Database, Vol. 2009, Article ID bap004, doi:10.1093/database/bap004

Original article

The Homeodomain Resource: a comprehensive collection of sequence, structure, interaction, genomic and functional information on the homeodomain protein family

R. Travis Moreland†, Joseph F. Ryan†, Christopher Pan and Andreas D. Baxevanis*

Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

***Corresponding author:** Tel: +1 301 496 8570; Fax: +1 301 402 6858; Email: andy@nhgri.nih.gov

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Submitted 7 November 2008; Accepted 14 March 2009

The Homeodomain Resource is a curated collection of sequence, structure, interaction, genomic and functional information on the homeodomain family. The current version builds upon previous versions by the addition of new, complete sets of homeodomain sequences from fully sequenced genomes, the expansion of existing curated homeodomain information and the improvement of data accessibility through better search tools and more complete data integration. This release contains 1534 full-length homeodomain-containing sequences, 93 experimentally derived homeodomain structures, 101 homeodomain protein–protein interactions, 107 homeodomain DNA-binding sites and 206 homeodomain proteins implicated in human genetic disorders.

Database URL: The Homeodomain Resource is freely available and can be accessed at http://research.nhgri.nih.gov/homeodomain/

Introduction

Homeodomain-containing proteins are transcription factors that play a critical role in various cellular processes, including body plan specification, pattern formation and cell fate determination during metazoan development (1). Members of this family are characterized by a helix-turn-helix DNA binding motif known as the homeodomain. X-ray crystallographic and NMR spectroscopic studies on several homeodomain-containing proteins (2–6) show that this motif is comprised of three α-helices that are folded into a compact globular structure with an N-terminal extension. Helices I and II lie parallel to each other and across from the third helix. This third helix is also referred to as the 'recognition helix', as it confers DNA-binding specificity on individual homeodomain proteins. Homeodomain-containing proteins may interact with each other to enhance or mediate transcriptional activity, either by the binding of multiple proteins to the same segment of DNA or through the formation of DNA-independent complexes. Nucleotide- and protein-level mutations associated with homeodomain proteins can lead to a number of congenital abnormalities [c.f. (7,8)]. The homeodomain structural motif is highly conserved across eukaryotic species, and the expansion and diversification of this family of proteins in various lineages has been shown to coincide with the advent of major morphological innovations (9–12).

Database, Vol. 2009, Article ID bap004, doi:10.1093/database/bap004

In recent years, studies utilizing high-throughput techniques have generated an extraordinary amount of information about these homeodomain proteins, but this information is not always easily accessible to the working biologist. For instance, recent large-scale genome sequencing efforts have led to the availability of complete collections of homeodomain proteins from an evolutionarily diverse set of species, but retrieving complete sets of homeodomain sequences from a particular species is not trivial. Likewise, while several large-scale projects aimed at computationally predicting protein–protein interactions through text mining and other similar approaches have been largely successful in terms of identifying potential relationships between proteins, identifying interactions specific to homeodomains remains an arduous task. In addition, the determination of 3D structures, identification of protein binding sites and our knowledge regarding the role of specific homeodomain proteins in disease causation has been steady, so keeping abreast of these discoveries remains challenging.

The Homeodomain Resource uses a combination of automated and manually verified extraction methods to yield a comprehensive collection of sequence, structure, interaction, genomic and functional information on the homeodomain family (13,14). In addition to a complete collection of homeodomains for 24 species (Table 1), the Homeodomain Resource contains information on DNA-binding targets, protein–protein interactions, 3D structures and homeodomains implicated in human disorders. Each annotation is manually curated, mapped to a specific protein and organism and fully cross-referenced to various external databases, including its primary citation in PubMed. Data are presented in an intuitive, user-friendly format and is keyword-searchable across all tables. Each reference in this database is rigorously selected to assure non-redundancy, and updates are performed on a continuous basis.

Examples of how data from the Homeodomain Resource have been used in various biological contexts to date include studies on the prediction of specific DNA-binding sites for homeodomain proteins (15), the analysis of non-conserved co-evolving positions within functional sites in a variety of protein families (16) and the interpretation of phage display selection experiments aimed at identifying elements within the engrailed homeodomain responsible for sequence-specific DNA binding (17). These data have also been used to help interpret features found within the structures of the stem cell transcription factor Nanog (18) and the *Drosophila* Bicoid–DNA complex (19). Finally, information from the Homeodomain Resource has been used as a reference to aid in understanding mutation data from patients with disorders such as idiopathic short stature and Leri-Weill dyschondrosteosis (20) and brachydactyly types D and E (21).

