place of publication of a taxonomic name is validated by sighting the actual publication. The specimen database includes a redetermination history. Both databases include links to external resources, such as literature DOIs, GenBank records, etc., and allow descriptions, notes and images to be linked to either names or specimens. Numbers of records are summarised in Table 1. In recent years, a special effort has been made to increase the proportion of specimens with DNA sequences available. More than 5000 GenBank records with sequences generated from PDD and ICMP specimens include hyperlinks back to the SCD database.

Funding support for the collections, databases and taxonomic research needed to ensure the data remains relevant is provided by the New Zealand Ministry of Business, Innovation and Employment through the Strategic Science Investment Fund Infrastructure Programme.

Data sharing with other providers

The NZFungi and SCD databases provide the largest resource for fungal nomenclatural data with a Southern Hemisphere focus. They provide a de
facto official list of fungi and plant pathogenic bacteria for New Zealand, and the international data standards employed allow these data to be shared with, and utilised by, other providers.

Fungal and bacterial names and associated images are harvested for the Global Biodiversity Information Facility (GBIF; http://www.gbif.org) checklists using the Integrated Publishing Toolkit and the Darwin Core Archive format. The fungal and bacterial names were provided to the Species 2000/Catalogue of Life regional hub and to the New Zealand Inventory of Biodiversity national checklist (Gordon 2012). Fungal and bacterial names are dynamically harvested by the New Zealand Organisms Register national digital species checklist (http://www.nzor.org.nz/). Data on new names is retrieved and cross-checked with IF database through web services.

Specimen data from PDD and ICMP is auto-harvested through the SCD database by the Atlas of Living Australia (http://www.ala.org.au/), the New Zealand Virtual Herbarium (http://www.virtualherbarium.org.nz/) and GBIF (http://www.gbif.org/country/NZ/summary). ICMP culture data is incorporated into the global culture collection aggregator StrainInfo (http://www.straininfo.net/).

Data delivery to users

The NZFungi2 web portal (http://nzfungi2.landcarerecords.org.nz/) provides public access to an integrated suite of data. To access the data requires a query based on a scientific name. If the searched name is not currently accepted for the fungus, then a link to the currently accepted name and its higher classification is provided, along with links to the bibliography, images, descriptions and specimens in PDD and ICMP associated with that name and all of its synonyms. Host relationships of the fungi are provided through links built into the bibliographic and specimen data. Hyperlinks are provided to the IF and Mycobank fungal names registration web pages for each fungal species name.

The NZFungi website enables host plant searching; for example entering the taxonomic name for potato (Solanum tuberosum) and selecting the associations tab will list all specimens in the ICMP and PDD collections isolated from potato, and all of the literature records of fungi and bacteria associated with potato in New Zealand (https://goo.gl/8CCaut). Figure 2 provides a screenshot showing a few of the literature and specimen records associated with potato.

![Solanum tuberosum L.](image)

Figure 2. A screenshot illustrating host plant searching on the NZFungi website. A few of the literature and specimen records of fungi associated with potato (Solanum tuberosum). For a full list of the records associated with potato, see https://goo.gl/8CCaut.
The collection-specific SCD portal (https://scd.landcareresearch.co.nz/) allows simple searches based on text from any field, or more complex searches using a field-specific syntax. Search results can be filtered by criteria such as taxonomic name, collecting locality, collection date, type status, etc. There are several levels of access and registered users can download specimen data in simple text-file (comma separated values) format for further analysis.

The non-professional, “citizen scientist” user group is catered for through two web interfaces that run off the legacy Microsoft Access database. The data delivered through these resources are now out of date but remain valuable because they allow users to access the data through simple pictorial keys, rather than scientific names. The Virtual Mycota (http://virtualmycota.landcareresearch.co.nz/) provides access to all taxa that have either descriptive or image data available. Most of the descriptions are from the scientific literature. The Fungal Guide (http://fungalguide.landcareresearch.co.nz/) treats only selected taxa, mostly common macrofungi, and provides fieldguide-like text prepared specifically for the Fungal Guide (Figure 3). Future developments will include a framework that runs off the current databases to deliver a continuously updated, electronic “Mycota” for New Zealand, incorporating existing published technical revisions as well as “popular” fungal guide content.

