The dysmorphic human-mouse homology database (DHMHD): an interactive World-Wide Web resource for gene mapping

C D Evans, A G Searle, A A Schinzel, R M Winter

Abstract

Genetic mapping and the examination of “candidate genes” for isolating loci associated with clinical syndromes can be greatly accelerated if there is information about where in the genome a particular locus might be situated. Such clues can come from homology to mouse mutants that have been mapped and knowledge of homology between mouse and human chromosomal segments. Further clues can come from chromosome aberrations giving a similar phenotype. However, these clues are often scattered widely in published reports, and even if they are collected together in catalogues or databases there is no rapid way of moving from one data type to another. The Dysmorphic Human and Mouse Homology Database (DHMHD) is designed to ease this data transition. DHMHD comprises detailed information from four separate sources and enables cross referencing through phenotypic and chromosome homology. The DHMHD system is a prototype which is now available online through the World-Wide Web.

(1 Med Genet 1996;33:289-294)

Key words: dysmorphic syndrome; cytogenetic aberration; mouse mutant syndrome; World-Wide Web.

1. Introduction

Recent advances in Internet software tools such as Gopher, WAIS (Wide Area Information Server), and the World-Wide Web (WWW) have encouraged a growth in online resources dedicated to medical genetics. For example, there are now a number of WWW resources linked to the Human Genome Mapping Project (HGMP), including Online Mendelian Inheritance in Man, or OMIM, a computerised format of McKusick’s catalogue, which was first available as a dial in system in the mid-eighties, and numerous human and mammalian genome databases such as the Genome Database (GDB) and the Mouse Genome Database (MGD). These systems permit remote database queries through a hypertext multimedia interface and allow chromosome images (ideograms) to be downloaded and displayed on a personal computer. This report describes a new World-Wide Web resource available through the Medical Research Council HGMP resource centre: The Dysmorphic Human and Mouse Homology Database (DHMHD). This application consists of three separate databases of human and mouse malformation syndromes together with a database of mouse/human homologous regions. The mouse and human malformation databases are linked together through the chromosome homology database. The purpose of the system is to allow retrieval of syndromes according to detailed phenotypic descriptions and to be able to carry out homology searches for the purpose of gene mapping.

2. Background

Dysmorphology is a field of medicine which has as one of its concerns the diagnosis of children born with multiple malformations or variations in quantitative characteristics (for example, head size) outside the normal range. A pattern of signs recognised as occurring together and thought to be pathogenetically related is collectively called a dysmorphic syndrome or malformation syndrome. Approximately 8 in 1000 children are born with multiple malformations, and in about half of these infants the condition is likely to have a strong genetic component. Most genetic conditions leading to abnormal physical development have not been mapped and the genes responsible have not been identified. All of these genes would be expected to play an important role in normal development. Thus, the elucidation of the genetic malformation syndromes will lead to greater understanding of normal embryological development, as well as explaining and preventing congenital abnormalities.

Genetic mapping is often the first step in the isolation of genes causing clinical syndromes. Examination of “candidate genes” for mutations is also becoming an important strategy. These methods are greatly accelerated if there is information about where in the genome a particular locus might be situated. Such clues can come from homology to mouse mutants that have been mapped and from knowledge of homology between mouse and human chromosomal segments. Further clues come from chromosome aberrations (that is, structural changes causing areas of the chromosome to be deleted, duplicated, or translocated) giving a similar phenotype. Finally, knowledge of the function, expression, and map location of “developmental” genes might be used to predict the effect of mutations of these genes. These
clues are often widely scattered in published reports, and even if they are collected together in catalogues or databases there is no rapid way of moving from one data type to another. This is in essence the purpose of DHMHD, which aims to provide a means of linking several databases through phenotypic and chromosome homology. Searching can be performed in a number of different ways. (A) By phenotype, by specifying a combination of malformations and asking for all mouse and human syndromes with this combination. (B) By chromosome homology, by asking for human syndromes located at a chromosome region homologous with a specific mouse chromosome region (and vice versa from human to mouse). (C) By phenotype and chromosome homology, by asking for all mouse or human syndromes with a specific combination of malformations mapping to homologous chromosome regions.

3. The databases
The databases that comprise DHMHD are derived from different sources and authors. These are as follows.

3.1 THE LONDON DYSMORPHOLOGY DATABASE (LDDB)
This is a database of over 3000 non-chromosomal malformation syndromes. The full database is microcomputer based and is published by Oxford University Press. It contains comprehensive data for each syndrome including references, a detailed abstract, and clinical photographs stored on CD-ROM. In the present application a subset of around 50% of the syndromes has been chosen. This is because many of the entries on the full database refer to single case reports or families which are important to register for clinical diagnostic purposes but which are not yet useful for gene mapping purposes. The syndromes are classified according to a master list of clinical features. This list is a subset of the full LDDB clinical feature list containing the major malformations or physical abnormalities. In DHMHD, the clinical features of each syndrome are given with a link to OMIM via the McKusick number.

