

The Melbourne Water macroinvertebrate database

Christopher J Walsh,
Michael J Sammonds & Joshphar Kunapo

Find and download data from Melbourne Water's macroinvertebrate database

Select data based on location, on sampling and processing methods, and on time period. Check locations on the map, and number and types of samples selected in table below. When you are happy with the selection, download the data as an xls file (containing sites, samples, biota, taxonomy, collection methods, and processing methods tables), together with a metadata pdf explaining the structure of the downloaded data.

When using data from this site, please cite the manual, which provides more information on the database and this page:
Walsh, C.J., Sammonds, M.J., Kunapo, J. (2019) The Melbourne Water macroinvertebrate database. Melbourne Waterway Research-Practice Partnership Report 19.5. School of Ecosystem and Forest Sciences, The University of Melbourne, Melbourne.
<https://tools.thewerg.unimelb.edu.au/mwbugs/>
The manual is available here in four formats

html pdf docx rmd

Subset site selection by:

- Marine segment downstream
- Major Rivers and their tributaries
- Stream name
- Project
- Local Government Authority

Select a subset (Press delete and start typing for predictive text)

Mornington Peninsula

Select type of data (dependent on collection/processing methods)

presence-absence (all collection methods)

Select taxon resolution of the data (dependent on processing methods)

family

Period of interest

01-Jan-1992 to 01-Jan-2019

189 samples from 45 sites selected.
The table below summarizes the sample types collected at each site in the selection. Sample type abbreviations are:

Abbreviation	Method
A-B-L-G	airlift, benthos, single sample; lab-subsample to 200, ID to genus
H-B-L-G	Hess sampler, benthos; lab-subsample to 200, ID to genus
RBA-C-F-L	RBA composite (sweep-kick); 30-min field-sort, ID to lowest taxon

School of Ecosystem and Forest Sciences, The University of Melbourne,
500 Yarra Boulevard, Burnley Victoria 3121, Australia



WATERWAY ECOSYSTEM
RESEARCH GROUP



Technical Report

20.1

Melbourne Waterway Research-Practice Partnership

Technical Report 20.1

The Melbourne Water macroinvertebrate database

Published: January 2020

Citation: *(for this report and for the use of any data downloaded from <https://tools.thewerg.unimelb.edu.au/mwbugs/>)*

Walsh, C.J., Sammonds, M.J., Kunapo, J. (2020) The Melbourne Water macroinvertebrate database. Melbourne Waterway Research-Practice Partnership Report 20.1. School of Ecosystem and Forest Sciences, The University of Melbourne, Melbourne. URL: (<https://tools.thewerg.unimelb.edu.au/mwbugs/>).

Contact author: Chris Walsh (cwalsh@unimelb.edu.au)

Cover photo: The database's web interface

Table of Contents

Summary.....	4
Introduction.....	5
Structure of the database.....	8
table <i>sites</i>	8
table <i>samples</i>	9
Table <i>biota</i>	12
Revision of the taxonomy table.....	13
Protocol for providing data to be entered into the database.....	20
Preparation of sites table.....	20
Preparation of samples table.....	21
Preparation of biota table.....	23
Compilation of data file for submission.....	24
Notes on using the web interface.....	25
References.....	27

Summary

Melbourne Water's investment in biological monitoring and ecological research since 1994 has generated a large set macroinvertebrate assemblage records from streams of the Greater Melbourne region. This rich resource of biological information has been used in many ecological studies and has underpinned the management of Melbourne's streams. We have collated the data into a database with an open-access web interface (<https://tools.thewerg.unimelb.edu.au/mwbugs/>), and we aim to update it as new data are generated, so that the data can be used to further advance ecological understanding and stream management strategies.

This document describes the structure of the database and its web interface; how to use them; and quality control protocols for adding data to the database. We have also developed a data-entry app (<https://tools.thewerg.unimelb.edu.au/bugDataEntry/>) for preparation of biological data for the database.

The new database advances on previous versions by being publicly available and searchable. It includes new data and corrects errors. It uses new site and sample-coding that integrates the data with new stream and catchment data (<https://tools.thewerg.unimelb.edu.au/mwstr/>), and more effectively captures differences in collection and processing methods.

The structure of taxonomic tables has been completely revised, allowing for easier adaptation to taxonomic revisions, and for the arrival of data using identifications from genetic analyses.

Introduction

Melbourne Water (MW) have been investing in biological monitoring and ecological research using stream macroinvertebrate assemblage composition since 1994. The resulting body of data is a rich resource of information on the biodiversity, ecology, and condition of Melbourne's streams, which has been used in many studies that have advanced stream ecology and has provided a strong evidence base for decisions on the management of Melbourne's streams. The data set has strong potential to contribute to further advances and to future management strategies. By collating the data into a database with an open-access web interface, we aim to maximise that potential.

The database includes macroinvertebrate assemblage composition data collected in the Melbourne Region since 1993. Most of the data were collected as part of monitoring projects commissioned by MW. The database also includes data collected from many research projects, most of which were, at least in part, funded by MW, and some data collected by the Victorian Environmental Protection Authority (EPA Victoria) in the region.

MW's biological monitoring program began in 1994 as the Streamwatch program with the aim of "*using biological indicators (aquatic macroinvertebrates) to determine the health of streams throughout the Port Phillip and Westernport region, including determining broad-scale changes in the longer-term (5-20 years) based on infrequent surveys*" (Butcher 2003).

The program began at a high point of interest in using macroinvertebrate assemblage composition as a biological indicator of stream health (Norris & Norris 1995). It coincided with an expansion of the statewide monitoring program of the EPA, which led to biological objectives for the rivers and streams of Victoria (EPA Victoria 2004) to support the State Environment Protection Policy (Government of Victoria 2004). In 1994 and 1995, the efforts of MW's and EPA Victoria's biological monitoring programs and a research study by the Cooperative Research Centre for Freshwater Ecology (CRCFE, Walsh *et al.* 2001) led to 173 sites being sampled across the region in that year alone, including 30 sites being sampled independently by different sampling teams within weeks of each other. Since that enthusiastic start, monitoring expenditure has been lower and generally better co-ordinated.

Since then, MW's annual monitoring program has continued with little interruption, albeit with occasional changes in focus and rationale. In most years since 1994, ~50 sites have been sampled using rapid bioassessment methods as part of the core monitoring program, with a largely different selection of sites each year. In the first decade, each year's monitoring results were recorded in technical reports (Smith, Vertessy & Hardwick 1997; Hardwick & Lewin 1999; Hardwick & Waller 1999a; Papas, Nicol & Crowther 2000; Crowther, Canale & Papas 2001; Papas *et al.* 2002; Crowther, Papas & MacKay 2003; WSL 2004, 2005). From 2005, data from the monitoring program were delivered without formal reports.

In addition to the core monitoring program, monitoring studies of individual rivers or regions were conducted during the first decade: Diamond and Stony Creeks (Cameron & Vertessy 1995, 1999); Woori Yallock Creek (1997-1998, no report found); Watts River (Coleman & Pettigrove 1998a); Andersons and Jumping Creeks (Coleman & Pettigrove 1998b); Dandenong Valley streams (Pettigrove & Coleman 1999); Lang Lang River

(Coleman & Pettigrove 2001); Westernport and Mornington Peninsula streams (Hardwick 1998); Plenty River (Hardwick & Waller 1999b); Merri and Darebin Creeks (AWT 1999); Moonee Ponds Creek (2000, no report found); streams of the Werribee catchment (2005, no report found); streams of the Maribyrnong catchment (2006-2007, no report found).

From 2006 on, the monitoring program concentrated more on repeated monitoring of sites aimed at addressing specific concerns such as: continued monitoring of the near-pristine sites of the upper Yarra catchment (including two sites that had been sampled annually by EPA Victoria until that time); change in condition in sites downstream of urban growth areas in the Merri and Toomuc catchments, and of the experimental catchments of Little Stringybark and Dobsons Creek (see below); recovery of streams following bushfires; and repeated monitoring of several focus sites across the region. In 2018, a formal sampling design aiming to sample across primary environmental gradients in the region (rainfall, urban land use, and forest cover) was employed to permit a comparison of traditional sampling and processing methods with emerging genetic methods (Walsh 2018).