A summary of the Landcare Research database online resources, plus other resources discussed in this article, is listed in Table 2.

Current use of databases and websites

The specimens in the collections and the NZFungi database are utilised by researchers from Landcare Research and other government-funded Crown Research Institutes, universities, government departments, commercial companies and many international institutes. New Zealand’s Ministry of Primary Industries uses the data to make decisions on biosecurity risk of organisms detected at the border, one basic criterion being whether the organism has been detected in New Zealand previously. The Environmental Protection Authority uses the data to assess applications to import new organisms into New Zealand as part of its administration of the Hazardous Substances and New Organisms legislation (http://www.epa.govt.nz/new-organisms/). The data inform decisions on the conservation threat classification of New Zealand fungi according to IUCN (http://iucn.ekoo.se/) and New Zealand Department of Conservation criteria (http://www.doc.govt.nz/nztcs).

RSS feeds (http://nzfungi2.landcareresearch.co.nz/feed) provide regular users, such as New Zealand’s biosecurity regulators, with notices of changes to the data; for example with respect to a change in a biostatus opinion, records of new species for New Zealand and changes to synonymy or higher classification.

Over the 12 months since November 2015, 13,454 unique individuals accessed the NZFungi website. Over this period, the site was visited 30,162 times, averaging 83 times daily. Of these visits, 57% were from returning visitors and 71% for New Zealand based users. Fungal Guide web pages over the same period were used by 8443 unique individuals 82% from New Zealand, with a peak during the autumn mushroom season. From GBIF, the PDD data has been downloaded 5300 times and the ICMP data 3500 times since 2013.

The taxonomic research from the Landcare Research fungal and bacterial systematics group has the basic aim of increasing the accuracy and completeness of the NZFungi database. At the same time, this research is dependent on the resources provided through the database. For example, when new species are described on the basis of a recent specimen, often old specimens representing the same species are already present in the collections, either unidentified to the species level or misidentified, and these provide valuable additional information on distribution and ecology (e.g. Sultan et al. 2011; McKenzie et al. 2013; Cooper 2014; Johnston et al. 2017). The breadth of the coverage of plant pathogenic bacteria in ICMP was the basis of a genome scale survey of genetic diversity across Pseudomonas syringae (Thakur et al. 2016). Our data formed the basis of the threat classification for New Zealand’s fungi (Hitchmough 2002).

There is huge value in the fact that many of the users of NZFungi database are from outside New
The visibility of our data encourages researchers undertaking global-scale taxonomic projects to include taxa and specimens representing New Zealand’s diversity (e.g. Carbone & Agnello 2013; Hustad & Miller 2015; Wilson et al. 2016). For New Zealand, this provides new information on its poorly documented fungal diversity (Buchanan et al. 2004). In turn, the researchers have access to the often unique Southern Hemisphere diversity that may be important for the construction of robust phylogenies, and essential for the development of classifications that are usable globally.

Recent meta-analysis-based research projects have utilised our data to identify the factors driving movement of plant pathogens globally and relating this to biosecurity risk (Bufford et al. 2016), and improved understanding of shifts in function and network structure of the mycorrhizal communities associated with exotic compared to native host species (Dickie et al. Forthcoming).
Conclusions

More than half of the Landcare Research annual budget for fungal and bacterial systematics goes towards maintaining the specimens in the PDD and ICMP collections and the associated data. These priorities have been driven by end-user demand, especially from the government agencies responsible for managing New Zealand’s biosecurity. This demand relates both to the quality, currency and accessibility of the data and to the authenticity provided by specimen-based validation. However, the investment in supporting information infrastructure must always be balanced against the generation of new data and knowledge through systematics research. This is especially relevant for the fungi of New Zealand, indeed for all countries, where the majority of the diversity remains undiscovered and uncatalogued.