3.2 MOUSE MALFORMATION DATABASE
This is an updated database of mouse malformation syndromes developed with the same system for phenotypic description used in LDDB. Links to mouse databases at other centres are provided, for example, The Mouse Genome Database, which incorporates The Encyclopaedia of the Mouse Genome and The Mouse Locus Catalogue.

3.3 HUMAN CYTOGENETIC ABERRATIONS
This a database of the clinical effects of human chromosomal aberrations. The data used in the current application is a subset containing the clinical features of all aberrations that have been reported more than once.

3.4 MOUSE/HUMAN CHROMOSOME HOMOLOGY
This is derived from a database which gives details (including associations with human diseases) of all mouse/human homologous genes which have been chromosomally located in both species. It has been compiled and is maintained by Dr Tony Searle at the Medical Research Council Radiobiology Unit, Didcot, UK.

4. DHMHD: a World-Wide Web resource
World-Wide Web operations are distributed across a client computer (where a user connects and browses online information) and one or more host computers which act as information servers. A host computer stores resource information along with the WWW server software necessary to provide access to it. A client (that is, a user with a WWW browser such as Netscape or Mosaic) makes a connection to a WWW source by specifying the correct Uniform Resource Locator (URL), an addressing mechanism based on the Internet Protocol. Resource information usually comprises a directory of hypertext documents which incorporate a number of multimedia formats such as text, graphics, and video. Additionally, and as DHMHD shows, WWW documents can provide an interface to databases located on the host computer. In order to allow client interaction with a database, the system must use server-side programs that conform to the Common Gateway Interface (CGI), the specification defining how queries should be passed to the database and how the associated results should be returned to the client. This configuration is used by DHMHD and is illustrated in fig 1. The URL for DHMHD is "http://www.hgmp.mrc.ac.uk/DHMHD/dysmorph.html". The user interface has been implemented using the HyperText Markup Language, or HTML, which is standard for WWW documents. A key feature of HTML is that it allows Uniform Resource Locators to be defined, or anchored, within a document. These form links that can connect to other WWW resources located on servers anywhere in the world (and thus the World-Wide Web is truly "hyperlinked"). In DHMHD, the main syndrome viewing facility provides hyperlinks to related genome databases, such as OMIM and MGD, although these systems are physically stored elsewhere. To the user, hyperlinks appear highlighted (or displayed underlined or with a different colour to normal text or both).

The CGI programs that communicate between the DHMHD database and the HTML interface have been programmed in Perl. The general method of database interaction is to specify a database query within an HTML page displayed at the client computer and send this query to the host when all query parameters are complete. The controlling CGI program residing on the server computer receives the query, processes it, and returns any results back
to the client in the form of another HTML document. As the returned document is in HTML format, it can also incorporate URL encoded links to related documents.

4.1 DHMHD SYSTEM OPERATIONS

The entry screen, or "home page", of DHMHD contains a number of hyperlinks which point to HTML documents providing general information about the project and allow user feedback through an electronic mail program. The home page also lists the three main functions currently provided by DHMHD: a view function, homology cross referencing, and phenotype-genotype search.

4.1.1 Viewing malformation syndromes and cytogenetic aberrations

The view facility of DHMHD permits a user to browse phenotypic anomalies listed for any of the human and mouse malformation syndromes and all features listed for patients with specific cytogenetic aberrations. Each condition is listed (in alphabetical order) as a hyperlink. To select a particular syndrome the user must click over the highlighted syndrome name with the (computer) mouse. This calls a CGI program which returns the phenotypic features associated with that syndrome. For human syndromes, the McKusick number is also displayed as a hyperlink. The URL for OMIM is anchored to this hyperlink, so if selected the associated OMIM record is retrieved from the database at Johns Hopkins
University Medical School, Baltimore, USA. Similarly, for mouse mutant syndromes, a hypertext link is provided to the Mouse Genome Database maintained at the Jackson Laboratory, Maine, USA. For both human and mouse syndromes the associated chromosome location(s) are also listed.

A similar view function is used for cytogenetic aberrations. The selection page is similar to the human syndrome list, only this time specific chromosome aberrations are listed according to the chromosome number and whether they are deletions or duplicated segments. On selection, all features listed for patients with the highlighted aberration are listed with an indication of their incidence across the affected cases. Fig 2 shows the WWW page displayed when the cytogenetic aberration del(7)p15–p13 is selected.

