03-200

Environment Canada Environnement Canada

Canada

The Electronic Compilation of Lipid Data From Post-1960 Aquatic Science Publications

By

N. Marenco, M. Arts, T. Spencer, A. Turnbull

NWRI Contribution No. 03-200

NATIONAL WATER RESEARCH INSTITUTE
INSTITUT NATIONAL DE RECHERCHE SUR LES EAUX

The Electronic Compilation of Lipid Data from Post-1960 Aquatic Science Publications

Nadia Marenco[1], Michael T. Arts[1,2], Tyler Spencer[1] and Andrew M. Turnbull[1]

NWRI Cont. # 03-200

[1] National Water Research Institute, Environment Canada, 867 Lakeshore Road, PO Box 5050, Burlington, Ont., L7R 4A6
[2]Author to whom all correspondence should be address

The Electronic Compilation of Lipid Data from Post-1960 Aquatic Science Publications

Nadia Marenco, Michael T. Arts, Tyler Spencer and Andrew M. Turnbull

ABSTRACT

One direct way to assess the "health" of aquatic ecosystems is to monitor changes in the quantity and quality of lipids in key species within the system. Lipids are a family of naturally occurring compounds which can be defined by their insolubility in water and solubility in non-polar solvents. The largest component of lipids are the fatty acids (FA). Fatty acids are the main constituents of cell membranes and, as such, mediate a wide variety of cellular processes. Numerous studies, over the last two decades, have established that certain FA are highly beneficial in terms of animal and human health. This revelation has led to a marked increase in their use in a broad spectrum of ecological studies. The objective of this project was to organize and compile lipid data from post-1960 aquatic-science publications into an easily accessible, manageable, searchable and expandable format. This report outlines the structure, content, format (procedure for electronic compilation) and intended use of the database entitled "LipidBase.mdb". This database allows researchers to make rapid comparisons for example, in concentrations of specific FA amongst species or within species located in different environments. Thus, it provides researchers with a flexible, heuristic tool to explore patterns in fatty acid signatures in a wide variety of organisms and tissues collected in diverse locations in both freshwater and marine locales.

Compilation électronique des données sur les lipides dans les publications en sciences aquatiques depuis 1960

Nadia Marenco, Michael T. Arts, Tyler Spencer and Andrew M. Turnbull

RÉSUMÉ

L'une des façons d'évaluer directement la « santé » des écosystèmes aquatiques consiste à évaluer la quantité et la qualité des lipides chez les espèces clés. Les lipides sont une famille de composés naturels que l'on peut définir par leur insolubilité dans l'eau et leur solubilité dans les solvants non polaires. Les acides gras (AG) sont la fraction des lipides la plus importante sur le plan quantitatif. Les AG étant les principaux constituants des membranes cellulaires, ils interviennent donc dans une large gamme de processus cellulaires. De nombreuses études au cours des deux dernières décennies ont établi que certains AG sont très favorables pour la santé humaine et animale. Cette révélation a donné lieu à une augmentation marquée de leur utilisation dans diverses études écologiques. Le présent projet avait pour objectif d'organiser et de compiler les données sur les lipides dans les publications en sciences aquatiques depuis 1960 sous une forme facilement accessible, gérable, consultable et extensible. Le rapport traite de la structure, du contenu, du format (méthode de compilation électronique) et de l'usage prévu de la base de données intitulée « LipidBase.mdb ». Cette base de données permet aux chercheurs d'établir des comparaisons rapides, par exemple les concentrations d'un certain AG chez plusieurs espèces ou chez une espèce habitant dans des milieux différents. Elle fournit donc aux chercheurs un outil heuristique souple permettant d'étudier les profils (signatures) d'AG d'une grande variété d'organismes et de tissus prélevés à divers endroits dans des habitats tant marins que dulçaquicoles.

NWRI RESEARCH SUMMARY

Plain language title

An electronic, searcheable, and expandable archive of lipid/fatty acid data from aquatic organisms.

What is the problem and what do sicentists already know about it?

Data on the fatty acids (lipids) contents of aquatic organisms has been collected in many places and researchers for about 5 decades now. However, this data has not been published in a consistent format nor has it been assembled in one place in any sort of comprehensive manner. We now know that some fatty acids are essential to animals in order to maintain their physiological competencies at an optimum level. We know that shortages of these essential fatty acids in their diets leads to health problems, poor survivorship and a general degradation in their physiological condition.

Why did NWRI do this study?

The scope and amount of lipid/fatty acid data present in the published data must be amalgamated into a cohesive, condensed, format if it is to be efficiently used for future analyses. Because of the importance of fatty acids to marine and freshwater ecosystems it was deemed advisable to create a searchable database containing fatty acid profiles from a wide range of aquatic organisms so that comparisons amongst/between species, locales, time periods etc can be made more efficiently.

What were the results?

We have produced a database with a commonly available software program (MicroSoft Access). The database is searchable using common logical operators (e.g. if, and, but, not). New data can be entered directly into the database or imported from another commonly available software program (Microsoft Excel). The database is consistent in the format of data entry so that intercomparisons between/amongst organisms can be made.

How will these results be used?

In a very real sense this database is a unique and expandable research tool. It will be used by NWRI researchers to provide our clients with a broader contextual understanding of how the fatty acid data that we provide on their particular organisms compares with similar data collected either on different organisms or on the same organisms collected in a different locale. The database can also be used heuristically i.e. to test hypothesis. For example, it may be desirable to know if levels of a particular fatty acid in an organism have changed over time or if one should, *apriori*, expect large difference between young and old organisms or between males and females.

Who were our main partners in the study?

Treasury Board provided the funding to hire the intern (Nadia Marenco, M.Sc.). YMCA administered this internship. The student was mentored by Dr. Michael Arts (NWRI).

Sommaire des recherches de l'INRE

Titre en langage clair

Une banque de données électronique consultable et extensible sur les lipides ou les acides gras d'organismes aquatiques.

Quel est le problème et que savent les chercheurs à ce sujet?

Des données sur la teneur en acides gras (lipides) des organismes aquatiques ont été recueillies à plusieurs endroits et provenant de plusieurs chercheurs sur une période d'environ 50 ans. Toutefois, ces données n'ont pas toutes été publiées sous la même forme et n'ont pas été rassemblées à un seul endroit. Nous savons maintenant que certains acides gras sont essentiels pour un fonctionnement physiologique optimal chez les animaux. Nous savons qu'une carence en ces acides gras essentiels dans le régime des animaux est responsable chez eux de problèmes de santé, menace leur survie et entraîne une dégradation générale de leur état physiologique.