Database description

The Homeodomain Resource has expanded significantly since its last release [Tables 1 and 2; (13)], and substantial enhancements have been made to the user interface to allow for easier navigation and overall usability. Unlike previous versions of the database, the current version connects all annotations in a relational framework (Figure 1), providing an integrated view of all the analyses associated with a particular homeodomain protein. This new system allows for a more powerful query engine that enables a user to query across multiple annotations in a single search (Figure 2). Homeodomain Resource accession numbers are assigned to each entry in the database to facilitate data

Table 1. Homeodomain Resource statistics, by species

Species name	Kingdom	Number of sequences
Arabidopsis thaliana	Plantae	88
Aspergillus nidulans	Fungi	6
Aspergillus niger	Fungi	8
Caenorhabditis elegans	Metazoa	95
Chaetomium globosum	Fungi	6
Coccidioides immitis	Fungi	7
Coprinopsis cinerea	Fungi	11
Coturnix japonica	Metazoa	1
Danio rerio	Metazoa	155
Dictyostelium discoideum AX4	Protozoa	14
Drosophila melanogaster	Metazoa	105
Gallus gallus	Metazoa	2
Homo sapiens	Metazoa	299
Laccaria bicolor	Fungi	9
Magnaporthe grisea	Fungi	7
Mesocricetus auratus	Metazoa	3
Mus musculus	Metazoa	356
Nematostella vectensis	Metazoa	130
Neurospora crassa	Fungi	6
Oncorhynchus tshawytscha	Metazoa	1
Paramecium tetraurelia strain d4-2	Protozoa	15
Rattus norvegicus	Metazoa	198
Saccharomyces cerevisiae	Fungi	9
Sclerotinia sclerotiorum	Fungi	8
Tetrahymena thermophila SB210	Protozoa	1
Trichomonas vaginalis G3	Protozoa	14
Trichoplax adhaerens	Metazoa	35
Ustilago maydis	Fungi	7
Xenopus laevis	Metazoa	2
Xenopus tropicalis	Metazoa	1

Species in bold denote those whose homeodomains were extracted from full genome scans.

sharing amongst the user community. These accession numbers take the format HDR*xn*, where *x* indicates the data category for the entry (e.g. s = structures) and *n* is a three-digit number identifying the entry. In addition, the database is more genome-centric, with an eye towards evolutionary studies. Whereas previous versions relied heavily on choosing proteins that had annotations from Swiss-Prot associated with them, this new edition places more emphasis on compiling complete sets of homeodomains from a diverse range of species. The combination of additional sequences, more comprehensive datasets, and greater data connectivity provides a much more powerful and robust resource to biologists.

Table 2. Homeodomain Resource statistics, by category

Protein-coding genes	1534
Pseudogenes	65
Distinct organisms	30
3D structures	93
Homeodomain proteins implicated in human genetic disorders	206
Homeodomain proteins with documented allelic variants	53
Homeodomain DNA-binding sites	107
Protein–protein interactions involving homeodomain proteins	101

Homeodomain protein sequence entries

The sequence dataset in the Homeodomain Resource was assembled by first utilizing data from a series of homeodomain surveys of metazoan genomes (22–24). Next, a hidden Markov model (HMM) was generated from these aligned sequences using the HMMer Toolkit (25), and the HMM was subsequently used to search RefSeq (26) to identify additional members of the homeodomain family. Alignments produced by HMMsearch (25) were parsed using Perl scripts; this was followed by manual alignment to the HMMsearch alignment using GeneDoc (27). Inspection (and manual adjustment) of the alignments become necessary if HMMsearch introduces gaps in biologically implausible locations within the sequence. One such example involves the sequence of HDRp1895, which is truncated at its N-terminus. HMMsearch introduced a gap of length 7 between the next-to-last (R52) and the