Melbourne Water also conducted studies addressing particular management questions, such as: the effects of resnagging in the Little Yarra River (McCasker 2002; Coleman 2006) and Eummemmering Creek (2003, no report found); supporting the Woori Yallock streamflow management plan (2013, no macroinvertebrate-specific report found); the management of environmental flows from the Merrimu Reservoir (Walsh, Bloink & Hehir 2014); and a broad-scale assessment of the capital works program from 2006 to 2015 (Water Technology, GHD & Streamline Research 2014).

In addition to all of the above monitoring data collected by MW, the database includes data collected from 1992 to 2009 by EPA Victoria from 239 sites across the region as part of their statewide monitoring program.

The database also includes data collected as part of research programs conducted in the region: the CRCFE projects 'Biological Assessment of Urban Streams' (Walsh *et al.* 2001), experimental assessment of riffle restoration in urban streams [Walsh & Breen (2001); further publication in preparation], the Yarra River ecological study (Walsh *et al.* 2007), assessment of the effects of stormwater treatment wetlands on stream ecology (Walsh 2004a); assessment of urban stormwater impacts on small stream ecosystems (Walsh 2004b, 2006). This last project led to the long-term monitoring of 11 sites on 7 streams to assess the effects of urban stormwater drainage retrofit on the stream health of Dobsons Creek, and Little Stringybark Creek and its tributaries (Walsh *et al.* 2015; with further papers in preparation). Associated with that project was a study of longitudinal variation in assemblage composition along streams of varying urban impact (White 2018), and an experiment on how urban impacts diminish the biodiversity benefit of increasing habitat complexity in streams (White & Walsh in prep).

Melbourne Water's data was first compiled into a central database in 2002 (Walsh & McCasker 2002). GHD managed the database until 2018. This report details the compilation of a revised database that:

- includes data not previously included; corrects errors; revises the site and sample-coding to integrate the data with new stream and catchment data (Kunapo, Walsh

& Sammonds 2019) and more effectively capture differences in collection and processing methods;

- revises the storage of taxonomic data in preparation for new data using identifications from genetic analyses; and
- is publicly available, including through a searchable web-interface.

The aims of this document are to:

- Describe the structure of the database, including instructions on how to use it;
- Describe quality control protocols used to enter data into the database (the database is an active resource, and will be updated as new data is generated);
- Describe the web interface for the database.

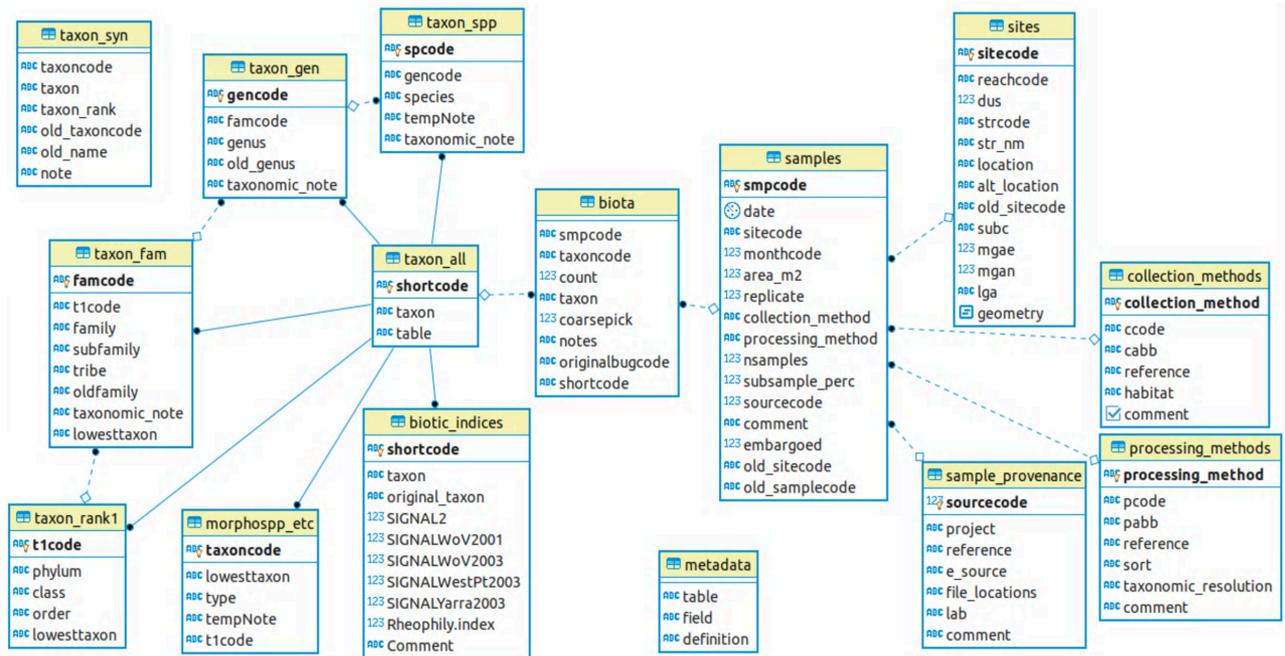


Fig. 1. Structure of tables of the Melbourne Water macroinvertebrate database and their relationships. Fields in bold indicate the field of that table, containing only unique values, which link (and must match) the field of the same name in a related table.

Structure of the database

Three tables constitute the core of the database (Fig. 1):

- The sites table lists every sampled site, identified by a unique sitecode;
- The samples table lists every sample taken at each site, identified by a unique samplecode;
- The biota table lists every taxon and its abundance identified from each sample.

Two additional tables are used to organize differences in collection methods and processing methods among samples, and a third is used to group samples by the primary project for which they were collected, with links to project reports (Fig. 1). Earlier versions of the database used a single taxonomic table with unique taxoncodes to link to the biota table. We use a different approach to taxonomy in this version of the database, as detailed below.

table sites

As of this report's publication, the sites table contained locational information on 1,154 sites across the region. It is a spatial table, with a geometry field, but the Map Grid of Australia 1994 Zone 55 coordinates of each site are also recorded in the mgae and mgan fields.

The sites table is linked to the Melbourne Water Stream Network database by the reachcode and strcode fields (streams table in the stream network database), and the str_nm field (in the stream_names table) (Kunapo *et al.* 2019). Reachcode values

comprise a three-character stream code (strcode) and a (usually) four-digit code equal to $1000 \times \log_{10}(\text{catchment area in } 100\text{-m}^2)$, so that codes beginning with 2 indicate total upstream catchment areas between 0.1 and 1 km², codes beginning with 3 indicate 1-10 km², codes beginning with 4 indicate 10-100 km², etc. For instance, the reach of the Yarra River at Whittons Reserve, Wonga Park (YAR-7350) has a catchment area of 2239 km²

Stream codes are generally intuitive abbreviations of the full stream name, and the stream network database stream_names table lists unique names for 31,397 streams, including clarifiers to ensure that ambiguously named streams have unique names (e.g. the eight Deep Creeks in the region: see Kunapo *et al.* (2019)). Names of streams can be explored at the stream network web site <https://tools.thewerg.unimelb.edu.au/mwstr/>.

Unique sitecodes in the sites table are achieved by combining reachcode with the single integer 'dus' field, which is the decile (0-9) of distance upstream from the bottom of each reach. A site at the bottom of a reach has a dus of 0, and one at the top of the reach 9. Thus throughout the region, sitecodes allow identification of the site location within <20m in most reaches (median reach length = 177 m).

The fields 'location' and 'alt_location' are, in most cases, transcribed from earlier versions of the database or from original source data. Many of the location descriptions are insufficient to be certain about reach locations, particularly upstream or downstream of roads. Where sampled reaches have been confirmed by field collectors, location descriptions have been clarified.