Pourquoi l'INRE a-t-il effectué cette étude?

Les diverses données publiées sur les lipides ou les acides gras doivent être réunies sous une forme cohésive et condensé pour qu'elles soient faciles à utiliser dans les analyses futures. Étant donné l'importance des acides gras pour les écosystèmes marins et dulçaquicoles, il nous a semblé utile de créer une base de données consultable renfermant les profils d'acides gras d'une large gamme d'organismes aquatiques de manière à pouvoir établir plus facilement des comparaisons chez une ou plusieurs espèces, dans un ou plusieurs habitats et pour une ou plusieurs périodes données.

Quels sont les résultats?

Nous avons créé la base de données avec un logiciel courant (Microsoft Access). La base de données est consultable à l'aide d'opérateurs logiques courants (p. ex. si, et, mais, non). Les nouvelles données peuvent être saisies directement dans la base de données ou importées d'un autre logiciel courant (Microsoft Excel). Les données de la base de données sont toutes formatées de la même façon, de sorte qu'il est possible d'établir des comparaisons chez un organisme ou entre plusieurs organismes.

Comment ces résultats seront-ils utilisés?

Cette base de données est vraiment un outil unique et extensible. Les chercheurs de l'INRE s'en serviront pour permettre à leurs clients de mieux comprendre comment les données sur les acides gras obtenues pour certains organismes se comparent à des données similaires obtenues pour des organismes différents ou pour le même organisme, mais prélevé dans un habitat différent. La base de données peut également être utilisée de façon heuristique, c.-à-d. pour vérifier une hypothèse. Par exemple, il pourrait être souhaitable de savoir si la concentration d'un acide gras en particulier chez un organisme a changé avec le temps ou si l'on peut *a priori* s'attendre à une grande différence entre un jeune et un vieil organisme ou entre un mâle et une femelle.

Quels étaient nos principaux partenaires dans cette étude?

Le Conseil du Trésor a fourni les fonds pour engager une stagiaire (Nadia Marenco, M.Sc.) Le YMCA a administré ce stage. L'étudiante a été dirigée par Michael Arts (INRE).

TABLE OF CONTENTS

ABSTRACT
RESEARCH SUMMARY
List of Figures
List of Tables

1.0 INTRODUCTION 1

2.0 DATABASE STRUCTURE AND CONSTRUCTION 5
2.1 Structure 6
2.2 Data Model 7
2.3 Database Table Structure 8
2.4 Units and Qualifiers 9

3.0 DATABASE CONTENT 9
3.1 Tables in "LipidBase.mdb" 10
3.1.1 Update Log 10
3.1.2 References 11
3.1.3 Glossary 12
3.1.4 Nomenclature 12
3.1.5 Sample ID 13
3.1.6 Sampling Site and Time 13
3.1.7 Type, Mass, Length, Age, Sex 14
3.1.8 FishBase Trophic Levels 16
3.1.9 Methods 17
3.1.10 Lipid Content 18

4.0 QA/QC PROGRAM 19
4.1 Cross-checking References 19
4.2 Scanned Data 20
4.3 Max-Min Check 20
4.4 Cross-Referencing 20

5.0 VIEWING, EXPORTING AND APPENDING DATA 21

6.0 DATABASE MAINTENANCE 22

7.0 SUMMARY 22

8.0 ACKNOWLEDGEMENTS 22

9.0 REFERENCES 24

10.0 CONTACT INFORMATION 26

List of Figures

Figure 1. Schematic diagram illustrating the relationships between the "LipidBase.mdb" database tables.......... 27

Figure 2. Schematic diagram illustrating the orthogonal format of a "LipidBase.mdb" database table. 28

List of Tables

Table 1. Update Log 11

Table 2. References 11

Table 3. Glossary 12

Table 4. Nomenclature 13

Table 5. Sample ID 13

Table 6. Sampling Site and Time 14

Table 7. Type, Mass, Length, Age, Sex 15

Table 8. FishBase Trophic Levels 17

Table 9. Methods 17

Table 10. Lipid Content 19

1.0 INTRODUCTION

Lipids are a family of naturally occurring compounds grouped together on the basis of their insolubility in water and solubility in non-polar solvents (e.g. alcohols, hydrocarbons, chloroform, benzene, and ethers). Lipids contain carbon, hydrogen and oxygen and, together with carbohydrates and proteins, are the main constituents of living cells.

Lipids include FA and their derivatives, neutral fats, waxes and steroids. A FA consists of a long non-polar chain of carbon atoms, with a polar carboxylic acid group (an organic acid containing the functional group -COOH) at one end. Most natural FA contain an even number of carbon atoms and have a variety of structures. In aquatic organisms, for instance, they commonly contain from 14 to 24 carbon atoms and have varying degrees of saturation. They may be saturated (SAFA), unsaturated (USFA), monounsaturated (containing one double bond) (MUFA), polyunsaturated (containing two or more double bonds) (PUFA) or highly unsaturated (HUFA, 20 or more carbons and two or more double bonds).

Fatty acids are classified into three major groups;

- Essential - cannot be synthesized at all or in amounts sufficient to meet the organism's growth and development needs. Typically these are long chain (20 carbons or more) PUFA.
- Conditionally indispensable - essential but widely available in the diet or else the organism's needs can be mostly met by endogenous synthesis from short chain precursors. Typically these are PUFA with 18 carbons

- Nonessential - constantly available in the diet or readily synthesized from short chain precursors in amounts sufficient to meet the organism's physiological needs.

Organisms that cannot make EFA in amounts sufficient to meet their needs must consume them "pre-formed" as part of their diet. There are two families of EFA, omega-3 (abbreviated as n3, n-3 or ω3) and omega-6 (abbreviated as n6, n-6 or ω6). The n3 and n6 FA contain a double bond three carbons and six carbons from the terminal methyl group, respectively.