Landcare Research recognised that it needed an information infrastructure that supports scientists and data curators with data-capture and data management systems, as well as providing outward-facing web-based information delivery systems. The infrastructure developed for our data is based on open-data standards and represents an investment of several millions of New Zealand dollars over the last 15 years. This infrastructure is part of an organisation-wide strategy to maximise benefit from resources recognised by the New Zealand government as Nationally Significant Collections and Databases (NSCD, http://natsigdc.landcareresearch.co.nz/). A dedicated team of informatics specialists was established to design, develop and support the resources needed to address the broad range of NSCD maintained by Landcare Research. Such an informatics team must include a broad range of skills, including individuals with research domain-specific expertise working closely with IT architects, analysts and developers. The infrastructure is required to provide a robust platform to service external end-user operational requirements, especially for biosecurity-related operations. At the same time, the infrastructure must be flexible and adaptable to ensure internal research needs are also satisfied, needs which change according to research priorities. In addition, information technology continues to change rapidly and systems can become quickly outdated. Investment in, and strategies for supporting, these different aspects of data management within a research environment require careful and continuous assessment and balancing.

Current investment in information infrastructure is prioritised towards delivering more accessible information for a wider audience on multiple devices. The eBiota framework (eFlora http://www.nzflora.info/) is intended to provide a synthesis of taxon-related information. The framework will allow us to deliver a continuously updated online Mycota for New Zealand, incorporating current nomenclatural and taxonomic opinion, distributions, existing published technical revisions and new “popular” content. The

| Website                                                                 | Address                                                                 |
|------------------------------------------------------------------------|-------------------------------------------------------------------------|
| Landcare Research fungal and bacterial systematics research group      | http://www.landcareresearch.co.nz/science/plants-animals-fungi/fungi     |
| NZFungi, Nga harore o Aotearoa – New Zealand Fungi Database            | http://nzfungi2.landcareresearch.co.nz                                  |
| Specimen Collection Database                                           | http://scd.landcareresearch.co.nz                                       |
| PDD fungarium                                                          | http://www.landcareresearch.co.nz/resources/collections/pdd              |
| ICMP culture collection                                                | http://www.landcareresearch.co.nz/resources/collections/icmp             |
| Virtual Mycota                                                         | http://virtualmycota.landcareresearch.co.nz/                            |
| Fungal Guide                                                           | http://fungalguidelandcareresearch.co.nz/                               |
| New Zealand Arthropod Collection                                       | http://www.landcareresearch.co.nz/resources/collections/nzac            |
| Allan Herbarium (CHR)                                                  | http://www.landcareresearch.co.nz/resources/collections/allan-herbarium |
| National Vegetation Survey databank (NV5)                              | http://nvs.landcareresearch.co.nz/                                      |
| National Soils Database (NSD)                                          | https://soils.landcareresearch.co.nz/                                   |
| New Zealand Organism Register (NZOR)                                   | http://www.nzor.org.nz                                                  |
| Index Fungorum                                                         | http://indexfungorum.org                                                |
| StrainInfo                                                             | http://www.straininfo.net/                                             |
| Atlas of Living Australia (ALA)                                        | http://www.ala.org.au/                                                 |
| New Zealand Virtual Herbarium (NZVH)                                  | http://www.virtualherbarium.org.nz/                                    |
| Catalogue of Life/Species2000                                          | http://www.catalogueoflife.org/                                        |
| Global Biodiversity Information Facility (GBIF) – New Zealand resources| http://www.gbif.org/country/NZ/summary                                 |
| Naturewatch                                                            | http://naturewatch.org.nz/                                             |
| iNaturalist                                                            | http://www.inaturalist.org/                                            |
latter content will be aimed at a new generation of online citizen scientists, including users of the increasingly popular Naturewatch platform (http://naturewatch.org.nz/), a regionalised version of the iNaturalist platform (http://www.inaturalist.org/).

Another current focus is on developing effective delivery of authentic, New Zealand relevant molecular data. User demand for this is driven by data quality issues with many GenBank accessions (Nilsson et al. 2014). The size of the PDD and ICMP collections provides an ideal resource to address this issue from a New Zealand perspective.

Acknowledgements

The authors were supported through the Landcare Research Systematics Portfolio with funding from the Science and Innovation Group of the New Zealand Ministry of Business, Innovation and Employment.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors were supported through the Landcare Research Systematics Portfolio with funding from the Ministry of Business, Innovation and Employment (Science and Innovation Group), New Zealand.

