PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Diatom Quality Control and Data Handling

M. A. R. Munro, A. M. Kreiser, R. W. Battarbee, S. Juggins, A. C. Stevenson, D. S. Anderson, N. J. Anderson, F. Berge, H. J. B. Birks, R. B. Davis, R. J. Flower, S. C. Fritz, E. Y. Haworth, V. J. Jones, J. C. Kingston and I. Renberg

Phil. Trans. R. Soc. Lond. B 1990 **327**, 257-261
doi: 10.1098/rstb.1990.0061

Phil. Trans. R. Soc. Lond. B **327**, 257–261 (1990)
Printed in Great Britain

Diatom quality control and data handling

By M. A. R. Munro[1], A. M. Kreiser[1], R. W. Battarbee[1], S. Juggins[1], A. C. Stevenson[2], D. S. Anderson[3], N. J. Anderson[4], F. Berge[5], H. J. B. Birks[6], R. B. Davis[3], R. J. Flower[1], S. C. Fritz[7], E. Y. Haworth[8], V. J. Jones[1], J. C. Kingston[9] and I. Renberg[4]

[1] *Palaeoecology Research Unit, University College London, 26 Bedford Way, London WC1H 0AP, U.K.*
[2] *Department of Geography, University of Newcastle upon Tyne, Newcastle NE1 7RU, U.K.*
[3] *Department of Botany, Deering Hall, University of Maine, Orono, Maine 04469, U.S.A.*
[4] *Department of Ecological Botany, University of Umeå, S-901 87 Umeå, Sweden*
[5] 2920 *Leira, Valdres, Norway*
[6] *Botanical Institute, University of Bergen, Allégaten 41, N-5007 Bergen, Norway*
[7] *Limnological Research Center, University of Minnesota, Minneapolis, Minnesota 55455, U.S.A.*
[8] *Institute of Freshwater Ecology, Ambleside, Cumbria LA22 0LP, U.K.*
[9] *Department of Biology, Queen's University, Kingston, Ontario, Canada*

The diatom data used for reconstructing pH within the Surface Water Acidification Project (SWAP) came from several different laboratories. The laboratories used agreed nomenclature and standardized identifications by using quality control techniques. A diatom database (DISCO) stored and processed counts and site information.

1. Introduction

The diatom section of the Surface Water Acidification Project (SWAP) Palaeolimnology Programme brought together data from many sources. One of the aims of the project was to construct a single large pH calibration dataset, by combining modern lake pH values with modern surface-sediment diatom assemblages and to use the calibration to reconstruct past lake pH values from the assemblages in sediment cores. However, reconstructions based on small regional subsets of chemical and diatom data from several laboratories must be combined in a way that resolves any differences in taxonomy or analytical technique, and selected extracts from this large combined dataset must be created in a form that can be read by the computer programs performing the calibration and reconstruction.

2. Standardization of diatom taxonomy

Because there is considerable variation in practice between diatomists in different laboratories it is essential in cooperative projects to establish agreed protocols for diatom taxonomy and nomenclature. In the SWAP project, diatomists from Norway, Sweden and the U.K. adopted an approach to diatom harmonization similar to that developed by diatomists involved in the Paleoecological Investigation of Recent Lake Acidification (PIRLA) project in the U.S.A. (Charles *et al.* 1987). This has included the use of taxonomic workshops, diatom slide exchange and the circulation of agreed taxonomic protocols both within SWAP and between the SWAP and PIRLA schemes.

(*a*) *Initial identification of problem areas*

The first SWAP diatom taxonomy workshop was held in March 1987. In preparation for this, each of the four laboratories involved circulated diatom slides and accompanying count sheets for three sediment samples to each of the other laboratories, choosing samples representing the range of soft-water floras encountered within SWAP. Each laboratory provided counts from all 12 slides and the results were compared at the taxonomy workshop. Figure 1*a* summarizes the results of one of these slides (Lingmoor Tarn), demonstrating the three main problems encountered: differences in nomenclature, splitting versus amalgamation of taxa and differing criteria used in the identification of a taxon. Figure 1*b* shows the result after revision at the workshop.