Carbons within a FA are identified by one of two nomenclature systems, n or delta (Δ). In the n-designation, carbon numbering begins from the methyl end. In the Δ-designation, carbon numbering begins from the carboxylic acid end. The number of carbon atoms in a FA is identified by the number preceding the colon. The number of double bonds in a FA is identified by the number following the colon. The position of the double bonds is also defined by one of the two nomenclature systems. Double bonds are usually spaced three carbons apart. Thus, n6 indicates a double bond in the n6, n9, n12 and n15 positions. For example, 20:4n6 = 20:4n-6 = 20:4Δ5, 8, 11, 14; there are 20 carbons in this FA and 4 double bonds. These double bonds are in the n6 (or Δ5), n9 (or Δ8), n12 (or Δ11) and n15 (or Δ14) positions depending on which nomenclature system is used.

Lipids are important body stores for aquatic organisms in several ways. Lipids supply energy for growth, development, prevention of over-winter starvation, are a source of EFA important in maintaining cell membrane function, are precursors for biologically active molecules that mediate an animal's stress response, cardiovascular function, immune function and neural development. Furthermore, there is a positive relationship between body lipid stores and reproductive success.

In the last two decades, numerous studies have established that certain fatty acids (EFA) are essential to human health (e.g., Stansby, 1984; Harris, 1985; Gerster, 1998; Connor, 2000). Fish and fish oils remain the major sources of EFA in the human diet. Supplementation with specific EFA appears to be useful as a treatment for certain neurological diseases, as well as to lower the levels of LDL (low-density lipoproteins; LDL is commonly known as "bad" cholesterol) in the blood (Stansby, 1984). In addition to lowering LDL, certain EFA also act to reduce platelet aggregation, decrease blood viscosity, prevent ischemic damage, and, to a limited extent, lower blood pressure (Stansby, 1984; Harris, 1985). Oil-rich fish products containing certain EFA have also been found to provide significant benefits in persons with various conditions including rheumatoid arthritis and inflammatory bowel diseases such as Crohn's Disease (Connor, 2000).

The British Nutrition Foundation has officially recommended a diet including 2-3 portions of fatty fish per week (equivalent to 3-4 g of fish oil per day) (Gerster, 1998). The current mean daily intake in a typical North American diet is far less than that. Consequently, companies marketing and selling marine fish (e.g. Brunswick Sardines, Connors Brothers Ltd., New Brunswick) have begun to aggressively market the "healthy" attributes of their products.

In addition, the United States Food and Drug Administration (FDA) is amending its regulations on food labelling to include trans fatty acids on the nutrition label immediately under the declaration of saturated fatty acids (FDA, 2003). This new regulation is intended to provide information that will assist consumers in maintaining healthy diets. The need for accessible data on the fatty acid (FA) composition of aquatic organisms has increased, not only because of their value to humans, but also because of their importance in animal husbandry (e.g. aquaculture).

Fatty acids are also important from an ecological perspective. Observed changes in aquatic ecosystems may be profound (e.g. the loss of the burrowing amphipod *Diporeia* from many areas in the Great Lakes) but the implications of these changes is not always clear. One direct way to quantify the impacts of such perturbations is to monitor changes in the quantity and quality of lipids in key species within an ecosystem. For example, numerous studies have shown that FA can be used as general indicators of health, stress, and reproductive condition in aquatic ecosystems (e.g., Sargent et al., 1987; Iverson, 1993; Virtue et al., 2000). Combinations of, or the whole suite of FA in a given tissue or in the whole organism are often referred to as an organism's FA signature (Iverson, 1993). Certain FA can be used as tracers to study an organism's position in the aquatic food web (i.e. trophic status) as well as markers differentiating amongst pelagic-, benthic- and terrestrially-derived carbon inputs to the diet (Sargent et al., 1987).

Fish FA compositions and contents are known to vary significantly amongst species and also in space and time. A number of factors, such as age, sex, body mass and length are known to be important contributors to this variation. Environmental factors, such as the nature and availability of food, as well as the season and location of catch are also critical (Gruger, 1967; Exler et al., 1975).

Given the crucial role of FA in aquatic ecosystems and in human health coupled with the large amount of relatively unorganized lipid/FA data in the literature we decided to compile, organize, and standardize lipid data from various post-1960 aquatic science publications into a manageable, accessible and expandable format that can be easily searched using logical Boolean operators (e.g. "if", "and", "not", "or" and "but"). The database "LipidBase.mdb" is intended to provide researchers and managers with a tool to easily access and use the data. Additionally, the database is intended to be flexible in order that new lipid data, when available, can be easily incorporated.

This report outlines the structure, content, format, procedure for electronic compilation and intended use of the database entitled "LipidBase.mdb. The organisms emphasized in "LipidBase.mdb" are aquatic invertebrates, mammals and especially fish however, it should be noted that known fatty acid signatures of any aquatic plant, animal, or microbial species could be included. The following documentation also provides a step-by-step guide to append, view and export data in relation to "LipidBase.mdb".

2.0 DATABASE STRUCTURE AND CONSTRUCTION

Prior to "LipidBase.mdb", published lipid data had not yet been compiled into a consistent, integrated, searchable and seamless format. In part this is because older data was published before computers were widely available but also because lipid researchers themselves utilize a variety of analytical approaches; for example, expressing lipids and FA as a function of wet-weight rather than the preferred method standardizing lipids and/or FA on a dry-weight basis. "LipidBase.mdb" begins to address this historical reality. In 2002 and 2003, we started the process of incorporating data obtained through literature searches of primary and secondary publications as well as other relevant data sets ("grey" literature) into "LipidBase.mdb".

2.1 Structure

"LipidBase.mdb" is assembled using Microsoft Access® 2000, a relational SQL (structured query language) compliant database software package that provides rapid electronic access to the data. Microsoft Access® is a user-friendly, readily available program that can store vast amounts of information without loss of speed. Its "querying power", derived from the relationships between the various database tables, makes Access® ideal for manipulating and comparing large data sets. Refined, user-specific searches can be generated in Access® with the use of logical operators, alone or in combination. These searches provide output unique to the questions posed.

"LipidBase.mdb" contains data compiled in a family of ten unique, but linked tables. It offers several tools to browse and view the data through forms, and is intended for users wishing to develop their own queries. The multi-table structure permits independent updates of specific data sets. The database currently occupies approximately 10 MB of hard-drive space. The minimum system requirements needed to run Access® 2000 or 2002 are a Pentium 200 MHz computer with 32 MB RAM and 170 MB of available hard disk space. The 'recommended' system requirements to run Access® 2002 are Microsoft Windows® XP Professional on a Pentium 300 MHz computer with 128 MB RAM.