The fields old_sitecode and subc refer to site codes used in earlier versions of the database and subcatchments in the earlier version of the stream network subcatchment spatial layer (Grace Detailed-GIS Services 2012). The lga field identifies the local government authority area in which each site resides.

table samples

As of this report's publication, the samples table contained details of 8,191 from 992 sites.

The primary field of the samples table is smpcode: a unique code comprising hyphen-separated monthcode, sitecode, replicate number, and a two-character code indicating collection and processing methods (ccode and pcode: see below).

For instance smpcode 072-GRD-5879-0-1-BH signifies a sample collected in month 072 (Dec 1995) from Gardiners Creek (GRD) at Solway Street Ashburton, where its catchment area is 76 km² ($\text{round}(10^{5.879} \times 10^{-4}, 0)$, zero indicating that the site is at the bottom of the reach). The sample is a single sample (or possibly the first replicate taken at the site) from a riffle using a standard bioassessment kick net, lab-sorted to a 300-count subsample, identified to lowest taxon. The conventions for these calculations are explained below.

Sitecode is matched to the unique sitecodes in the sites table: see the section above for an explanation of the sitecode logic. The month code is a three-digit integer indicating the number of months since December 1989 (e.g. Jan 1990 = 01, Jan 1992 = 25). The date that the sample was taken is recorded in the date field. Monthcodes can be

calculated in R from dates and translated back into dates using two functions in the file "bugDatabaseFunctions.R".

```
# translate a date into a monthcode
calcMonthDate(lubridate::ymd("1995-12-10"))
## [1] 72
# translate a monthcode to a date
calcMonthNo(72)
## [1] "Dec 1995"
```

The replicate field is used to distinguish samples taken from the same habitat using the same collection method in the same month. For most samples there are no such replicates, and replicate = 1.

Collection method distinguishes both the collection method and the habitat sampled. Each collection method is represented in the samplecode by a single character (ccode). The collection_method field links to the collection_methods table, which details all collection methods recorded in the database (with potential for augmentation with other methods as the database grows: Table 1).

Table 1. The collection_methods table. ccode is used to identify collection methods in samplecodes. The entries in reference refer to bibtex entries in the associated reference database bugDatabase.bib.

ccode	cabb	collection_method	reference	habitat
A	RBA-E-	RBA edge (sweep)	epa_vic_2003	edge
B	RBA-R-	RBA riffle (kick)	epa_vic_2003	riffle
C	RBA-C-	RBA composite (sweep-kick)	epa_vic_2003	edge and riffle
D	RBA-C-	RBA edge-riffle combined	epa_vic_2003	edge and riffle
E	RBA-E-	RBA two edges combined	epa_vic_2003	edge
F	H-B-	Hess sampler, benthos	hess_1941	benthos
G	H-B-	micro-Hess sampler, benthos		benthos
H	A-B-	airlift, benthos, single sample	hellowell_1978	benthos
I	S-W-	snag bag, large woody debris, single sample	growns_etal_1999	large woody debris
J	A-M-	airlift, benthos, samples combined	hellowell_1978	benthos
K	S-W-	snag bag, large woody debris, samples combined	growns_etal_1999	large woody debris
L	B-B-	Boulton suction sampler, benthos, single sample	boulton_1985	benthos
M	B-B-	Boulton suction sampler, benthos, samples combined	boulton_1985	benthos
N	S-W-	sweep and jab over natural large woody debris	ghd_2013	large woody debris
O	S-W-	sweep and jab over artificial large woody debris	ghd_2013	large woody debris

Processing method distinguishes both the method used to sort the sample following collection and the taxonomic resolution to which the sample was identified. Each processing method is represented in the samplecode by a single character (pcode). The processing_method field links to the processing_methods table, which details all processing methods recorded in the database (with tentative, as yet unused fields for anticipated new data arising from DNA methods: Table 2).

For quantitative collection methods, the area sampled is recorded in the samples table field 'area_m2'. For some quantitative sampling methods, replicate samples were combined before processing. In such cases (e.g. the airlift and snag-bag samples of the Yarra Ecological study Walsh *et al.* (2007)), the number of sample units in the combined sample is recorded in the field 'nsamples', and the area_m2 field records the total area sampled by all sample units. For samples that have been subsampled, the percentage subsample is recorded in the subsample_perc table. It is important to note that the counts recorded in the biota table are raw counts, and estimates of total abundance in subsampled samples need to be calculated as: count*100/subsample_perc.

Table 2. The processing_methods table. pcode is used to identify processing methods in samplecodes. The entries in reference refer to bibtex entries in the associated reference database bugDatabase.bib.

pcode	pabb	processing_method	reference	sort	taxonomic_resolution
A	F-F	30-min field-sort, ID to family	epa_vic_2003	field	family
B	L-F	lab-subsample to 200, ID to family	walsh_1997	lab	family
C	L-F	lab-subsample to 300, ID to family	walsh_1997	lab	family
D	L-F	lab-subsample to 400, ID to family	walsh_1997	lab	family
C	L-G	lab-subsample to 200, ID to genus	walsh_1997	lab	genus
E	L-G	lab-subsample to 400, ID to genus	walsh_etal_2007	lab	genus
F	F-L	30-min field-sort, ID to lowest taxon	epa_vic_2003	field	lowest
G	L-L	lab-subsample to 200, ID to lowest taxon	walsh_1997	lab	lowest
H	L-L	lab-subsample to 300, ID to lowest taxon	walsh_1997	lab	lowest
I	L-F	lab complete sort, ID to family		lab	family
K	L-L	lab complete sort, ID to lowest taxon		lab	lowest
L		different bar-coding methods....			
M	D-L	eDNA samples?			
R	R-F	residue of subsampling to family			
S	R-L	residue of subsampling to lowest taxon			
N	N	no sample taken			
O	L-G	lab-subsample to 300, ID to genus		lab	genus
J	L-G	lab complete sort, ID to genus		lab	genus

Both the `collection_methods` and `processing_methods` tables include a more intuitive abbreviation field (`cabb` and `pabb`, respectively). These abbreviations are used by the web interface (see below) to summarize the methods used for collections of samples.

The `sourcecode` field links to the field of the same name in table 'sample_provenance'. Samples, as much as possible, have been allocated to a single project which commissioned their collection, and the `sample_provenance` table lists relevant details about that project including the name of the project, publications arising from each project (bibtex citations linked to `bugDatabase.bib`), information on the source and format of the data as supplied before it was entered into the database, and the laboratory that collected and processed each project. For some sets of samples, allocation to individual projects was not simple. For instance, samples that have been allocated to Phase III of the Little Stringybark Creek project (`sourcecode` 3), were also collected as part of MW's annual monitoring program. In such cases projects were allocated on the basis of the most likely avenue for publication using the data.

Data for some samples will not be included in the publicly available version of the data if preparation of publications using the data is in progress. Such samples will be identified in the `samples` table using the 'embargoed' field. If the sample data is being held back for publication 'embargoed' = 1.

Finally, the fields 'old_sitecode' and 'old_samplecode' are included in the `samples` table to maintain a link with old versions of the database.

Table *biota*

As of this report's publication, the `samples` table contained 186,279 records of taxa from 8,173 samples.

The `biota` table links to the `samples` table by the `smrcode` field. Every `smrcode` has a list of taxa identified by `taxoncode`, and a count. The count value is the raw count of specimens: if the sample was subsampled, the estimated abundance needs to be calculated by multiplying count by $100/\text{subsample_perc}$ (found in the `samples` table). *Data provided for the database should not be multiplied out.*

When samples are subsampled, it is standard practice to scan the whole sample for any large, rare taxa. Such specimens are identified using the 'coarsepick' field (= 1 if the specimen is the result of such a scan).

