

# A database of fossil occurrences in the Ordovician of Baltoscandia

ØYVIND HAMMER



Hammer, Ø. 2003–04–30: A database of fossil occurrences in the Ordovician of Baltoscandia. *Bulletin of the Geological Society of Denmark*, Vol. 50, pp. 5–9. Copenhagen. © 2003 by Geological Society of Denmark. ISSN 0011–6297.

A database recording published fossil occurrences in the Ordovician of Baltoscandia has been constructed and made available on the Internet (<http://asaphus.uio.no>). The database presently contains 10340 entries, each recording the first and last appearances of a single species at a single locality, together with other stratigraphical, taxonomic and geographical information. The data are taken from 141 publications, most of them from 1980 or later but also included are some older key references. The database will be continuously maintained, revised and expanded. In addition to serving as a reference for researchers working on the Ordovician of Baltoscandia, the database has been used to produce biodiversity curves for different fossil groups. These curves reflect the global increase in biodiversity through the Ordovician, but also show interesting local fluctuations.

*Key words:* Ordovician, Baltoscandia, Baltica, database, diversity.

Øyvind Hammer [[ohammer@nhm.uio.no](mailto:ohammer@nhm.uio.no)], Geological Museum, Boks 1172 Blindern, N-0318 Oslo, Norway. 15 June 2001.

The Baltoscandian region has yielded most of the Ordovician fossil specimens known from the palaeo-continent of Baltica. Fossils have been collected in the region for more than two hundred years, and detailed bed-by-bed sampling has been carried out at many localities in particular during the last twenty years. This database project aims to make these data available on a web site, allowing researchers to search for published occurrences of a specified taxon, or occurrences within a geographical region or stratigraphical unit.

The fundamental unit of the database is the first and last appearance data (FAD and LAD) of a single taxon (usually species) at one locality as recorded by a single publication. The inclusion of abundance data would have been an advantage, in particular because it would have allowed more stringent comparisons of taxon counts across localities and stratigraphical levels, for example by rarefaction analysis (Adrain *et al.* 2000). Abundance data are however not presently included, partly due to the excessive work involved in entering such data into the computer and partly because most publications do not give such information.

The database is not limited to any particular fossil group. It contains records of macrofossils such as trilobites, brachiopods, graptolites, corals, cephalopods,

gastropods, bivalves, hyolithids, echinoderms, bryozoans, sponges and macroalgae, and microfossils such as conodonts, ostracodes, chitinozoans and acritarchs. Data have been taken from publications of fossil occurrences in Sweden, Norway, Estonia, western Russia, Lithuania, Latvia and Denmark (Bornholm).

FADs and LADs are entered into the database with the stratigraphical precision available in the publication, which ranges from very poor (only the formation is known) to very good (level in centimeters within a section). This allows a better stratigraphical resolution than that possible at stadial level, as is common in many diversity studies.

The database is openly available at the web site [asaphus.uio.no](http://asaphus.uio.no). It is searchable on the basis of taxonomy, geography or stratigraphy, and the web pages allow the on-line plotting of range charts.

## Database structure and technical implementation

The relational database consists of five tables: Occurrences, literature, tie points and apparent absolute ages of tie points and formations. Central to the database is the table of occurrences, containing records of