Access® can theoretically hold hundreds of MB of data, however Access® databases containing thousands of records that begin to see heavy traffic will run at limited speeds. Access® is an inexpensive database solution for small projects, such as this one, but is not intended as a database for internet applications. In fact, a handful of users running simultaneous queries can crash a large Access® database. If a database such as this becomes extremely large (hundreds of MB of data) the data contained in "LipidBase.mdb" can be exported to other database software (e.g., SQL Server® and Oracle®).

2.2 Data Model

"LipidBase.mdb" is based on the relationship between the "Reference Manager ID" and the "Sample ID". The vast majority of data obtained from scientific publications can be referenced to originate from both a "Reference Manager ID" and/or from a "Sample ID". Each used publication was manually labeled with its own "Reference Manager ID", a unique value that was assigned, for indexing purposes, to each aquatic-science publication whose data was incorporated into the database. This ID number is identical to its "filing" number in a reference manager database generated with the Reference Manager Professional Edition® software (ResearchSoft, CA, V.10.0). Within this Reference Manager database, the reference information (e.g., title, author(s), year of publication, journal name, volume, issue, page numbers) of each publication was digitally input before being filed away for future use.

Because each publication contained, in most cases, data pertaining to several different aquatic species, each species had to be additionally "indexed" in order to maintain the referential integrity within the database. Consequently, the "Sample ID" was designed as an ID for each unique species or, for example, to differentiate between similar species that were obtained from different sampling sites or during different seasons.

To give but one example, let us assume the first publication used was indexed with a "Reference Manager ID" of 2345 and the publication contained fatty acid data on 5 species of fish. After assigning the "Reference Manager ID", each fish would have been assigned a unique "Sample ID". For instance, the first fish would have a "Sample ID" of 1, and the second a "Sample ID" of 2 and so on up to 5, until all species were indexed. Because all fish in this one publication necessarily share the same "Reference Manager ID", we know the data originates from the same publication. However, because each fish have a different "Sample ID", we know the fish differ from one another; in this case they are a different species.

Most data measurements or descriptions, with the exception of a few tables, such as the "Nomenclature" and "Update Log" tables, are referenced to a "Reference Manager ID" and/or "Sample ID". The links between the tables resulting are shown in Figure 1.

2.3 Database Table Structure

Most tables in "LipidBase.mdb" have been assembled according to the requirements of the second or third normal form (2NF or 3NF). This means tables are "orthogonal" or normalized, with each record devoted to a specific data point. An example of this format is shown in Figure 2. Although this may not be an intuitive format for reviewing data, it offers advantages for database searching and storage. More conventional tables can be constructed using Access® query and report capabilities or by exporting to spreadsheet software. Detailed information on database principles and structure can be found in O'Neil (1994).

2.4 Units and Qualifiers

Numerous units are used within "LipidBase.mdb". Because there is no one standard format of quoting fatty acid concentrations, and because the publications used are from a wide variety of sources, the data are presented with a wide variety of units. Most tables must therefore contain one or several fields displaying units of the data (e.g., %, mg/g).

In several tables, additional fields, such as "Comment", have been added to identify sources of the data as well as any relevant "qualifiers" (i.e. important fragments of information that cannot be incorporated into any other fields, but need inclusion in the database for completeness). For instance, if and when a particular fatty acid was recorded as a non-detect ("ND"), this 'text' could not be input into the 'numerical' "FA_Amt" field. It would thus be input into the "Comment" field.

3.0 DATABASE CONTENT

This section contains a description of the contents of each table in "LipidBase.mdb". Provided below is a list of all tables in "LipidBase.mdb". The definition of each field in all tables is included in each subsection. The fields included are outlined under "Field Name". The "Data Type" provides information on the data format in the field (e.g., text, numeric, date/time, etc). The "Description" provides a description of the field.

In broad terms, the contents of the database are:

- Update Log
- References
- Glossary

- Nomenclature
- Sample ID
- Sampling Site and Time
- Type, Mass, Length, Age, Sex
- FishBase Trophic Levels
- Methods
- Lipid Content

3.1 Tables in "LipidBase.mdb"

There are obvious limitations to the data provided in each table due to the different sources of the data (country, journal, background of researcher, available technology and methods of analysis at date of study, etc.) however an attempt was made, in part through the judicious selection of source data, to be as thorough and consistent as possible. The following sections describe the tables in "LipidBase.mdb" and provide details on the information they contain as well as their limitations.

3.1.1 Update Log

The "Update Log" table (Table 1) is not necessary for database functionality, because no fields in this table act as a "primary key field". It outlines the editorial details (e.g., editor's name, date of edit, etc.) of the database and acts as the sole "sign-in" sheet of database updates and use.

Table 1. Update Log

Field Name	Data Type	Description
Editor	Text	Name of database editor
Date	Date/Time	Date database updated
File Name	Text	Name of database file
Updated	Text	Yes: database updated; No: database not updated
Saved Changes	Text	Yes: changes to database saved; No: changes to database not saved
Backup Made	Text	Yes: backup of database made on CD; No: no backup of database made
CD No	Text	Number of CD database backup made on
CD Name	Text	Name of CD database backup made on

3.1.2 References

Technically, the "References" table is not necessary for database functionality either, because thus far, most of the data from each table are linked through the "Sample ID" and not the "Reference Manager ID". However, the "Reference Manager ID" field in this table was designed a "primary key field" and may be linked to all tables in the database containing this field. This table provides information on the scientific publications (e.g., primary and secondary authors, journal name, publisher, etc.) that were used for data entry into the database (Table 2).

Table 2. References

Field Name	Data Type	Description
Reference Manager ID	Number	Reference Manager ID of publication data obtained from
Primary Author	Text	Primary author on paper
Secondary Author(s)	Text	Secondary author(s) on paper
Reference	Text	Journal, edition, volume, issue, pages
Year	Number	Year of publication
Publisher	Text	Journal publisher

Included Yet	Text	Yes: data in paper included; No: data in paper not yet included; In part: only portion of data in paper included
In File	Text	Yes: have in file; No: don't have in file

3.1.3 Glossary

The "Glossary" table was created to resolve any ambiguities associated with the definition of the "lipid type" (i.e., FA_Type) entries and/or other terminology contained in the database. It is not necessary for database functionality, because no fields in this table act as a "primary key field". This table exists solely to provide common abbreviations, names and definitions for lipids and fatty acids in the database (Table 3).