If the original supplied data provided taxon names, this is included in the 'taxon' field of the `biota` table. The originally supplied `taxoncode` is also recorded in the 'originalbugcode' field.

`Taxoncode` values are consistent with past conventions. All are eight-characters long: the first two or three characters indicating phylum, class or order; the first four characters indicating family (or in some cases sub-family or tribe), the first six characters indicating genus; and the first eight characters indicating species. Specimens not identified to species level have trailing 9s or 0s to indicate that they have not been identified beyond a certain level.

`Shortcode` is the `taxoncode` with trailing 9s or 0s removed. This field links to the `shortcode` field in `taxon_all` (see below).

Revision of the taxonomy table

In preparing the database, we revised the structure of taxonomic data from earlier versions of the database to improve ease of updating taxonomic names and relations as they are revised, and of keeping track of those revisions. The original taxonomy table (called EPA bugcodes in some database versions, and referred to as the source table below) had >8,800 rows for every possible code, including variants on partially unidentified specimens, and 15 fields with inconsistent treatment of taxonomic rank above genus-level, and inconsistent means of recording code and taxon-name revisions.

The eight new taxonomic tables (Fig. 1) take advantage of the hierarchical logic of the Victorian 8-character EPA taxoncodes developed by John Dean, that have become widely used across Australia. The database taxonomic tables can be accessed via an excel file available at https://tools.thewerg.unimelb.edu.au/data/mwbugs/mwbugs_taxonomy_tables.xlsx.

Taxon names can be derived from codes and vice-versa using R functions, discussed below. The data entry app at <https://tools.thewerg.unimelb.edu.au/bugDataEntry/> uses the tables to translate inputted taxon names into taxoncodes, and compiles a 'biota' table suitable for input into the database.

The four main taxonomic table list all Australian stream macroinvertebrate taxa at:

- Phylum, Class, Order (table `taxon_rank1`);
- Family, sub-family, tribe (table `taxon_fam`);
- Genus (table `taxon_gen`);
- Species (table `taxon_spp`);

linked in hierarchical order by an increasingly long 'shortcode' derived from the first portion of the 8-digit EPA taxoncode. Species and other taxa that have not been fully assigned to the taxonomic hierarchy are stored in a separate table (`morpho_spp`). The shortcodes from these five tables are compiled into the `taxon_all` table which lists all taxa (and the child table that they come from), and shortcodes are linked to the `biota` table. The shortcode field of `taxon_all` is also linked to the `biotic_indices` table, which lists a range of biotic indices of stream health. Finally, the unlinked `taxon_sym` table lists synonyms for taxa that have been re-named or revised. The following gives more details on the structure of each table.

1. Table `taxon_rank1` lists all unique taxa above family level (phylum, class, and order). These are variously identified by the first 1, 2 or 3 characters in the taxoncode, abbreviated as the t1code. Table 3 shows several examples of these variations. The full table contains 90 entries.

Table 3. A selection of the 90 rows in the taxon_rank1 table to illustrate the variations of the t1 code.

t1code	phylum	class	order	lowesttaxon
IA	Porifera			Porifera
IB	Cnidaria	Hydrozoa		Hydrozoa
IF2	Platyhelminthes	Turbellaria	Rhabdocoela	Rhabdocoela
K	Mollusca			Mollusca
KG	Mollusca	Gastropoda		Gastropoda
O	Arthropoda	Crustacea		Crustacea
OJ	Arthropoda	Crustacea	Copepoda	Copepoda
OJ3	Arthropoda	Crustacea	Cyclopoida	Cyclopoida
Q	Arthropoda	Insecta		Insecta
QT	Arthropoda	Insecta	Trichoptera	Trichoptera

2. taxon_fam lists all families, subfamilies, and tribes (where relevant) specified by the first four characters of the taxoncode, abbreviated as the famcode. taxon_fam is linked to taxon_rank1 by t1code. Revisions to family names are handled by recording the old family name of renamed families in the oldfamily field, and where the revision involved more complex splitting, it is explained in the taxonomic_note column (Table 4). For this level and below, Rotifera (t1code = "J") have been omitted, because the taxon codes in the source table were not as consistent as other groups, and rotifers are rarely identified in macroinvertebrate samples. The table contains 386 rows.

Table 4. A selection of the 386 rows in the taxon_fam table to illustrate the table structure, and the nature of revision notes.

t1code	famcode	family	subfamily	tribe	oldfamily	taxonomic_note	lowesttaxon
KG	KG02	Tateidae			Hydrobiidae		Tateidae
KG	KG13	Pomatiopsidae			Hydrobiidae	Coxiella moved to this family	Pomatiopsidae
QC	QC03	Sphaeriusidae			Microsporidae	Revised to combine Microsporidae and Sphaeriidae	Sphaeriusidae
QD	QDAF	Chironomidae	Orthoclaadiinae				Orthoclaadiinae
QD	QDAG	Chironomidae	Chironominae	Pseudochironomini			Pseudochironomini
QO	QO16	Corduliidae				Most genera and species transferred to new Families	Corduliidae
QT	QT25	Leptoceridae					Leptoceridae

3. `taxon_gen` lists 1,078 genera, specified by the first 6 characters of the taxoncode, abbreviated as `gencode`. The table is linked to `taxon_fam` by `famcode`. As for families, revisions are recorded by noting the old name used for a revised genus in the `old_genus` field. In its current state, the table includes superceded genus names, as noted in the `taxonomic_note` field (Table 5).

Table 5. A selection of the 1078 rows in the `taxon_gen` table to illustrate the table structure, and the nature of revision notes.

famcode	gencode	genus	old_genus	taxonomic_note
IB02	IB0201	Cordylophora		
LO05	LO0506	Pristinella		Transferred to Pristina
QD06	QD0602	Dixella	Paradixa	
QDAF	QDAF02	?Acricotopus		
QE02	QE0201	Baetis		Genus does not occur in Australia
QE06	QE0606	Tillyardophlebia	Leptophlebiidae Genus D	
QE08	QE0803	Irpacaenis	Caenidae Genus C	

4. `taxon_spp` lists 4,275 species, specified by full 8-character taxoncodes (also called bugcodes). It is linked to `taxon_gen` by `gencode`, and to `taxon_rank1` by `t1code`. This table only contains taxa with matches to entries in the `taxon_gen` and `taxon_rank1` tables. It contains some undescribed morphospecies. Morphospecies without links to `taxon_gen` are listed in the next table (`morphospp_etc`). Revisions at the species level have not been fully checked at the time of writing: the field `tempNote` contains notes on revisions taken directly from the source table (Table 6).

Table 6. A selection of the 4275 rows in the `taxon_spp` table to illustrate the table structure, and the nature of revision notes.

gencode	scode	species	tempNote	taxonomic_note
KP0104	KP010401	Velesunio ambiguus		
LH0104	LH010401	Glossiphonia australiensis		
QO1618	QO161808	Synthemis ofarrelli	[Now QO230802]	
QT0605	QT0605B1	Cheumatopsyche sp.AV10	[= QT060503 Ch.kakaduensis]	
QT2507	QT2507B1	Oecetis EPA sp.10	(= EPA sp.1, QT2507A1)	
QT2511	QT251102	Triplectides australicus		

5. morphospp_etc lists 644 lowesttaxon entries in the source table of three types:
 - a. 'voucher' taxa, that were identified as morphospecies by various analysts. These taxa have not been well curated, and likely contain substantial redundancy. Their primary value is for within-study use, where they are unlikely to be ambiguous identifications of other taxa.
 - b. 'partially unidentified taxa' such as two morphologically similar genera that are difficult to tell apart (e.g. *Physa/Physastra*);
 - c. 'valid taxa not in taxonomy hierarchy', Taxa that, in the source table, were given non-standard taxoncodes, and thus do not fit into the new schema easily. Further work is needed to decide if these can or should be incorporated into the other tables with more correct codes, but this will require ensuring that their codes have not been used in the biological data of the database first.

