BIOINFORMATICS **APPLICATIONS NOTE** Vol. 14 no. 9 1998
Pages 827–828

An Internet linkage and mutation database for the complex phenotype asthma

Matthias Wjst and Thomas Immervoll

GSF National Research Center for Environment and Health, P.O.B. 1129,
D-85758 Oberschleissheim, Germany

Received on June 18, 1998; revised on July 17, 1998; accepted on August 19, 1998

Abstract

Summary: *The paper presents details of database construction, website installation and server architecture of the asthma and allergy gene database.*
Availability: *Database and server templates are available on request from the first author.*
Supplementary Information: *The URL of the asthma and allergy gene database is http://cooke.gsf.de*
Contact: *Matthias Wjst, wjst@gsf.de*

Asthma affects up to 10% of the population in Western countries (Wjst *et al*., 1997). There is no doubt that a major genetic factor exists, however, the inheritance does not follow a simple Mendelian pattern (‘complex phenotype’). This has given rise to many different linkage and candidate gene studies (Sandford, 1996). It is becoming more and more difficult to get an overview on this wealth of information, presented at many meetings, and in many different journals. Even major databases like GDB or OMIM contain only part of the relevant information, or are unable to summarize recent results. As this problems also emerges in other fields of genetic research, several disease specific databases like that for coagulation factor IX, cystic fibrosis, PKU and PAH have been constructed (Cotton, 1996). They follow more or less some common standards (Ad Hoc Committee, 1996). For complex human diseases like atherosclerosis, myocardial infarction, diabetes, schizophrenia or asthma, however, no public databases are available despite intensive research for more than 10 years. We therefore summarize our experience with the construction of the asthma and allergy gene database.

Server architecture

The aim of the software development was a precise, fast, accurate, stable and secure system. Because a first prototype of a relational database with asthma linkage data (Genahunt 1.1) was already implemented under MS-Access 7.0, this database was moved to an Intel Pentium-based machine with 128 Mb memory and a 3 GB hard drive, including a UPS power management system. The Windows NT 4.0 operating system was installed, as well as Netscape Fasttrack 2.1 software. Fasttrack is low-cost server software that showed high throughput values in recent server tests. The local network is connected through a firewall to the Internet and uses the standard TCP-IP protocol.

Cold Fusion 3.0 (CF, Allaire Corp. http://www.allaire.com) was used as an interface between server and database (Forta and Drucker, 1997). The first http requests a query page from the server. If the user has filled in the form, a small Javascript program checks to see if all data are entered correctly, and the page is then sent back to the server. Response pages containing only HTML formatted text can be transmitted without modification back to the client. If the server recognizes a CF template, however, it will send the response page to CF first, where the CF meta language (CFML) tags are processed and translated back into HTML. Special CFML tags allow SQL queries to be defined for retrieval of data from ODBC compliant data sources. The final result is then transmitted back to the client. For the additional figures, small Java programs (‘applets’) are transmitted to the client (JavaChart, Visual Engineering http://www.ve.com). Together with the accompanying data the client builds up the figures on the local computer. Both client-side data verification and remote rendering of figures offer the advantage of a reduced server load. However, because of multiple errors reported with the Java engine in the test phase, the use of applets has been temporarily disabled. Figures are created at the moment at the server side in GIF format using a publicly available C++ routine (Fly 1.4.2 available from Martin Gleeson at http://www.unimelb.edu.au/fly/fly.html).

To guarantee data security the latest patches of the software ventures were installed and audit trails enabled for file accesses and system functions. Any user profiles, however, were avoided to ensure privacy of information retrieval. For data safety, backups are run biweekly after checking for computer viruses and consistency of the database tables. There are always mirror copies of the actual database that are not available to the public. Data and database design are protected by the European Data Protection Directive of October 24, 1998.

Website installation

Recommendations for constructing medical websites have been included (International Committee of Medical Journal

Editors, 1997). For the page layout, standard HTML 3.0 syntax was used to support most of the popular browsers. Users are required to register at the database by giving their name and address on their first visit. If the registration is complete, a password is sent by email to the user. Unregistered users can see only parts of the website. This procedure has been chosen because data submission or data annotation will require an exact sender identification. Other menu items are a member registry, and a mailing list where messages can be mailed to all subscribed users. The main linkage and mutation database is supplemented by on-line help for data retrieval and data submission.