Table 3. Glossary

Field Name	Data Type	Description
Abbreviation	Text	Lipid abbreviation
Name	Text	Name of lipid
Definition	Text	Definition of lipid

3.1.4 Nomenclature

The "Nomenclature" table was created to resolve any ambiguities associated with the abbreviation of "lipid type" (i.e., FA_Type) entries. It is not necessary for database functionality, because no fields in this table act as a "primary key field". This table contains standard lipid abbreviations used in the "Lipid Content" table of the database (Table 4). When updates are made to the "Lipid Content" table either in the form of changes or additions, the "Nomenclature" table should be consulted and the standard lipid abbreviation used for consistency.

Table 4. Nomenclature

Field Name	Data Type	Description
FA_Type	Text	Standard lipid abbreviation used in this database

3.1.5 Sample ID

The "Sample ID" table (Table 5) is the primary table necessary for database functionality, because the "Sample ID" field acts as a "primary key field". This table contains the "Reference Manager ID" and the "Sample ID". Each "Reference Manager ID" can have several corresponding sample ID's. However, a unique "Sample ID" can only have one "Reference Manager ID" (see Section 2.2: Data Model).

Table 5. Sample ID

Field Name	Data Type	Description
Reference Manager ID	Number	Reference Manager ID of publication data obtained from
Sample ID	Number	Sample ID of aquatic animal

3.1.6 Sampling Site and Time

The "Sampling Site and Time" table contains, when and if available, information on the sampling site(s), longitude, latitude, UTM coordinates, and sampling season of the data entries (Table 6).

In general, it is known total lipid decreases in the fall and winter to a minimum value in late winter and then begins to increase in the spring. Because of the variability of total lipid however, it is important that both the season and location of the catch is reported.

This information can be useful when attempting to extrapolate average values or for making broader correlations.

Despite the importance of some of these factors in determining the nutritional quality of aquatic lipids, few authors have stated the season or even the precise location from which their samples were obtained. Wherever latitude and longitude values were provided, they were converted to UTM coordinates by using online "converter" programs (e.g., Geographic/UTM Coordinate Converter, June 20, 2003. Charles L. Taylor). These programs work much like currency converters. The algorithms used for the conversion are "built-in" to the program such that no calculations have to be performed by the user.

Table 6. Sampling Site and Time

Field Name	Data Type	Description
Sample ID	Number	Sample ID of aquatic animal
Sampling Site	Text	Sampling site of aquatic animal
Latitude	Text	Latitude of sampling site
Longitude	Text	Longitude of sampling site
UTM Zone	Text	UTM zone
GPS Datum	Text	UTM datum
UTM E (m)	Text	UTM easting in meters
UTM N (m)	Text	UTM northing in meters
Sampling Month or Season	Text	Sampling month or season
Comment	Text	Miscellaneous comment

3.1.7 Type, Mass, Length, Age, Sex

The "Type, Mass, Length, Age, Sex" table is not necessary for database functionality, because no fields in this table act as a "primary key field". However, this table contains information, when and if available, on the animal type, genus, species, common name, trophic level (see Section 3.1.8: Daniel Pauly Trophic Levels), body mass, length, sex and age of the data entries (Table 7). Despite the importance of these factors in determining the nutritional quality of fish lipids, few authors have stated the body mass, length, age and/or even sex of their samples.

Both the scientific and common names of the aquatic species, when available, were included. Unfortunately, there may have been instances of nomenclature confusion due to the lack of standardization in the nomenclature of aquatic organisms. Until the nomenclature is standardized more thoroughly, the work of Froese and Pauly (2003) was used for species identification (see Section 3.1.8: FishBase Trophic Levels).

Table 7. Type, Mass, Length, Age, Sex

Field Name	Data Type	Description
Sample ID	Number	Sample ID of aquatic animal
Aquatic Organism Type	Text	Type of aquatic animal (fish/mammal/invertebrate)
Family	Text	Family of aquatic animal
Genus	Text	Genus of aquatic animal
Species	Text	Species of aquatic animal
Common Name	Text	Common name of aquatic animal
ISSCAAP	Number	"International Standard Statistical Classification of Aquatic Animals and Plants" level of aquatic animal
Trophic Level	Number	Trophic level of aquatic animal
seTL	Number	Standard error of trophic level of aquatic animal
Trophic Level Reference	Text	Trophic level reference information
Body Mass_Low (g)	Number	Body mass of aquatic animal in grams (low end of range)
Body Mass_High (g)	Number	Body mass of aquatic animal in grams (high end of range)
Body Mass (g)	Text	Body mass of aquatic animal in grams
Body Mass_SD (g)	Text	Standard deviation of body mass of aquatic animal in grams
Body Mass_Comment	Text	To indicate if the mass is a weight wet or dry weight
Length_Low (cm)	Text	Length of aquatic animal in cm (low end of range)

Length_High (cm)	Text	Length of aquatic animal in cm (high end of range)
Length (cm)	Text	Length of aquatic animal in cm
Length_SD (cm)	Text	Standard deviation of length of aquatic animal in cm
Sex	Text	Sex of aquatic animal
Age_Low (yrs)	Text	Age of aquatic animal in years (low end of range)
Age_High (yrs)	Text	Age of aquatic animal in years (high end of range)
Age (yrs)	Text	Age of aquatic animal in years
Age_SD (yrs)	Text	Standard deviation of age of aquatic animal in years
Comment	Text	Miscellaneous comment

3.1.8 FishBase Trophic Levels

Ecologists measure an organism's niche in terms of its trophic level; a feeding level, as often represented in a food chain or food web. Trophic levels can be assigned according to ratios of stable carbon and nitrogen isotopes. Primary producers comprise the bottom trophic level, followed by primary consumers (herbivores), then secondary consumers (carnivores feeding on herbivores), and so on. In marine environments, the "bottom" level contains primarily seaweed and phytoplankton, that serve as food for level two organisms, whose predators, in turn, make up level three, and so it goes up the marine food web to the apex, killer whales at trophic level five.

The "FishBase Trophic Levels" table is not necessary for database functionality, because no fields in this table act as a "primary key field". For species specifically targeted in "LipidBase.mdb", information on trophic level may be directly accessed in the previous table (Table 7). However this table contains trophic level data for species in addition to those provided in Table 7. The work of Froese and Pauly (2003) was used for aquatic species identification.