4.1.2 Homology cross referencing

The homology cross referencing facility permits the user to search for either mouse or human syndromes based on chromosome location, or search for homologous human and mouse chromosome segments directly. The listed options form two sets of complementary programs that allow a search to be directed from the perspective of mouse or human disorders. To search for syndromes that are associated with a chromosome region, an HTML form (comprising input text fields, radio buttons, and checkboxes) is used to set up the query. The database of human dysmorphic syndromes is then checked for any syndromes that have a listed chromosome location that overlaps with the specified region. Matching syndromes are returned as a hyperlink list, which is the list of human syndromes that have an associated chromosome location overlapping with chromosome 7 on the short arm (p). Selection of one of these hyperlinks will call the view program described above. An alternative mode of operation is to search human-mouse homologies. As with syndrome queries, complementary programs are provided in order to search from human or mouse chromosome locations. Fig 3.1 shows part of the human chromosomes page which lists hyperlinks for specific human chromosome regions associated with chromosome 7. Selection of any one of these links will call a mapping program that searches for homologous mouse chromosome regions. Human-mouse homologous regions are listed along the left hand side of the returned page, along with homologous loci (see fig 3.2). The (hyperlinked) homologous loci point to relevant OMIM or MGD entries stored at the Johns Hopkins University and Jackson Laboratory sites respectively. * If one of the listed homologous chromosome regions is selected, then the syndrome search program returns mouse syndromes that match the relevant chromosome region. For example, referring to fig 3.2, if the mouse chromosome region 13 CEN-A3 (listed as homologous to human position 7p13) was selected, all mouse syndromes that have an associated chromosome location overlapping with this region will be returned as a hyperlink list that will in turn call the mouse syndrome view function.

4.1.3 Phenotype-genotype searches

The most complex facility of DHMHD involves searching the database for disorders that match a phenotype-genotype combination. The main search page is a query specification form which allows selection of a set of up to five phenotypic features as well as an optional

* Recent updates to the MRC HGMPI World-Wide Web server have included a (mirror) copy of MGD which is updated regularly. DHMHD now links to the UK mirror of MGD in order to improve speed for UK users.
The dysmorphic human-mouse homology database (DHMHD)

Figure 4.1 Result of query “What human and mouse syndromes share the features patchy depigmentation of hair and sensorineural deafness?”

Figure 4.2 Result of query “What human dysmorphic and cytogenetic syndromes with polydactyly/bifid hallux on human chromosome 7 have locations in homologous regions to mouse syndromes with polydactyly/bifid hallux?”

The user may want to ask the question “Is a particular feature (for example, polydactyly/bifid hallux) found in connection with syndromes in homologous regions in man and mouse?” An additional feature that allows this search facility is the link option. This is denoted by a separate checkbox on the main search screen which can be switched on when specifying which of the three databases to search. When set, this option will cause the search program to examine homologous mappings between any of the disorders found to match the phenotype selection criteria. Fig. 4.2 shows the resulting page when the link option is set for the search polydactyly/bifid hallux and chromosome 7. The top part of the screen informs the user that Greig syndrome is connected with mouse syndrome Xt-extra toes by the homologous mapping between the human chromosome region 7p13 and the mouse chromosome region 13 CEN-A3. Greig syndrome is also connected with human 7p13 through the cytogenetics database.

5. Discussion

At the outset of the DHMHD project, one of the main goals was to provide an online resource comprising data that had previously only been available in stand alone PC format. The World-Wide Web had a number of attractions in achieving this aim. Firstly, the World-Wide Web offered a generic solution. Server and client software is available for numerous hardware and operating system configurations, and is constantly under development. Also, a vast proportion of software is public domain, and there are many freely available WWW software tools which ease the development of interactive online resources of this type. There are disadvantages, however, particularly with regard to the limitations of HTML when complex interaction is required as it is with DHMHD. This pertains mainly to the phenotype-genotype search facility, in which the user is required to set up a database query from many options. There is too much detail for one main search form, and this has forced the liberal use of CGI programs in order to provide a look up facility using dynamically created HTML documents (a technique known as creating documents on-the-fly). This mechanism really only provides “real time” interaction when performance (that is, speed in terms of network connections) is good. It is anticipated that problems of this type will be overcome in time and with inevitable improvements with regard to WWW software and performance.

The DHMHD project is continuing and the system will be extended by incorporating chromosome image maps (ideograms) and the locations of candidate genes within these genetic maps. These will be genes that could possibly give rise to malformation syndromes when mutated. The definition of a candidate gene will be very loose and will include genes for any transcription factors, growth factors or receptors, structural proteins, etc. Hypertext links between the gene symbols and OMIM, GDB, and MGD will be created. This facility
will allow users rapidly to search for possible candidate genes near to or within a candidate chromosome region which has been identified as possibly containing a gene or genes causing malformations from the phenotype databases.

LDDB and HCDB data have been made available by Oxford University Press. The DHMHD project is funded by a research award from the Medical Research Council. A G Searle acknowledges assistance from the Human Genome Mapping Project.

4 Searle AG. Tables of generic homology. Mouse Genome 1995;93:102-44.