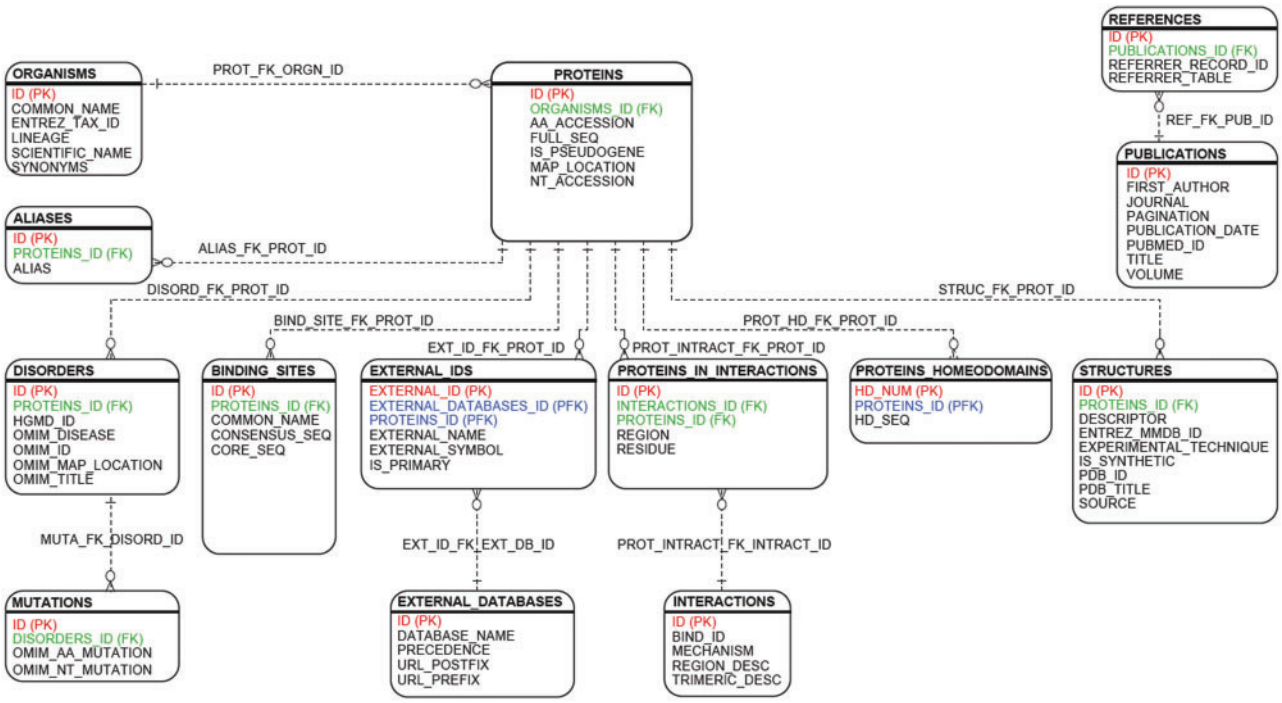

Figure 1. HDR relational framework. The relational design connecting the Homeodomain Resource's 14 primary tables are illustrated in the figure. Primary keys are indicated in red, foreign keys in green and keys characterized as both primary and foreign in blue. The new database design is centered around data found in the 'Proteins' table. All proteins are lineage-specific and linked to the 'Organisms' table. A single protein may contain one or more homeodomains related to the 'Proteins_Homeodomains' table. DNA-binding targets, protein–protein interactions, 3D structures and homeodomains implicated in human disorders are normalized and linked to the 'Proteins' table. External annotation from multiple databases are integrated via the 'External_Ids' table. Database entries are referenced with their primary citation via the 'Publications' table.