Table 8. FishBase Trophic Levels

Field Name	Data Type	Description
Scientific	Text	Scientific name of aquatic animal
English	Text	English name of aquatic animal
ISSCAAP	Number	"International Standard Statistical Classification of Aquatic Animals and Plants" level of aquatic animal
TL	Number	Trophic level of aquatic animal
SeTL	Number	Standard error of trophic level of aquatic animal
Reference	Text	Trophic level reference information

3.1.9 Methods

This table contains information, wherever available, on the field sampling methods, lipid extraction methods, as well as details on the gas chromatographic (GC) analyses (Table 9). Without this information, the lipid database would be incomplete.

The sampling and analytical procedures varied from paper to paper, and in many cases, information in the literature on the methods that were used to analyze the fatty acids were not always detailed. Publication of exact details of the analytical procedures (e.g., the size of the column, carrier gas, type of detector, etc.) would facilitate the comparison of data from different sources, as well as permit the intra-specific comparison of the various analytical methodologies.

Table 9. Methods

Field Name	Data Type	Description
Sample ID	Number	Sample ID of aquatic animal
Field Sampling Method	Text	Field sampling method used to obtain aquatic animal
Lipid Extraction Method	Text	Lipid extraction method used to obtain lipids from aquatic animal
FAME Method	Text	Fatty acid methyl ester method used to obtain FAMEs from aquatic animal
GC Type	Text	Gas chromatograph used in analyses
Detector	Text	Detector used with GC in analyses

GC Column Type	Text	GC column type used in GC analyses
Column Length (m)	Number	Length of column used in GC analyses in meters
Column ID (mm)	Number	Inner diameter of column used in GC analyses in millimeters
Column Film Thickness (um)	Number	Thickness of column film in micrometers
Column Coating	Text	Thickness of column film in micrometers
Carrier Gas	Text	Carrier gas used in GC analyses
Comment	Text	Miscellaneous comment

3.1.10 Lipid Content

This table contains all the fatty acid data provided by the various authors for those species detailed in the tables discussed above. It contains, when and if available, information on the animal tissue sampled, the fatty acid type, as well as the fatty acid amount, corresponding standard deviation and units.

The type of data reported varied considerably amongst publications. For some species, only the total lipid or lipid class composition were provided. Publication of exact details of the aquatic species (mass, length, age, sex, etc.) would facilitate the comparison of fatty acid data from different sources. Publications that provide these details should be given high priority for future inclusion to "LipidBase.mdb".

The total lipid content and fatty acid composition vary with the tissues analyzed, prompting creation of the "Tissue Sampled" column. Typically, sampling was from a pool of several fish of the same species or from a selected number of individual fish from which an average value was calculated.

Table 10. Lipid Content

Field Name	Data Type	Description
Sample ID	Number	Sample ID of aquatic animal
Tissue Sampled	Text	Tissue sampled for lipids of aquatic animal
FA_Type	Text	Fatty acid type
FA_Amt	Number	Fatty acid amount
FA_Amt SD	Number	Standard deviation of fatty acid amount
FA_Units	Text	Fatty acid amount units
Comment	Text	Miscellaneous comment

4.0 QA/QC PROGRAM

A thorough QA/QC program was implemented throughout data incorporation as well as upon completion. In general, the error identifying and trapping process was designed to identify: (a) error resulting from the manipulation of the data from its original form (usually digital) into the database; and (b) potential errors in the original data leading to anomalous values. Because much of the data provided for incorporation into "LipidBase.mdb" was compiled from hard copies of aquatic-science publications, subtle errors not resulting in anomalous data points were impossible to identify. Therefore, error trapping only addressed the relative accuracy of the data.

4.1 Cross-checking References

The systematic indexing system used for each publication involving the use of a reference database generated with the Reference Manager® software ensured that each publication used for "LipidBase.mdb" could be easily located and accessed for future research and data verification. Moreover, it guaranteed that any additions made to the database involved data from publications whose data had not yet been incorporated. Avoiding data duplication in this way was accomplished by cross-checking the new reference with those already stored in the Reference Manager® reference database. Reducing duplicates as much as possible was necessary to produce a concise database of minimal size.

4.2 Scanned Data

A large portion of the information in the "Glossary" table was only available in hard copy and was scanned in for incorporation into the database. The scanned data was filtered in Microsoft Excel® to identify and flag anomalies in nomenclature. Flagged data was either corrected or deleted prior to incorporation into "LipidBase.mdb".

4.3 Max-Min Check

Database tables were queried to display maximum and minimum values. Where raw data was available, the maximum and minimum values were compared to the raw data to ensure no errors developed during the process of importing and/or merging the data into "LipidBase.mdb". A comparison of the number of records was also carried out to ensure records were neither accidentally lost nor created.

4.4 Cross-Referencing

Cross-referencing queries were generated to ensure that data presented were indexed to a "Reference Manager ID". Each "Sample ID" must have a corresponding and existing "Reference Manager ID". A "Sample ID" can only have one "Reference Manager ID".

5.0 VIEWING, EXPORTING AND APPENDING DATA

Viewing the data in "LipidBase.mdb" can be accomplished in several ways. For instance, the database can simply be opened, and the table of choice selected by double-clicking for perusal. This presents the table, as is. However, if the user wishes to view only a certain portion of the entire data set, running a query specifying the needed parameters is the most efficient way. Queries can either be created in Microsoft Access® in "Design view" or by using the "Query wizard". The Microsoft Access® Help menu can be consulted for more information.

Exporting data can also be accomplished several ways. Within Microsoft Access®, the table can be selected, and the "Export" function used by accessing it under the File menu. Alternatively, the user has the option of copying and pasting the desired data into a spreadsheet. Depending on the amount of data required and the project, this may suit the requirements of the user.

Appending data is a more time consuming process if it is being imported from a spreadsheet file format such as Microsoft Excel®. First and foremost, the fields must be formatted such that a date and time field only contains the date and/or time, numeric fields contain only numbers and text fields contain only text and so on. In addition, if appending into an existing table, field names must be identical to those in the Access® table being appended into. Once all the formatting is complete, data can be appended by accessing the "Import" function under "Get External Data" under the File menu. Incidentally, this is the same way new data can be imported from Excel® into Access®. The alternative way of appending new data requires that the users open the desired table in Microsoft Access® and directly input the new data.