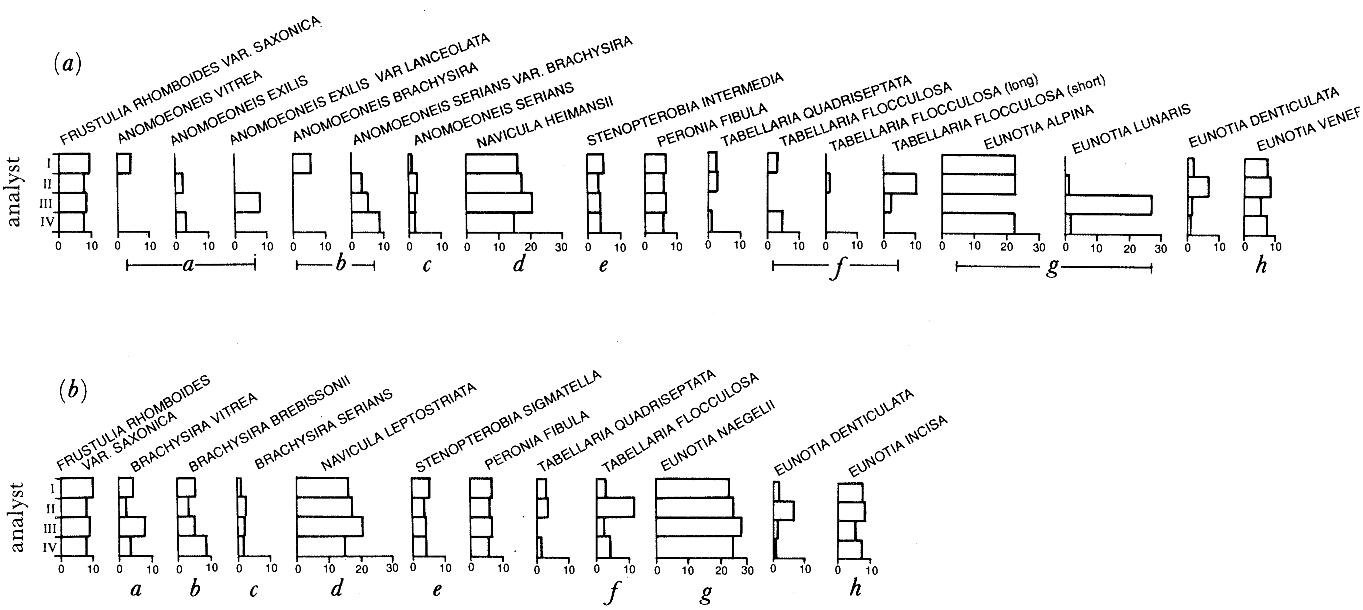

FIGURE 1. (*a*) Dominant taxa in the Lingmoor Tarn slide illustrating problems of nomenclature in groups *a*, *b*, *c*, *d*, *e*, *g*, *h*, problems of splitting versus amalgamation in groups *a*, *f* and the use of differing identification criteria in group *g*; (*b*) dominant taxa in the Lingmoor Tarn slide after full taxonomic and nomenclatural revision. (Horizontal scale is percentage occurrence.)

The agreements from the workshop were circulated to all the workshop participants in a SWAP taxonomic guide. Nomenclature was based on the Checklist of British diatoms (Hartley 1986), which formed the framework for a coded checklist of British diatoms (Williams *et al.* 1988). The SWAP taxonomic guide also included agreements on definitions for boundaries between certain species and their varieties, in some cases following inspection of type material. If possible, published descriptions were referred to. Failing this, the criteria for identification were agreed between the participants.

(*b*) *Applying and refining the SWAP taxonomic guide*

To test the 1987 workshop protocols three slides from lake sediment samples representing the range of pH values encountered within SWAP (one sample from each of the ranges pH < 5, pH 5–6 and pH > 6) were circulated to all SWAP diatomists, who counted the slides without prior access to the countsheets of other diatomists. The results were discussed at a workshop in July 1988.

Many potential problems had been avoided by following the protocols agreed at the previous workshop. However, because of the inclusion of taxa additional to those already encountered, some further problems were raised. Most of these were resolved and a revised edition of the SWAP taxonomic guide was circulated to all diatomists. This workshop also produced a guide for the handling of unknown diatoms.

A major problem encountered at the 1988 workshop was one of identification within the genus *Aulacoseira*. This genus presents particular taxonomic problems because of the difficulties of matching girdle (side) views of the diatom valves with valve (front) views.

(*c*) *Focusing on problems within the genus* Aulacoseira

As a result of the 1988 workshop it was decided to concentrate on the taxonomy of the *Aulacoseira* genus. Three samples containing *Aulacoseira* were selected from SWAP sites and slides were circulated to all SWAP diatomists, who counted 300 valves of *Aulacoseira* from each slide and discussed the results at the diatom workshop in February 1989. There were no problems with the most abundant *Aulacoseira* taxa, such as *A. lirata* and *A. distans* var. *nivalis*. For the less abundant taxa, e.g. *A. lirata* var. *alpigena*, *A. subarctica* and *A. subborealis*, definitions were agreed by using either published definitions or criteria agreed at the workshop. Previously counted surface and core samples containing the problematic *Aulacoseira* taxa were consequently recounted.