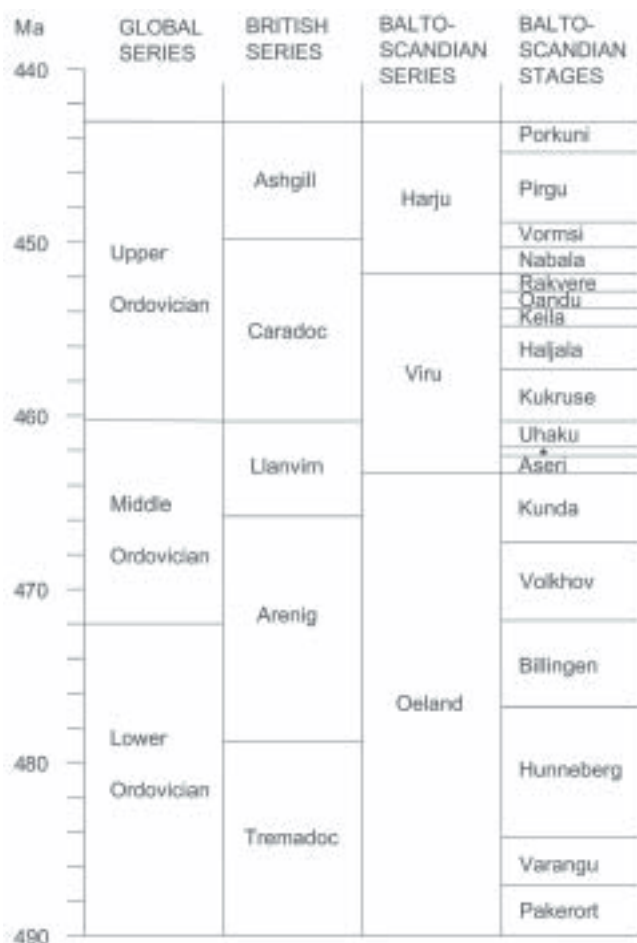


Fig. 1. Main stratigraphical units and their apparent ages as used in the database. Modified from Männil & Meidla (1994), Cooper (1999) and Webby *et al.* (in press). \* = Lasnamägi Stage.

the first and last appearance of one species at a single locality according to a given publication. Each record includes fossil group (e.g. Trilobita, Brachiopoda), family, genus, subgenus, species, subspecies, species, author and year. Geographical location data include country, area and locality. First and last appearances are each specified with an indication of formation, member, bed, biozone and level in meters within the section, if available. Finally, the record contains author and year of the publication, date of entry into the database, and a text field for comments. In the 'comments' field information such as 'cf.' or 'aff.' may be entered. In cases where the taxon name in the database differs from the original name in the publication, due to synonymy changes, the original name is also given in the 'comments' field. In many publications, in particular the older ones, much of this information is not given, and the corresponding fields are left blank in the database. Sometimes the occurrence is only known at the level of formation, and the data

fields for member, bed, biozone and level in meters within section must then be left blank. All taxonomic, geographical and stratigraphical names are given in plain text. The common practice of using codes adds complexity, and involves extra work in looking up the codes when entering the data. Global replacements of names can easily be performed with a search and-replace operation rather than changing the name in a code/name table.

The main challenge when constructing such databases is that of taxonomy. The naming of species is in constant flux, and synonymy lists are often long and confusing. As far as possible, the occurrence records in the Baltoscandia database have been updated with the names in recent use, in some cases with the help of experts on the particular groups (see the Acknowledgments section). Still, it is unavoidable that a number of species will be registered under two or more synonymous names.

The occurrence records contain links to entries in a literature table, with full references to all publications. This table in itself constitutes a useful bibliography for the palaeontology of the Ordovician of Baltoscandia.

There are three tables providing stratigraphical information for correlation purposes. The tie point table contains the levels in meters of at least two tie points in each of the sections (localities) referenced in the table of occurrences. These tie points (often biozone boundaries) have names that provide links to correlated levels in other sections. When other markers are not given within the section, lithological boundaries are used as tie points, giving possible errors caused by diachronous units. Levels of the bottom and top of the studied sections are also given. The table of apparent absolute ages of tie points contains approximate absolute ages, in millions of years before present, of all tie points. This table is used for the scaling of range charts and for the calculation of diversity curves. The age model and correlation chart as given in Figure 1 is based on Männil & Meidla (1994) and Cooper (1999) with later revisions (Webby *et al.* in press). A similar table exists for the approximate absolute ages and durations of formations and members.

Range charts are produced dynamically on the web pages using a simple correlation scheme. The FAD and LAD ages are estimated by linear interpolation between the absolute ages of the tie points bracketing the occurrence in the section. When the FAD or LAD is below the lowest or above the highest tie point in the section, the age is similarly estimated by extrapolation. These procedures assume constant sedimentation rates, which is of course inaccurate in many cases, but still allow a higher resolution than provided