6.0 DATABASE MAINTENANCE

Fragmentation of the database can occur when data is imported or exported, deleted or moved within the database causing the database file to occupy excessive hard-drive space. To alleviate this problem, the database must be "compacted". To compact the database, close any open tables, select "Tools", "Database utilities", and "Compact database". The compacted database can be saved under the same name or under a new filename. Because the process of compacting involves writing to temporary files, a minimum free hard-drive space of 2.5 times the size of the database is recommended.

7.0 SUMMARY

This report is an overview of "LipidBase.mdb", a database tool created to address a general need in aquatic lipid research. A manual for utilizing this new tool is not provided, though descriptions of all the contents in the database are. To resolve any ambiguities associated with the abbreviation of field names and data entries, a detailed description of each field within "LipidBase.mdb" is included. Further details on data interpretation can be obtained from the referenced reports.

8.0 ACKNOWLEDGEMENTS

Suzanne Budge of the Biology Department of Dalhousie University (Halifax, NS) provided some data for "LipidBase.mdb" in electronic form. The Treasury Board, through the YMCA of Toronto, Ontario, Canada, provided funding for Nadia Marenco's internship at Environment Canada. We also wish to acknowledge the help of numerous librarians at the National Water Research Institute for their invaluable assistance in processing interlibrary loans.

9.0 REFERENCES

Conner, WE., 2000. Importance of n-3 fatty acids in health and disease. The American Journal of Clinical Nutrition, 71(1S), 171S-175S.

Exler, J., Kinsella, JE., Watt, BK., 1975. Lipids and fatty acids of important finfish: new data for nutrient tables. J. Am. Oil Chem. Soc., 52, 154-159.

Food and Drug Administration, 2003. Federal Register, Part III, Rules and Regulations, 21 CFR Part 101, 68(133), 41434-41506.

Froese, R., and Pauly, D., eds, 2003. Fishbase. World Wide Web electronic publication. www.fishbase.org, version 07, August 2003.

Geographic/UTM Coordinate Converter, June 20, 2003. Charles L. Taylor, http://home.hiwaay.net/ntaylorc/toolbox/geography/geoutm.html.

Gerster, 1998. Can adults adequately convert alpha-linolenic (18:3n-3) to eicosapentaenoic acid (20:5n-3) and docosahexaenoic acid (22:6n-3). International Journal of Vitamin and Nutrition Research, 68, 159-173.

Gruger, EH. Jr., 1967. Fatty acid composition. In *Fish Oils: Their Chemistry, Technology, Stability, Nutritional Properties and Uses*, ed. ME Stansby. AVI Publishing Co., Westport, CT, chap. 1, pp. 3-30.

Harris, WS., 1985. Health effects of omega-3 fatty acids. Contemp. Nutr. (General Mills), 10(8), 2 p.

Iverson, SJ., 1993. Milk secretion in marine mammals in relation to foraging: can milk fatty acids predict diet? Symp. Zool. Soc. Lond. 66, 263-291.

O'Neil, P., Database - Principles, Programming, Performance, San Francisco, USA: Morgan Kaufmann Publishers, Inc., 873pp., 1994.

Sargent, JR., Parkes, RJ., Mueller-Harvey, I., and Henderson, RJ., 1987. Lipid biomarkers in marine ecology. In *Microbes in the sea*. Edited by MA Sleigh. Wiley and Sons, New York. Pp. 119-138.

Stansby, ME., 1984. Fish or fish oil in the diet and heart attacks. Mar. Fish. Rev., 46(2), 60-63.

Virtue, P., Mayzaud, P., Albessard, E., Nichols, P., 2000. Use of fatty acids as dietary indicators in northern krill, *Meganyctiphanes norvegica*, from northeastern Atlantic, Kattegat, and Mediterranean waters.

10.0 CONTACT INFORMATION

Dr. Michael T. Arts
National Water Research Institute
Environment Canada
867 Lakeshore Rd., PO Box 5050
Burlington, ON, L7R 4A6
Tel.: (905) 336-6460
Fax: (905) 336-4699
Email: Michael.Arts@ec.gc.ca

Tyler Spencer
National Water Research Institute
Environment Canada
867 Lakeshore Rd., PO Box 5050
Burlington, ON, L7R 4A6
Tel.: (905) 336-4728
Fax: (905) 336-4699
Email: Tyler.Spencer@ec.gc.ca

Nadia Marenco
150 Farnham Ave. #208
Toronto, Ontario
M4V 1H5
Tel.: (416) 925-0702
Email: nmaren@hotmail.com