Database construction

The main database relies on a relational structure. Custom data types, primary and foreign keys, index fields and relations were all planned to avoid redundant data structure and repeated entries and to anticipate further growth (Bryant, 1998). Multi-user database accesses also require extensive control structures to maintain consistency of the database. Logfiles for data entries help to trace incorrect or broken links. All data submissions are done after exclusive row-level locking including a roll-back procedure if the transaction fails for any reason.

The 'parent' table is the study table. It relates in a 1:*n* mode to the linkage and 1:1 to the mutation and *n*:1 to the population, phenotype and reference table. The term 'study' is therefore defined in this context as the examination of at least one individual (in most cases, however, one ethnic group) for a single phenotype for one or more linked markers or one single mutation. It appears that most of the research papers have to be broken into different studies before they can be entered in the database. The first genome-wide search for asthma genes (Daniels *et al*., 1997), for example, resulted in eight different studies. If the authors use different statistical techniques for analysis of the same data only the results with most significant values are entered to avoid any conflict with redundant data.

A further principle of the database construction was parsimony. Only data relevant to the main topic was collected. Information not stored in the database could be easily retrieved by hyperlinks to external web databases. As a new feature 'self-structuring' database entry forms were developed. For that purpose the entry form is constructed after querying the database for already existing field contents. This facilitates complex meta-analysis later, since tables will contain only distinct entries of a controlled vocabulary. Another advantage is the reduction of typing errors, which is especially important with marker and mutation symbols.

User interface

The Java console for the figures requires Netscape Navigator release 3.0 or higher, or Internet Assistant 3.0. The options in the browser client have to be set *on* for both Java and Javascript. The address, http://cooke.gsf.de, then has to be typed in the location window on the top of the browser. A menu appearing on the left side is available during the whole session, and helps to navigate through the website.

Clicking on a hyperlink starts in most cases an (invisible) database query. The resulting pages are generated 'on the fly', meaning that the result of a query always represents the actual state of the database. Many of the reports are constructed as 'drill-down' applications: The result lists are presented with hyperlinks that can start another database query. The user can therefore start with a query of single chromosomal region, select a specific study, look at that phenotype, look where that phenotype has also been used and „walk' around the whole data. More detailed information not found in the local database can be retrieved by links to either the Genome Database GDB (Baltimore) or to the PubMed Database at the US National Library of Medicine (Bethesda). At present 88 linkage studies with 7164 linkage positions and 72 mutations in 24 genes are contained in the database.

Acknowledgements

We wish to thank S. Emfinger for correcting the manuscript, G. Fischer, B. Forta and D. McMurdie for their technical support, R. Cotton for pointing us to important references, N.Morton for encouraging us to start this service and many more colleagues, in particular F. Kaufmann, who has given us helpful comments on the website.

References

Ad Hoc Committee on Mutation Nomenclature (1996) Update on nomenclature for human gene mutations. *Hum Mut*., **8**, 197–202.

Bryant,S.P. (1998) Managing pedigree and genotype data. In Bishop,M.J. (ed.), *Guide to Human Genome Computing*. Academic Press, San Diego, pp. 49–74.

Cotton,R.G.H. (1996) Locus specific mutation databases. In Landegren,U. (ed.), *Laboratory Protocols for Mutation Detection*. Oxford University Press, Oxford, pp. 5–7.

Daniels,S.E., Bhattacharrya,S., James,A., Leaves,N.I., Young,A., Hill,M.R., Faux,J.A., Ryan,G.F., Le Söuef,P.N., Lathrop,G.M., Musk,A.W. and Cookson,W.O.C.M (1996) A genome-wide search for quantitative trait loci underlying asthma. *Nature*, **383**, 247–250.

Forta,B. and Drucker,S. (1997) *The Cold Fusion Web Database Construction Kit*. Que Publisher, Indianapolis, pp. 31–34.

International Committee of Medical Journal Editors (1997) Policies for posting biomedical journal information on the Internet (editorial). *JAMA*, **277**, 1808.

Sandford,A., Weir,T. and Paré.P. (1996) The genetics of asthma. *Am. J. Res. Crit. Care Med*., **153**, 1749–1765.

Wjst,M., Wassmer,G. and Seuchter,S. for the ECRHS Group (1997) Genes for asthma? A pooled analysis of the European community respiratory health survey. *Am. Rev. Resp. Crit. Care Med*., **156**, 1773–1780