This series of workshops and quality control exercises has enabled all diatom data generated within SWAP to be compatible between laboratories and to be suitable for storage and manipulation in a computerized database.

3. The computer database

The diatom database (DISCO) at University College London combines archives of diatom counts, chemical analyses and catchment descriptions for several projects, including SWAP. It uses the commercial program ORACLE and the standard database language SQL. Diatom counts, taxonomic information and chemistry can be entered on or retrieved from the database, by filling out entries on forms that appear on the computer screen. The PIRLA project also required a large computer database (Ahmad & Charles 1988). It also uses a commercial program (SIR) and stores data from many other sources (including chemical analyses and chrysophyte counts). However most of the PIRLA database is hierarchic, with the diatom counts at a lower level in the hierarchy than the site information, whereas DISCO is a relational database, consisting of tables of data with no pre-established structure. The data themselves are used to connect the different tables. Figure 2 shows how common site and sample codes can link different tables to establish a structure similar to the PIRLA database. Many other arrangements are possible.

The database includes the version of the Hartley (1986) diatom checklist coded by Williams *et al.* (1988). Williams' coding scheme allocates a code to each taxon on the list by using a number to represent the original name of the taxon. This does not encode any information about the genus or species name, so the codes are suited to following a taxon through revisions of nomenclature. Some names do not fit this scheme: names of valid taxa that have not been

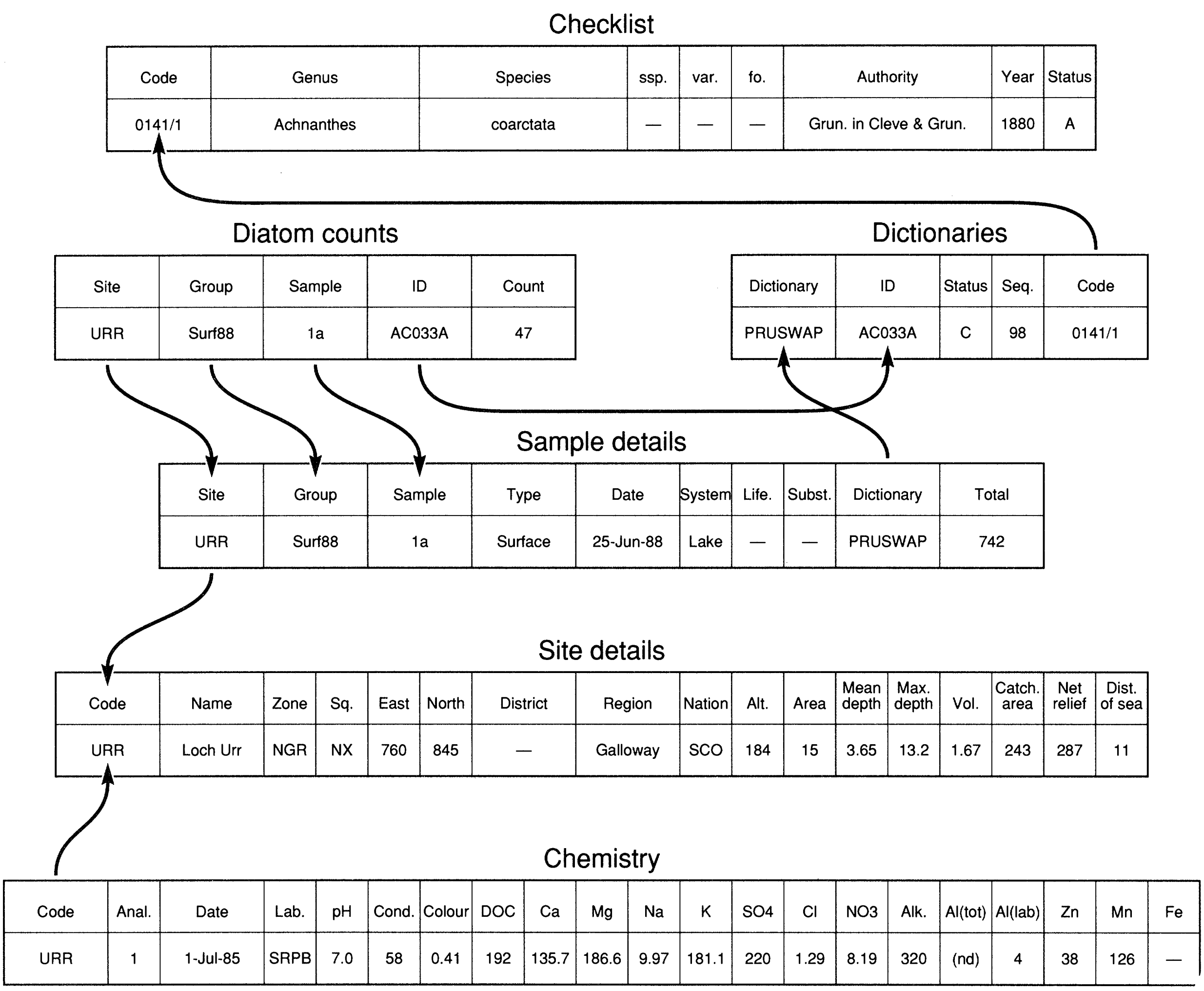

FIGURE 2. The rectangles represent entries in various database tables, the arrows show how common entries can connect the tables. Other connections are possible.