The 8-character taxoncodes of this table have been retained for now, to permit matching with extant data. This remains a work in progress.

6. taxon_all is simply a compilation of the shortcode and taxon fields from the preceding taxonomic tables (The name of the shortcode differs among tables to permit the hierarchical linking of the tables: t1code for taxon_rank1, famcode for taxon_fam, gencode for taxon_gen, spcode for taxon_spp, and taxoncode for morphospp_etc). taxon_all has a third field, 'table', indicating the source table for each taxon.
7. taxon_syn lists any taxonomic changes to permit tracking of deprecated taxon names.
8. biotic_indices lists all taxa (and their shortcodes) that have been used in a range of biotic indices of stream health which assign sensitivity grades to taxa. The sensitivity grades listed in this table are:
 - a. SIGNAL2 (Chessman 2003);
 - b. SIGNALWoV2001 and SIGNALWoV2003, variants on the SIGNAL scores used in the State Environment Protection Authority (Waters of Victoria) (SEPP Wov; EPA Victoria 2004);
 - c. SIGNALWestPt2003 SIGNAL score for streams of the Westernport catchment variation to the SEPP WoV (Government of Victoria 2001);
 - d. SIGNALYarra2003 SIGNAL score for streams of the Yarra catchment variation to the SEPP WoV (Victoria 1999);
 - e. Rheophily.index, an index of sensitivity to flow stress (Bond, Thomson & Reich 2012).

As noted above, none of the above tables include the trailing 9s or 0s that are typically used in specifying (say) families. Instead we have written functions that read such codes and correctly convert them to their names (and taxonomic hierarchy), in the file

“bugDatabaseFunctions.R”

(<https://tools.thewerg.unimelb.edu.au/data/mwbugs/bugDatabaseFunctions.R>).

The data entry app (<https://tools.unimelb.edu.au/bugDataEntry/>) aids compilation of sample data by taking inputted taxon names, converting them to the correct taxoncodes and compiling them into a table consistent with the biota table.

A large proportion, but not all, of the database follows a useful convention that is not widely followed. Taxoncodes that end in trailing 00s indicate specimens that were identified to the specified supra-specific level (genus, family, etc.) because that was the taxonomic level for that taxon used in the sample or study. Taxoncodes that end in trailing 99s indicate specimens that were damaged or immature and were identified to a level higher than was used for other specimens of that taxon in the sample or study. A convention that is retained for past data is coleopteran specimens ending in “9I” are larvae, and dipteran specimens ending in “9I” are pupae. We propose that all future data follow these conventions (with the last convention being optional).

Thus, a vector of taxoncodes can be imported into R, and converted to their taxonomic hierarchy using the function “codeTaxonomy()”. As an example, the taxoncodes of specimens collected in the 2018-2019 Melbourne Water biomonitoring program as received can be converted to taxon names, producing output as in Table 7.

Furthermore, the naming conventions in the supplied data can be checked for consistency with the database. In the supplied 2018-2019 MW biomonitoring data there were 21 mismatches between the names with the supplied data and the database data. The example output in Table 8 shows that the mismatches were a combination of taxonomic revisions not being up-to-date in the supplied data (e.g. Sphaeriusidae, Platyncemididae, Cordylophoridae), of non-standard terminology (e.g. Talitridae sp.(Unidentified)), and of misspellings (e.g. Neurorthidae, rather than Nevorthidae), allowing useful checks of data correctness.

Table 7. Example output from the codeTaxonomy() function, and code for running it.

```
testBiota <- as.data.frame(readxl::read_excel("MMMacroProject2018-2019_FinalBugData.xlsx", sheet = 5))
#Note that fields for taxon code and taxon name need to be "taxoncode" and "Lowesttaxon", consistent with database
testBiota <- unique(testBiota[c("bugcode", "lowesttaxon")])
names(testBiota) <- c("taxoncode", "lowesttaxon")
bugnames <- codeTaxonomy(testBiota$taxoncode)
#show first 6 taxa
flextable::regulartable(head(bugnames))
```

taxoncode	phylum	class	order	family	subfamily	tribe	species	lowesttaxon
MM999999	Arthropoda	Arachnida	Acarina					Acarina (Unident.)
KG069999	Mollusca	Gastropoda		Planorbidae	Ancylinae			Ancylinae (Unident.)
QO999998	Arthropoda	Insecta	Odonata					Epiproctophora (Unident.)
QDAA9999	Arthropoda	Insecta	Diptera	Chironomidae	Aphroteniinae			Aphroteniinae (Unident.)
QT239999	Arthropoda	Insecta	Trichoptera	Atriplectididae				Atriplectididae (Unident.)
OT019999	Arthropoda	Crustacea	Caridea	Atyidae				Atyidae (Unident.)

Table 8. Example code for running the function checkTaxonNames(), and its output, showing mismatches between supplied taxon names and database names.

```
mismatches <- checkTaxonNames(testBiota)
#show first 6 taxa
flextable::regulartable(tail(mismatches))
```

taxoncode	phylum	class	order	family	subfamily	tribe	lowesttaxon	suppliedNames
OP019999	Arthropoda	Crustacea	Amphipoda	Talitridae			Talitridae (Unident.)	Talitridae sp.(Unident.)
QO049999	Arthropoda	Insecta	Odonata	Platycnemididae			Platycnemididae (Unident.)	Protoneuridae (Unident.)
QN049999	Arthropoda	Insecta	Neuroptera	Nevrorthidae			Nevrorthidae (Unident.)	Neurorthidae (Unident.)
QD099911	Arthropoda	Insecta	Diptera	Ceratopogonidae			NO MATCH	Ceratopogoninae (Unident.)
OR259999	Arthropoda	Crustacea	Isopoda	Oniscidae			Oniscidae (Unident.)	Oniscidae sp.(Unident.)
IB029999	Cnidaria	Hydrozoa		Cordylophoridae			Cordylophoridae (Unident.)	Clavidae (Unident.)

Protocol for providing data to be entered into the database

The database is an active resource, and will be updated with relevant data as it becomes available. If you have macroinvertebrate assemblage data that could be suitable for inclusion in the database, please contact Chris Walsh at cwalsh@unimelb.edu.au.

Data for the database should be prepared in a format consistent with the database structure (Fig. 1). If your standard approach to data preparation and curation is to use a spreadsheet program such as Microsoft Excel, a useful approach would be to download some data from the database web interface (see next section), and use the downloaded xlsx file as a template for data preparation. The data entry app (<https://tools.unimelb.edu.au/bugDataEntry/>) may also prove useful.

Any data supplied should be accompanied (in a cover letter) by information about the project that led to the samples being collected. Did the project have a name? Who funded it? Are there any publications that report on or use the data? (If so please provide citation data, and preferably a copy of the publications.) Who collected the data (names and organization)? Who processed the data (names and organization)? What quality assurance protocols were used in the data collection and processing? This information will be used to populate the `samples_provenance` table and to assign a sourcecode to the samples.

Preparation of sites table

The sites table should preferably be a separate spatial table: e.g. an ESRI shapefile, a MapInfo file, or an sf object saved as an Rdata file in R. Alternatively, it can be supplied as a standard table in excel, as long as the `mgae` and `mgan` fields are populated and correct. The following is a process for preparing the sites data using R and its `sf` package. Correct alignment of the sites with the stream network and allocation of correct sitecodes based on the Melbourne Water Stream network (`mwstr`) reachcodes can also be achieved in a GIS package. To do this, you will need the `mwstr` spatial data (see Kunapo *et al.* 2019 for links to the data). Alternatively, correct streamcodes can be derived manually using the interactive tool at <https://tools.thewerg.unimelb.edu.au/mwstr/>.

We assume that you have a spatial layer of the site locations. Each site should have:

- a draft site code (perhaps a field called 'old_sitecode');
- a stream name (field `str_nm`), ensuring that the stream names exactly match the correct stream in the `mwstr` `stream_names` table. This can also be searched by the stream selection drop-down box at <https://tools.thewerg.unimelb.edu.au/mwstr/>;
- a location description (field `location`, as informative as possible, including distance upstream or downstream from a landmark that is unlikely to change).