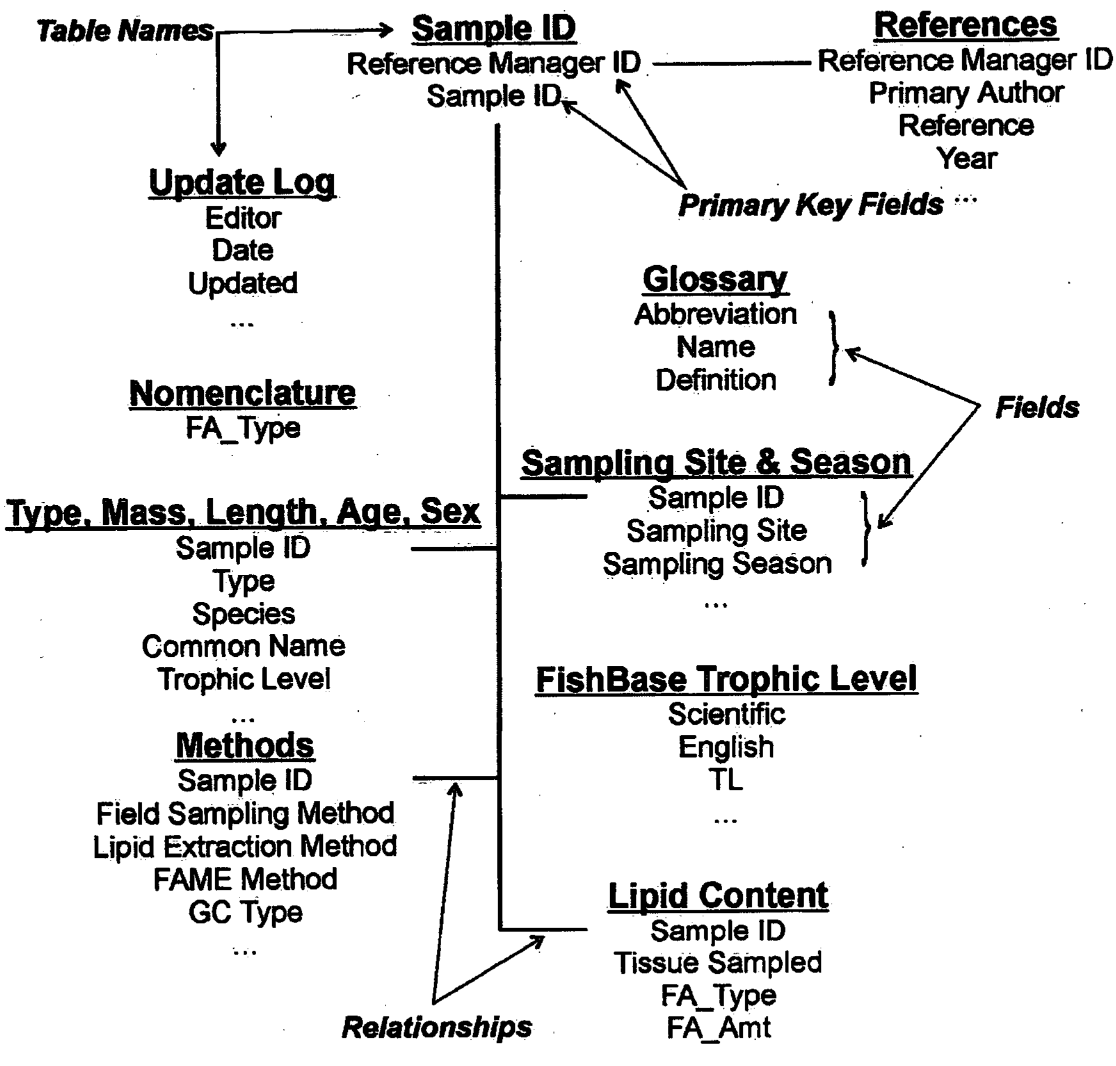

Figure 1. Schematic diagram illustrating the relationships between the "LipidBase.mdb" database tables.

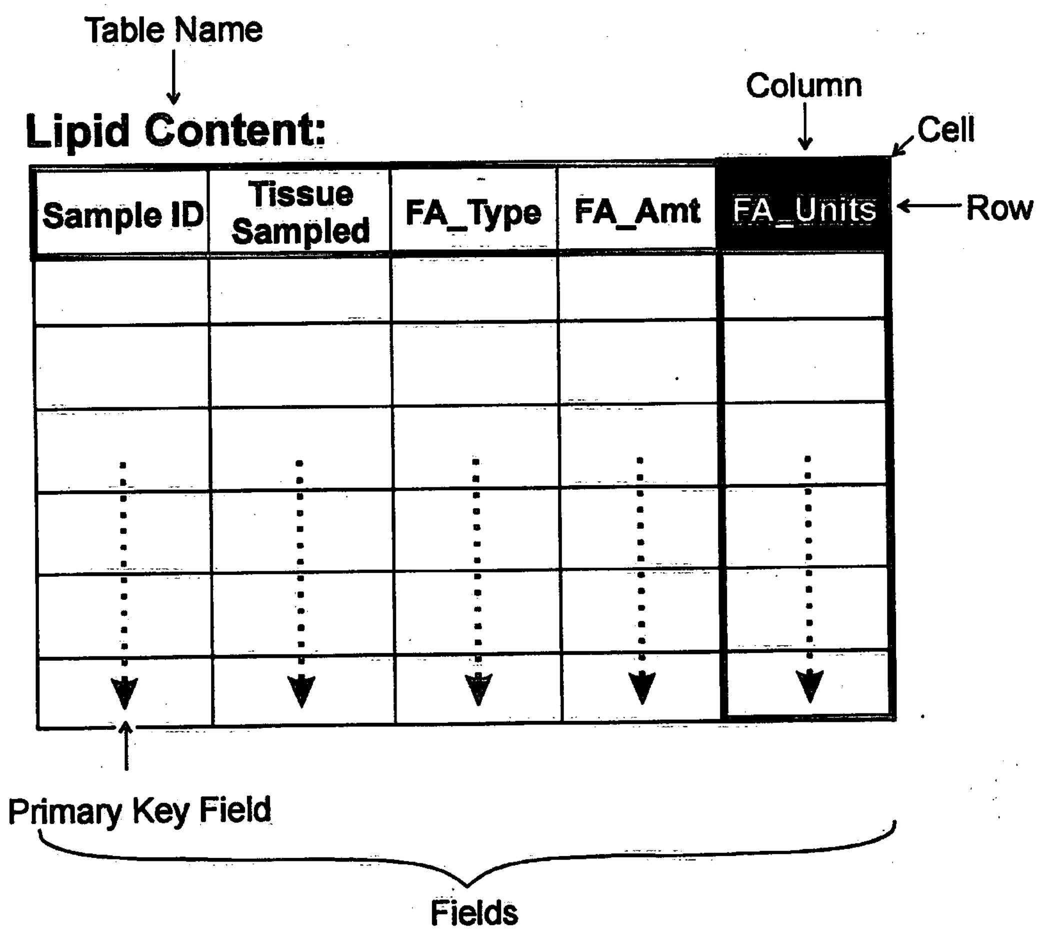

Figure 2. Schematic diagram illustrating the orthogonal format of a "LipidBase.mdb" database table.

PRINTED IN CANADA
IMPRIMÉ AU CANADA
ON RECYCLED PAPER
SUR DU PAPIER RECYCLÉ

National Water Research Institute
Environment Canada
Canada Centre for Inland Waters
P.O. Box 5050
867 Lakeshore Road
Burlington, Ontario
L7R 4A6 Canada

National Hydrology Research Centre
11 Innovation Boulevard
Saskatoon, Saskatchewan
S7N 3H5 Canada

NATIONAL WATER RESEARCH INSTITUTE
INSTITUT NATIONAL DE RECHERCHE SUR LES EAUX

Institut national de recherche sur les eaux
Environnement Canada
Centre canadien des eaux intérieures
Case postale 5050
867, chemin Lakeshore
Burlington, Ontario
L7R 4A6 Canada

Centre national de recherche en hydrologie
11, boul. Innovation
Saskatoon, Saskatchewan
S7N 3H5 Canada

Environment Canada Environnement Canada

Canada