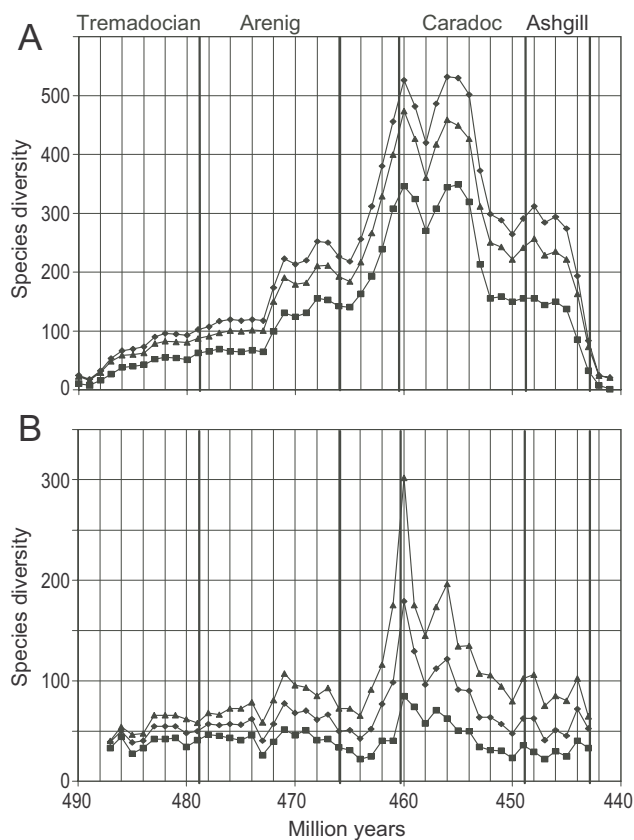


Fig. 2. Range-through species diversity for all groups, calculated for 1-million year intervals (see text). A: Original diversity (top curve, diamonds) and upper (triangles) and lower (squares) limits of 95% confidence interval from 1000 replicates with bootstrapped localities. B: Rarefaction to 19 quasi-samples (localities) within each 1-million year interval. Mean (middle curve, diamonds) and upper and lower limits of 95% confidence interval for 1000 replicates.

by the tie points themselves. At least, this method preserves the rank order of datums within one section. In cases where the level in meters within the section is not known, the FAD and LAD ages are estimated as the absolute ages of the bottom and top of the formation or member, probably inflating the real range. Attempts are made to keep the stratigraphical tables updated according to new publications, as these tables are vital for the production of diversity curves and range charts.

The database is running on an SGI O2 (UNIX work-

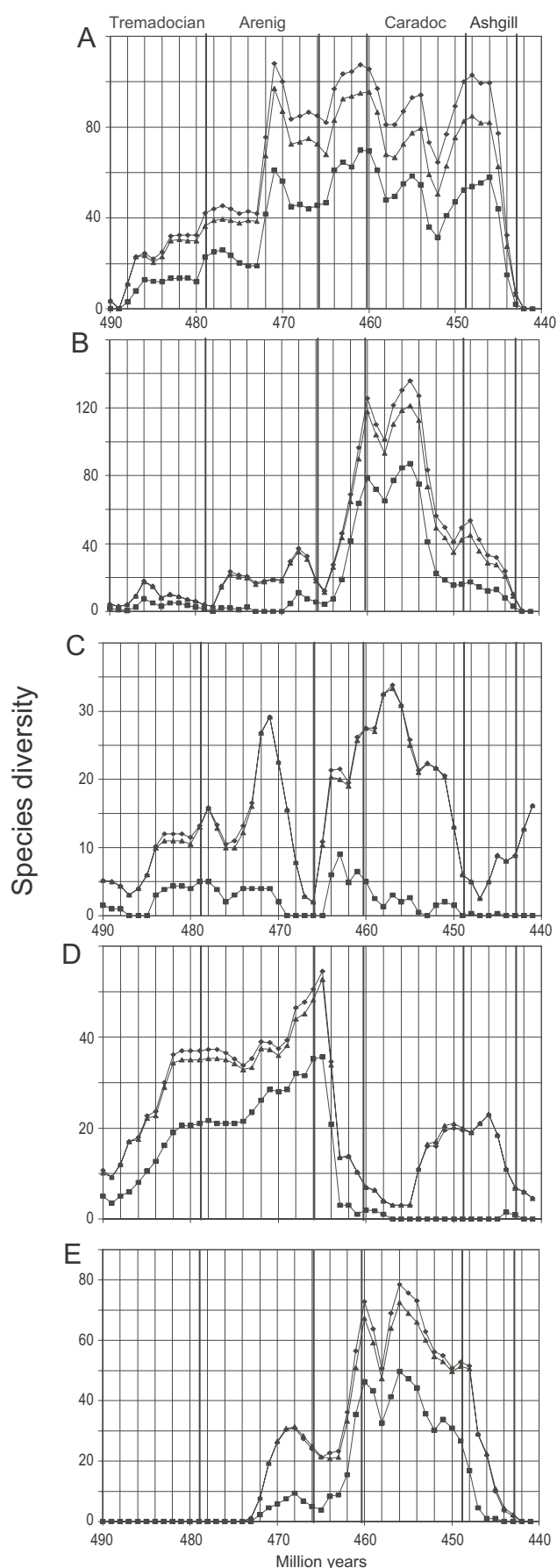


Fig. 3. Range-through species diversity curves for selected groups, calculated for 1-million year intervals (see text). Top curve (diamonds) in each plot is the original diversity. Lower curves (triangles and squares) represent upper and lower limits of 95% confidence intervals from bootstrapping of localities. A: Trilobites. B: Brachiopods. C: Graptolites. D: Conodonts. E: Ostracodes.

station). The Apache web server includes an interpreter for 'PHP', a server-side scripting language. The generation of dynamical web pages, including interfacing to a database server running MySQL, is programmed in PHP. All the software (Apache, PHP and MySQL) is freely available under the UNIX, Linux and Windows operating systems.