recorded in Britain (so Hartley excluded them), names that refer to aggregate taxonomic categories that do not have any place in a checklist of single taxa (such as *Navicula* spp.) and names needed for newly discovered taxa, or taxa that diatomists can recognize consistently but cannot identify as any checklist entry. It includes these in a supplementary code list. The database also stores tables of chemical analyses for calibration data set values that were used for improving reconstruction methods (Birks *et al.*, this symposium). So far only pH is used, but other data are included, such as: conductivity, colour, dissolved organic carbon, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, SO_4^{2-}, Cl^{-}, NO_3^{-}, alkalinity, Al (total), Al (labile), Zn, Mn, Fe, Cu, Pb, dissolved SiO_2 and total P. These will be used for future development of diatom-based reconstruction methods.

4. Data processing

Diatomists submitting counts to the database provided an outline code dictionary for their data. The dictionary was checked to ensure that it contained valid checklist or supplementary list codes, and was compared to previous lists from the same laboratory. The counts were then converted to percentages of the total number of diatoms counted in each sample; any amalgamations and other re-definitions of taxa were performed. The database also provided a list of the more frequently occurring taxa for each group of samples, defined as those present in at least two of the samples and having a frequency of more than 1% in at least one sample. These lists were compared to help identify unresolved taxonomic problems.

The SWAP calibration data-set was created by merging full sets of surface sediment percentages in this way for the regional datasets from Scotland, Wales, Cumbria, Norway and Sweden. All the aggregates to genus level and above (e.g. *Navicula* spp.) were deleted and the remainder of this list was used to select the taxa to be exported to the calibration programs. The core percentages were then exported to the reconstruction programs as a series of individual site datasets.

The chemical values added to the database were screened in several stages. The program for adding new values included simple checks on the range of the values for various determinands. At least two people checked a complete paper listing of the database values against the original laboratory report. The values were standardized (zero mean, standard deviation = 1), by using both site and regional means, and values outside the range ± 2 were noted. In some cases histograms of the distributions were plotted as a further check, as the values for some determinands were not normally distributed. Values of pH were excluded from the data-set only if they were extremely discordant (such as a value of 3.6 from Llyn Bugeilyn whose other pH values were 5.0, 4.6 and 5.1), or if exceptional conductivity and ionic concentration values suggested that they were influenced by a sea-salt event. Geometric mean pH values were used for calibration (arithmetic means of H^+ concentrations). The groups of analyses used to calculate the means were matched with the calibration dataset to ensure that re-sampled sites (which had two sets of diatom assemblages) had two sets of pH values. The sampling date for the chemistry was normally within three years of the diatom sampling date.

The database will allow further exploration of diatom–water-quality data as samples from a wider range of environmental gradients are included and the full range of chemical data is screened in the same way as the pH values.

References

Ahmad, H. & Charles, D. F. 1988 *PIRLA data base management system user's manual* (2nd ed.). Paleoecological Investigation of Recent Lake Acidification (PIRLA), report no. 32, 266 pp.

Charles, D. F., Whitehead, D. R., Anderson, D. S., Bienert, R., Camburn, K. E., Cook, R. B., Crisman, T. L., Davis, R. B., Ford, J., Fry, B. D., Hites, R. A., Kahl, J. S., Kingston, J. C., Kreis, R. G., Mitchell, M. J., Norton, S. A., Roll, L. A., Smol, J. P., Sweets, P. R., Uutala, A. J., White, J. R., Whiting, M. C. & Wise, R. J. 1987 The PIRLA project (Paleoecological Investigation of Recent Lake Acidification): preliminary results from the Adirondacks, New England, N. Great Lakes and N. Florida. *Wat. Air Soil Pollut.* **31**, 355–366.

Hartley, B. 1986 A check-list of the freshwater, brackish and marine diatoms of the British Isles and adjoining coastal waters. *J. mar. biol. Ass. U.K.* **66**, 531–610.

Williams, D. M., Hartley, B., Ross, R., Munro, M. A. R., Juggins, S. & Battarbee, R. W. 1988 *A coded checklist of British diatoms*. London: ENSIS Publishing.