This is an example of the process for preparing site data to match the database's site table, using data imported for sourcecode 58. Ideally, the data should be supplied with correct sitecodes derived using the stream network. If you are unable to do this, then include a field called `sitecode` in the table, and give each site a unique code equivalent to the draft site code.

```

#import site spatial Layer
new_sites <- sf::st_read("pyrites_sites.shp")
#import mwstr stream-line Layer (assuming it is in the same folder as above)
mwstr <- sf::st_read("mwstr_v1.1.shp") #you may need to change the version number
stream_names <- read_excel("mwstr_tables_v1.1.xlsx", sheet = "stream_names") #ditto
# # Or if you have access to the database
# mwstr <- sf::st_read(mwstr_db, query = "SELECT * FROM streams")
# stream_names <- RPostgreSQL::dbReadTable(mwstr_db, "stream_names")
# sites <- sf::st_read(mwbugs_db, query = "SELECT * FROM sites")
new_sites$strcode <- stream_names$strcode[match(new_sites$str_nm, stream_names$str_nm)]
sum(is.na(new_sites$strcode)) #0 if not zero, some names are not perfect matches with stream_names table

#use snapSiteToReach() function in bugDatabaseFunctions.R to move sites to stream line and
assign them a reachcode
#and a dus (thereby a sitecode)
new_sites1 <- new_sites[0,]
for(i in 1:dim(new_sites)[1]){
  new_sites1 <- rbind(new_sites1,
    snapSiteToReach(point = new_sites[i,], #sf POINT object with a single point
      str_nm = new_sites$str_nm[i]))
}
#new_sites1[order(new_sites1$dist_moved, decreasing = TRUE),]
max(new_sites1$dist_moved) #19.8 m - acceptably close as to not raise suspicions of error
sum(new_sites1$strcode != new_sites1$str_nm_assumed) #0 if not zero, then there is a name mismatch
new_sites1$sitecode <- paste(new_sites1$reachcode,new_sites1$dus,sep = "-")
sum(duplicated(new_sites1$sitecode)) #if not zero, there are duplicate sitecodes.
#If they are not the same site, manually rename them by adjusting dus if possible.

#Check no existing sites in database that should be considered the same site.
plot(sites$geometry[sites$reachcode %in% new_sites1$reachcode], col = "green")
plot(new_sites1$geometry[new_sites1$reachcode %in% sites$reachcode], col = "red", cex = 2,
  add = TRUE)
plot(mwstr$geometry[grepl("GDM",mwstr$reachcode)], col = "blue", add = TRUE)
plot(mwstr$geometry[grepl("PYR",mwstr$reachcode)], col = "blue", add = TRUE)
#plotting them shows that three of the new_sites1 should be given the same sitecode as existing sites
#make minor adjustments to two of their sitecodes to (subtract 2 to the dus value) to make them equal existing sites
new_sites1$sitecode[match(c("GDM-5569-8","GDM-5604-8"),new_sites1$sitecode)] <- c("GDM-5569-6","GDM-5604-6")
sum(new_sites1$sitecode %in% sites$sitecode) #3 sites already in sites table.

#Now prepare sites table to match database sites table.
new_sites1$alt_location <- NA #Unless you have alternative location descriptions
new_sites1$old_sitecode <- new_sites1$site #Keep record of sitecodes used in preparation of dataset
new_sites1$mgae <- sf::st_coordinates(new_sites1)[,1]
new_sites1$mgan <- sf::st_coordinates(new_sites1)[,2]
new_sites1$lgla <- "Moorabool"
#rearrange to match order of fields in sites
new_sites1 <- new_sites1[match(names(sites),names(new_sites1))]

```

Preparation of samples table

The samples table should contain relevant details of all samples collected (and those unable to be collected if the stream was dry) in the project. If the sample contains no biotic data (for instance if the stream was dry, and the sample was unable to be collected), then that should be noted in the comment field for that sample.

The samples table should contain the fields:

- smpcode (see below).

- date. The date field should be formatted as a date in yyyy-mm-dd format.
- sitecode. All values must have a matching value in the sites table.
- monthcode. This can be calculated from the date using the calcMonthDate() function in bugDatabaseFunctions.R in R, or by the following formula in Excel (assuming the date is entered in cell A1): "=VALUE((YEAR(A1) - 1990)*12 + MONTH(A1))".
- area_m2. Surface area of the sample, for quantitative collection methods only. Leave blank (or NA in R) if sampled using RBA or another qualitative method.
- replicate. The default value for replicate is 1. If more than one sample were collected from the same habitat using the same collection method, then the second sample should have replicate = 2, etc. Note it is possible for the same sample to have been processed in different ways to arrive at different biological data. (For instance, a single sample could be field-sorted, and the residue kept and sorted in the lab, permitting assembly of a lab-sorted version of the sample). In such cases, the two versions of the sample would have the same replicate, but different method suffixes. (e.g. Walsh (2006)).
- collection_method. All values must have a matching value in the collection_methods table.
- processing_method. All values must have a matching value in the processing_methods table.
- nsamples. If not a composite sample from multiple sample units, this should = 1.
- subsample_perc. The percentage of the sample that was subsampled for sorting. Leave empty if field-sorted (NA in R), 100 if fully sorted in the lab.
- sourcecode. Left empty: we will allocate appropriate sourcecodes on entering the data into the database.
- comment. Any relevant explanatory comments about the sample.
- embargoed. Probably 0. Let us know if you want the data embargoed for a period.
- old_sitecode. Must either have a matching value in the sites table (a record of sitecodes used in preparation of dataset), or be empty.
- old_samplecode. Like old_sitecode, this can be used to keep a record of the sample codes used in preparation of the dataset.

With all other fields compiled, the smpcode field can then be compiled by concatenation:

- in R, "paste(samples\$monthcode,"-", samples\$sitecode,"-", samples\$replicate,"-", cm\$ccode[cm\$collection_method == samples\$collection_method], pm\$ccode[pm\$processing_method == samples\$processing_method], sep =""")" (where cm and pm are the collection_methods and processing_methods tables, respectively);

- in Excel, '=A1&"-"&A2&"-"&A3&"-"&A4&A5', assuming that monthcode, sitecode, replicate, ccode, and pcode are in the first five columns of row A.

This is an example of the process for preparing samples data to match the database's site table, using data imported for sourcecode 58.

```
#Data provided for this study was undesirably minimal
cm <- RPostgreSQL::dbReadTable(mwbugs_db, "collection_methods")
pm <- RPostgreSQL::dbReadTable(mwbugs_db, "processing_methods")
samples <- RPostgreSQL::dbReadTable(mwbugs_db, "samples")
spv <- RPostgreSQL::dbReadTable(mwbugs_db, "sample_provenance")

save(samples1, file = "pyrites_samples.rda")
new_samples <- get(load("pyrites_samples.rda"))
new_samples$old_sitecode <- new_samples$site
new_samples$sitecode <- new_sites1$sitecode[match(new_samples$old_sitecode, new_sites1$old_sitecode)]
#calculate monthcode using calcMonthDate() in bugDatabaseFunctions.R
new_samples$date <- as.Date(new_samples$date)
new_samples$monthcode <- calcMonthDate(new_samples$date)
#samples are all rapid bioassessment samples Lab-sorted to 300 individuals, identified to family
#2 samples at all sites, some riffle-edge, some two edges...
new_samples$collection_method <- "RBA edge (sweep)"
new_samples$collection_method[tolower(new_samples$hab) == "kick"] <- "RBA riffle (kick)"
new_samples$processing_method <- "lab-subsample to 300, ID to family"
new_samples$area_m2 <- NA
new_samples$nsamples <- 1
new_samples$subsampling_perc <- new_samples$ssPerc
new_samples$sourcecode <- 58
new_samples$comment <- NA
new_samples$embargoed <- 0
new_samples$old_samplecode <- new_samples$sample
new_samples$replicate <- 1
new_samples$smpcode <- with(new_samples,
  paste(monthcode, "-", sitecode, "-", replicate, "-",
    cm$ccode[match(collection_method, cm$collection_method)],
    pm$pcode[match(processing_method, pm$processing_method)],
    sep = ""))
new_samples$replicate[duplicated(new_samples$smpcode)] <- 2
new_samples$smpcode <- with(new_samples,
  paste(monthcode, "-", sitecode, "-", replicate, "-",
    cm$ccode[match(collection_method, cm$collection_method)],
    pm$pcode[match(processing_method, pm$processing_method)],
    sep = ""))
new_samples <- new_samples[match(names(samples), names(new_samples))]
```