## Biodiversity curves

The diversity curves produced from the database (Figs. 2 and 3) have been described elsewhere (Hammer in press), and only a brief presentation will be given here. The curves were calculated using the range-through assumption, meaning that continuous presence of a species from FAD to LAD is assumed. The number of species is counted within consecutive one-million-year time slices, such that a species ranging through a time slice counts as one unit, while a species with FAD or LAD within the time slice counts a half unit (Foote 2000).

Two types of randomization test were executed in order to estimate confidence intervals for diversities. First, a bootstrap-type analysis was carried out, where 1000 resampled replicates of the original database were produced, each with randomly selected localities (with replacement). Upper and lower 5% percentiles for the diversity count within each time slice were calculated from the set of resampled replicates. Second, a rarefaction procedure was used, where each time slice was represented by a random selection of localities covering that time slice, selected without replacement. For the total database, the smallest number of localities found in any time slice is 19, and this was chosen as the standardized locality count for rarefaction. These tests indicate that the biota as a whole and some groups such as trilobites are well and evenly sampled, while the curves for other groups may be seriously degraded by lack of data. Such problems seem to be particularly serious for graptolites, for brachiopods in the Lower Ordovician and for conodonts in the Upper Ordovician (Fig. 3).

The Ordovician radiation is even more pronounced in the Baltoscandian curves than in the global curves of Sepkoski (1995) and others. This may be at least partly due to the drift of the Baltica plate from high to low latitudes through the Ordovician (Torsvik *et al.* 1992). As an example, trilobite biodiversity is apparently falling through the Ordovician in many regions of the world, while Baltica and Avalonia display increasing trilobite diversities (Adrain *et al.* in press).

The curves from Norway and Sweden (not shown)

Table 1: Number of entries (FAD and LAD at one location) in the database, for different fossil groups and countries.

|              | Norway | Sweden | Estonia | Other | Total |
|--------------|--------|--------|---------|-------|-------|
| Trilobites   | 608    | 734    | 977     | 372   | 2691  |
| Brachiopods  | 257    | 277    | 1238    | 116   | 1888  |
| Ostracods    | 57     | 125    | 521     | 933   | 1636  |
| Conodonts    | 201    | 966    | 72      | 0     | 1239  |
| Bryozoans    | 25     | 57     | 735     | 16    | 833   |
| Molluscs     | 77     | 90     | 351     | 0     | 518   |
| Chitinozoans | 102    | 272    | 0       | 14    | 388   |
| Graptolites  | 103    | 222    | 41      | 63    | 429   |
| Other        | 164    | 161    | 386     | 7     | 718   |
| Total        | 1594   | 2904   | 4321    | 1521  | 10340 |

appear to correlate with sea-level change, with high diversities at low stands. Whether this is due to primary differences in diversity or to shifts in preservation potential caused by the appearance of carbonate facies at low stands is so far unclear. Furthermore, these diversity 'cycles' appear to have a periodicity of about 14 million years, with four main peaks throughout the Ordovician.

The problems and pitfalls connected with the production of diversity curves from taxon counts are well known, and can at best be minimized, not eliminated. These problems include imperfect taxonomy (double counting of synonyms), incomplete data due to differential preservation and collection, imperfect stratigraphical correlation and technical problems such as edge effects and the influence of turnover rates on counts within intervals (Foote 2000). Benton (1999) has discussed these and other problems, but nevertheless defends the production of diversity curves.

## Conclusion

The database is continuously being updated and extended. In July 2002 it contained 10340 occurrence records taken from 141 publications. This included 3656 species or subspecies (cf. and aff. are not counted as separate species). Table 1 shows the number of recorded occurrences in different countries and fossil groups.

It is hoped that this database will facilitate rapid information retrieval by researchers working on the Ordovician of Baltoscandia. The data have also been used for the construction of diversity curves through the period, and can potentially be used for the quantitative assessment of palaeobiogeographical aspects such as region provincialism and biogeographical gradients. Finally, the database is suitable for quanti