References

Buchanan PK, Beever RE, Galloway DJ, Glare TR, Johnston PR, Ehc M, de Menna ME, Pennycook SR, Ridley GS, Smith JMB, et al. 2004. The fungi of New Zealand - an introduction. In: McKenzie EHC, editor. Introduction to fungi of New Zealand. Fungi of New Zealand volume 1. Hong Kong: Fungal Diversity Press; p. 1–48.

Bufford JL, Hulme PE, Sikes BA, Cooper JA, Johnston PR, Duncan RP. 2016. Taxonomic similarity, more than contact opportunity, explains novel plant-pathogen associations between native and alien taxa. New Phytol. 212:657–667.

Carbone M, Agnello C. 2013. Notes on Urnula campylospora (Pezizales). Ascomycete Org. 5:25–32.

Cooper JA. 2014. New species and combinations of some New Zealand agarics belonging to clitopilus, lyophyllum, gerhardtia, clitocybe, hydnangiun, mycena, rhodocollybia and gerronema. Mycosphere. 5:263–288.

Dickie IA, Cooper JA, Bufford JL, Hulme PE, Bates ST. Forthcoming. Loss of functional diversity and network modularity in alien plant-fungal symbioses. AoB Plants (Oxford).

Gordon DP. 2012. New Zealand inventory of biodiversity, volume 3: kingdoms bacteria, protozoa, chromista, plantae, fungi. Christchurch: Canterbury University Press; p. 1–603.

Hitchmough R. 2002. New Zealand threat classification system lists 2002 threatened species occasional publication 23. Wellington: Department of Conservation; p. 1–210.

Hustad VP, Miller AN. 2015. Studies in the genus glutinoglossum. Mycologia. 107:647–657.

Johnston PR. 2006. New Zealand’s nonlichenised fungi - where they came from, who collected them, where they are now. Natl Sci Museum Monographs. 34:37–49.

Johnston PR, Rogers JD, Park D, Marton NA. 2017. Entbalbostroma erumpensgen. et sp. nov. (Xylariaceae) from phormium in New Zealand. Mycotaxon. 131:765–771.

McKenzie EHC. 2004. The history of taxonomic mycology in New Zealand. In: McKenzie EHC, editor. Introduction to fungi of New Zealand. Fungi of New Zealand volume 1. Hong Kong: Fungal Diversity Press; p. 49–80.

McKenzie EHC, Padamsee M, Dick M. 2013. First report of rust on alnus in New Zealand is melampsoridium betulinum, not M. hiratsukanum. Plant Pathol Quar. 3:59–65.

Nilsson RH, Hyde KD, Pawlowska J, Ryberg M, Tedersoo L, Aas AB, Alias SI, Alves A, Anderson CL, Antonelli A, et al. 2014. Improving ITS sequence data for identification of plant pathogenic fungi. Fungal Divers. 67:11–19.

Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP. 2010. Names are key to the big new biology. Trends Ecol Evol. 25:686–691.

Pennycook SR. 1989. Part II. Fungal plant diseases recorded in New Zealand. Plant diseases recorded in New Zealand, Vol. 2. Auckland: Plant Diseases Division, DSIR; p. 5–502.

Sultan A, Johnston PR, Park D, Robertson AW. 2011. Two new pathogenic ascomycetes in Guignardia and rosendscheldiella on New Zealand’s pygmy mistletoes (Korthalsella: viscaeaceae). Stud Mycol. 68:237–247.

Thakur S, Weir BS, Guttman D. 2016. Phytopathogen genome announcement: draft genome sequences of 62 pseudononas syringae type and pathotype strains. Mol Plant Microbe Interact. 29:243–246.

Wilson AW, Hosaka K, Mueller GM. 2016. Evolution of ectomycorrhizas as a driver of diversification and biogeographic patterns in the model mycorrhizal mushroom genus laccaria. New Phytol. doi:10.1111/nph.14270

Young JM, Dye DW, Bradbury JF, Panagopoulos CG, Robbs CF. 1978. A proposed nomenclature and classification for plant pathogenic bacteria. NZ Journal Ag Res. 21:153–177.