Preparation of biota table

If you are beginning with written lab-sheets containing taxon names, then the data-entry app (<https://tools.unimelb.edu.au/bugDataEntry/>) is a useful resource for preparing a biota table with the required structure and content. The table should be provided as a table with 5 or 6 columns:

1. smpcode, with each value having a matching entry in the samples table. The only entries in the samples table that should not have matches in the biota table are those samples that contained no biota (for instance because the stream was dry at the time of sampling).

2. taxoncode, with each value being a valid taxoncode (i.e. an 8-character code equivalent to a shortcode value in the taxon_all table, filled with trailing 9s or 0s if the shortcode is <8 characters long).
3. taxon. While not absolutely necessary, including taxon names in the table is a useful level of redundancy to aid checking for any inconsistencies between taxoncodes and taxon names.
4. count. The number of specimens of each taxon counted. For subsampled samples, these should be raw counts (**Do not multiply the counts by 100/subsample size**). For any coarse-picked specimens, multiply their count by the subsample proportion. (For instance, if two *Cherax* specimens are picked out of the sample before sorting a 15% subsample, enter a count value for these two specimens as $2 * 15 / 100 = 0.3$, and set the coarsepick field for this specimen to 1). If a further *Cherax* specimen is found in the subsample, enter that specimen as a separate entry, with a count of 1 and coarsepick = 0. Thus, in this case *Cherax* would have two entries in the one sample.
5. coarsepick. Default value is zero. Should only be 1 for specimens that have been picked from a sample before (or after) subsampling.
6. notes. Any relevant notes on the specimen. For instance, explain why a specimen could not be identified to the desired taxonomic level because (was it immature or damaged?)

Compilation of data file for submission

The three core tables (new_sites, new_samples, new_biota) should be supplied in an appropriate format (e.g new_samples, and new_biota could be 2 separate csv files, a 2-sheet excel file, and new_sites could be an ESRI shapefile, or new_samples and new_biota could be saved as data.frame objects, and new_sites as an sf object, all saved as an .Rdata file) together with a cover letter providing information about the project that led to the data (see above).

Notes on using the web interface

The web interface for the database (<https://tools.thewerg.unimelb.edu.au/mwbugs/>) allows users to select subsets of data from the database by a range of criteria. Sites can be selected based on:

- The marine segment they drain to;
- The major stream catchment they are in (Melbourne Water's Healthy Waterways Strategy breaks the region up into units that they term Catchments and Subcatchments, but these are not hydrologically correct catchments: an option to select by these units will be made available in the future);
- Stream name;
- Project (as defined in the sample_provenance table);
- Local Government Authority.

In all cases, except when selecting on project, samples can be selected based on:

- Data type;
- Taxonomic resolution;
- Time period.

For 'Projects', all of the data collected for that project is provided without an option for subsetting.

Summary information on the number and types of samples at each site can be accessed on an interactive map, and, if the selected subset contains data from 50 or fewer sites, a summary table of the numbers and types of samples is shown.

The web interface compiles selected data into an Excel file which can be downloaded, together with an explanatory document providing a summary of the data downloaded, including licence information, metadata, and cautionary notes on using the downloaded data. *You should read the explanatory document before using downloaded data.* The compiled data in the download file differs from the database data in several ways.

1. The samples table includes a SIGNAL score field calculated from the biota data to provide a quick stream-health indicator.
2. The count field in the biota table is adjusted depending on the 'data type' chosen:
 - a. if "presence-absence (all collection methods)" is selected, then all samples in the chosen sites are selected, and counts are converted to 1.
 - b. if "unbiased count (lab-sorted methods, for rarefaction analysis)" is selected, then only lab-sorted samples are selected. For subsampled samples any coarse-picked specimens are excluded, and raw counts are provided to permit rarefaction estimates of taxon richness (e.g. Gotelli & Colwell 2001).

- c. if “abundance per unit effort (lab-sorted methods)” is selected, only lab-sorted samples are also selected. But in this case, coarse-picked specimens are included, and counts are multiplied by $100 \times \text{subsamp_perc}$ to provide a standardised estimate of abundance in each sample.
 - d. if “number per m² (quantitative methods)” is selected, then only samples collected by quantitative methods (e.g. airlift sampler, Hess sampler or snag-bag) are included. Counts are multiplied by $100 \times \text{subsamp_perc}$ and divided by the sample area in m² to give a measure of density.
3. Taxa in the biota table are combined if taxonomic resolution of “family” or “genus” is selected. This is done by converting the last 4 characters of each taxoncode to “9999” for family level, or the last two characters to “99” for genus level, and then summing counts for all resulting unique taxoncodes. If “lowesttaxon” is selected, taxoncodes are unchanged, but only samples with processing_method indicating they were identified to lowest taxon are selected. If “genus” is selected, samples identified to either genus or lowest taxon are selected. If “family” is selected, all samples are selected.
4. The taxonomy table in the download file is a compiled list of all taxa in the dataset using the function code `Taxonomy()` (see above).

References

(The full reference bibTex database used by this report and the web interface can be downloaded at <https://tools.thewerg.unimelb.edu.au/data/mwbugs/bugDatabase.bib>.)

AWT. (1999) *Merri and Darebin Creeks Biological Monitoring Program*. AWT report No., Australian Water Technologies, Melbourne, Australia.

Bond, N., Thomson, J. & Reich, P. (2012) *Macroinvertebrate Responses to Antecedent Flow, Long-Term Flow Regime Characteristics and Landscape Context in Victorian Rivers*. Report, National Water Commission.

Boulton, A.J. (1985) A sampling device that quantitatively collects benthos in flowing or standing waters. *Hydrobiologia*, **127**, 31–39.

Butcher, R. (2003) *Developing a Long-Term Environmental Fixed Sites Monitoring Program for Melbourne Water*. Biological Sciences, Monash University, Melbourne.

Cameron, A. & Vertessy, D. (1995) *Macroinvertebrate Survey of the Diamond Creek Basin and Stony Creek, May 1994*. WES Scientific Report, Water EcoScience, Melbourne.

Cameron, A. & Vertessy, D. (1999) *Aquatic Macroinvertebrate Survey of the Diamond Creek Basin. Final Report*. ALS Consultants, Melbourne.

Chessman, B.C. (2003) New sensitivity grades for Australian river macroinvertebrates. *Marine and Freshwater Research*, **54**, 95–103.

Coleman, R. (2006) *Ecological and Physical Responses to Large Woody Debris Reintroduction in the Little Yarra River, Victoria, Australia*. Thesis.

Coleman, R. & Pettigrove, V. (1998a) *Waterway Assessment in the Yarra Catchment: The Health of the Watts River*. Waterways; Drainage, Melbourne Water Corporation, Melbourne.

Coleman, R. & Pettigrove, V. (1998b) *Waterway Assessment in the Yarra Catchment: The Health of the Andersons and Jumping Creeks*. Waterways; Drainage, Melbourne Water Corporation, Melbourne.

Coleman, R. & Pettigrove, V. (2001) *Waterway Assessment in the Western Port Catchment: The Health of the Lang Lang River*. Waterways Group, Melbourne Water Corporation, Melbourne.

Crowther, D., Canale, G. & Papas, P. (2001) *Melbourne Water Biological Monitoring Program 2000-2001*. Arthur Rylah Institute for Environmental Research, Department of Sustainability; Environment, Melbourne.

Crowther, D., Papas, P. & MacKay, S. (2003) *Melbourne Water Biological Monitoring Program 2002-2003*. Arthur Rylah Institute for Environmental Research, Department of Sustainability; Environment, Melbourne.

EPA Victoria. (2003) *Rapid Bioassessment Methodology for Rivers and Streams*. Guideline for environmental management, Publication No., Environment Protection Authority Victoria, Melbourne.

- EPA Victoria. (2004) *Biological Objectives for Rivers and Streams–Ecosystem Protection*. Report, Environment Protection Authority Victoria.
- GHD. (2013) *Ecological Benefits of Artificial Large Woody Debris in Hoddles Creek, Spring 2012*. Report for Melbourne Water, Melbourne Water.
- Gotelli, N. & Colwell, R. (2001) Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.
- Government of Victoria. (2001) *Variation to State Environment Protection Policy (Waters of Victoria): Insertion of Schedule F8, Waters of the Western Port and Catchment*. Report, The Craftsman Press.
- Government of Victoria. (2004) State Environment Protection Policy (Waters of Victoria) 23/2/1988, No. S13 Gazette 26/2/1988, as varied to 2004.
- Grace Detailed-GIS Services. (2012) *Directly Connected Imperviousness Compilation for Melbourne Water Selected Catchments*. Report.
- Growns, J.E., King, A.J. & Betts, F.M. (1999) The snag bag: A new method for sampling macroinvertebrate communities on large woody debris. *Hydrobiologia*, **405**, 67–77.
- Hardwick, R. (1998) *Waterway Assessment in the Western Port Catchment: The Environmental Health of Western Port Peninsula Streams*. AWT Report No. 267/98 for Waterways and Drainage, Melbourne Water, Australian Water Technologies, Melbourne.
- Hardwick, R. & Lewin, K. (1999) *Autumn 1996-Spring 1997 Macroinvertebrate and Sediment Toxicants Monitoring Program*. AWT report No., Australian Water Technologies, Melbourne, Australia.
- Hardwick, R. & Waller, K. (1999a) *1998-1999 Melbourne Water Monitoring Program*. AWT report No., Australian Water Technologies, Melbourne, Australia.
- Hardwick, R. & Waller, K. (1999b) *Plenty River Ecological Assessment*. AWT report No., Australian Water Technologies, Melbourne, Australia.
- Hellawell, J.M. (1978) *Biological Surveillance of Rivers. A Biological Monitoring Handbook*. Water Research Centre, Stevenage, UK.
- Hess, A. (1941) New limnological sampling equipment. *Limnological Society of America Special Publication*, **6**, 1–5.
- Kunapo, J., Walsh, C.J. & Sammonds, M.J. (2019) *Melbourne Water Stream Network. Version 1.0*. Report, School of Ecosystem; Forest Sciences, The University of Melbourne.
- McCasker, N. (2002) *Macroinvertebrate Colonization of Re-Introduced Large Woody Debris in the Little Yarra River*. Thesis.
- Norris, R.H. & Norris, K.R. (1995) The need for biological assessment of water quality: Australian perspective. *Australian Journal of Ecology*, **20**, 1–6.
- Papas, P., Crowther, D., Nicol, M., Canale, G., McKay, S. & Schreiber, S. (2002) *Melbourne Water Biological Monitoring Program - 2001-2002*. Arthur Rylah Institute for

Environmental Research, Department of Natural Resources; Environment, Victoria, Melbourne, Australia.

Papas, P.J., Nicol, M. & Crowther, D. (2000) *Melbourne Water Biological Monitoring Program - 1999/2000*. Arthur Rylah Institute for Environmental Research, Department of Natural Resources; Environment, Victoria, Melbourne, Australia.

Pettigrove, V. & Coleman, R. (1999) *Waterway Assessment in the Dandenong Valley: The Health of Corhanwarrabul, Monbulk and Ferny Creeks*. Melbourne Water Corporation, Melbourne.

Smith, T., Vertessy, D. & Hardwick, R. (1997) *StreamWatch Biological Monitoring Program. Annual Report 1996*. Melbourne Water Corporation, Melbourne.

Victoria, G. of. (1999) *Variation to State Environment Protection Policy (Waters of Victoria): Insertion of Schedule F7, Waters of the Yarra Catchment*. Report, The Craftsman Press.

Walsh, C.J. (2004a) *Impacts of Stormwater Treatment Wetlands on Stream Macroinvertebrates: A Study of Four Wetlands Constructed by Melbourne Water on Streams in Eastern Melbourne*. Report for Melbourne Water, Water Studies Centre, Monash University, Melbourne.

Walsh, C.J. (2004b) Protection of in-stream biota from urban impacts: Minimise catchment imperviousness or improve drainage design? *Marine and Freshwater Research*, **55**, 317–326.

Walsh, C.J. (2006) Biological indicators of stream health using macroinvertebrate assemblage composition: A comparison of sensitivity to an urban gradient. *Marine and Freshwater Research*, **57**, 37–47.

Walsh, C.J. (2018) *Melbourne Water Macroinvertebrate Spring 2018 Monitoring Program Proposal*. Report, School of Ecosystem; Forest Sciences, The University of Melbourne.

Walsh, C.J., Bloink, C. & Hehir, G. (2014) *Preliminary Report on Macroinvertebrate Assemblages of Pyrites and Goodmans Creek in Relation to the Management of Merrimu Reservoir: Patterns in 2013 in Early Stages of Flow Manipulation*. Report, Waterway Ecosystem Research Group, The University of Melbourne.

Walsh, C.J. & Breen, P.F. (2001) A biological approach to assessing the potential success of habitat restoration in urban streams. *Verhandlungen Internationale Vereinigung für Theoretische und Angewandte Limnologie*, **27**, 3654–3658.

Walsh, C.J., Fletcher, T.D., Bos, D.G. & Imberger, S.J. (2015) Restoring a stream through retention of urban stormwater runoff: A catchment-scale experiment in a social-ecological system. *Freshwater Science*, **34**, 1161–1168.

Walsh, C.J. & McCasker, N. (2002) *Melbourne Water Macroinvertebrate Database 1994 – 2002. A Report on Its Compilation, with Recommendations for Its Use and for Future Biological Monitoring*. Report, CRC for Freshwater Ecology.

Walsh, C.J., Sharpe, A.K., Breen, P.F. & Sonneman, J.A. (2001) Effects of urbanization on streams of the Melbourne region, Victoria, Australia. I. Benthic macroinvertebrate communities. *Freshwater Biology*, **46**, 535–551.

Walsh, C.J., Waller, K.A., Gehling, J. & Mac Nally, R. (2007) Riverine invertebrate assemblages are degraded more by catchment urbanization than by riparian deforestation. *Freshwater Biology*, **52**, 574–587.

Water Technology, GHD & Streamline Research. (2014) *Capital Works Monitoring Program*. linked site summary tool, Melbourne Water, Melbourne, Australia.

White, J.Y. (2018) *The Effects of Urban Land Use, Riparian Forest Cover, and Instream Habitat Complexity on Macroinvertebrate Assemblages*. PhD Thesis, The University of Melbourne, Melbourne Australia.

White, J.Y. & Walsh, C.J. (in prep) Catchment-scale urbanization diminishes effects of habitat complexity on instream macroinvertebrate assemblages. *Ecological Applications*.

WSL. (2004) *Melbourne Water Biological Monitoring Program–2003/2004*. Report for Melbourne Water Catchment and Waterways, WSL consultants Pty Ltd, Melbourne.

WSL. (2005) *Melbourne Water Biological Monitoring Program–2004/2005*. Report for Melbourne Water Catchment and Waterways, WSL consultants Pty Ltd, Melbourne.