Database, Vol. 2009, Article ID bap004, doi:10.1093/database/bap004

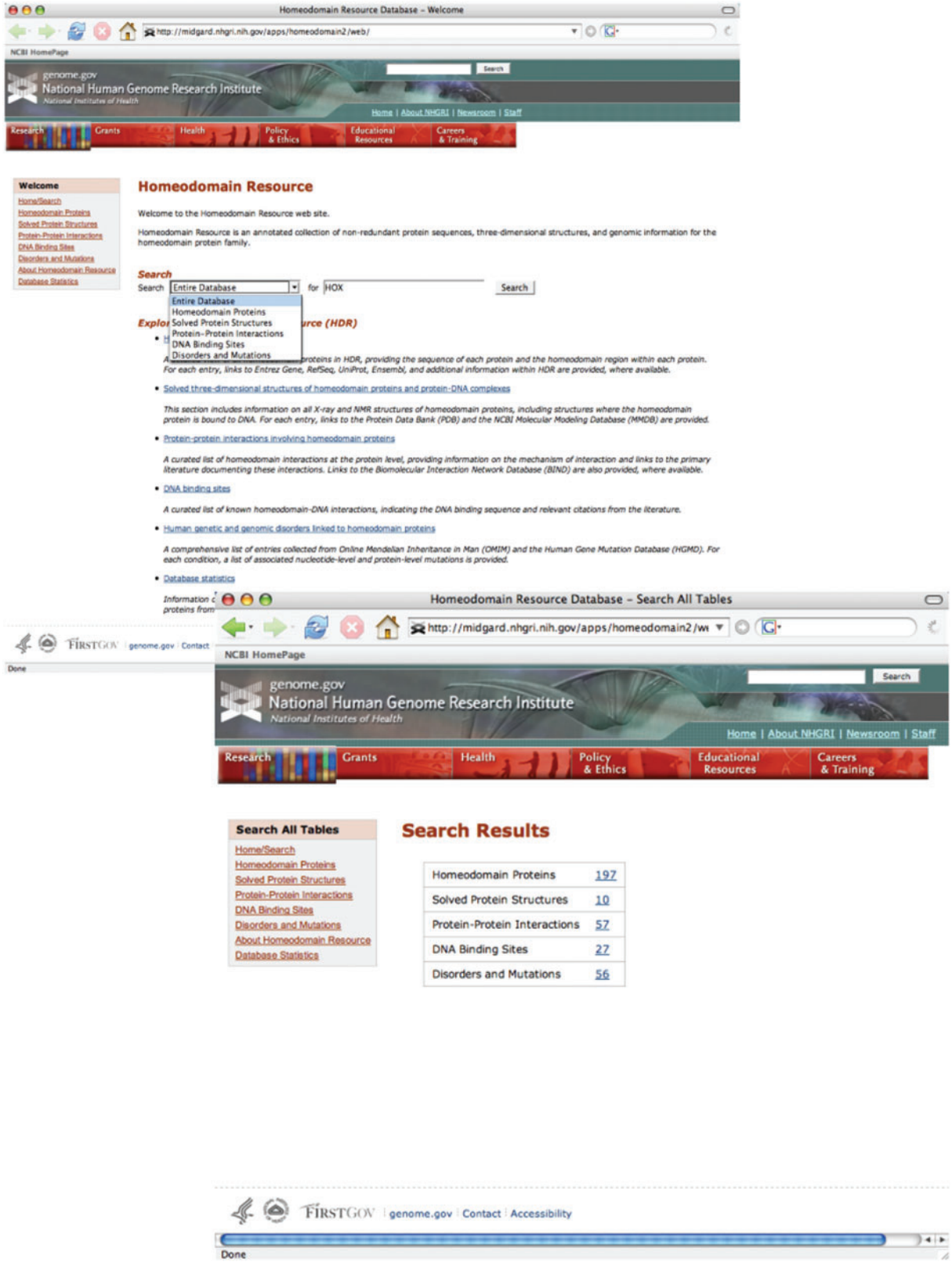

Figure 2. The Homeodomain Resource provides a simple search query interface, allowing the user to either query part or all of the Resource (top). Selecting 'Entire Database' from the pull-down menu returns a summary screen, indicating how many entries of each type were identified (bottom). Clicking on any of the hyperlinked numbers in the table takes the user directly to that set of results. In addition, overall navigation within the site has been improved with the addition of sidebar tools and links to complete datasets in each homeodomain category.

last (Q53) residue in the sequence; in this case, the gap was removed, placing R52 directly next to Q53, thereby producing a better-quality alignment. These alignments are then added, along with annotations from Entrez Gene (28), to the Homeodomain Resource. The International Protein Index (29) was used to match Entrez Gene identifiers with entries from other external resources, such as the Mouse Genome Database (MGD; 30) and the Zebrafish Information Network (ZFIN; 31), where possible. As of December 2008, 24 fully sequenced genomes have been sampled: 8 metazoan (4 vertebrate and 4 invertebrate), 11 fungi, 4 protozoan and 1 plant (Table 1). This process yielded 1534 protein entries. Individual protein entries are hyperlinked to a detailed view that presents gene- and protein-level annotation, full-lineage taxonomy and both the full-length and homeodomain-only sequences. Annotations that refer to external resources are hyperlinked to their source database (e.g. Entrez Gene).

The complete set of homeodomain proteins can be downloaded in FASTA format as either full-length sequences or homeodomain alignments. Alternatively, a customized dataset can be built either by selecting sequences resulting from a query or by manually selecting sequences from the entire dataset. Query results can be sorted to facilitate the construction of custom datasets. The ability to retrieve a complete set of aligned homeodomains from a range of species makes the Homeodomain Resource an invaluable first step in a phylogenetic analysis. For example, a researcher wanting to know the phylogenetic affinity of a previously undescribed homeodomain from a fungus could download an aligned dataset of homeodomains from several fungal species, align the undescribed homeodomain to this dataset and then run one or more phylogenetic algorithms on this alignment. Users interested in an evolution-based classification of homeodomain-containing proteins are also encouraged to explore HomeoDB (32), a complementary database focusing on homeobox gene phylogenetic classification.

Structures of homeodomain proteins and protein–DNA complexes

The homeodomain structures are manually compiled from the NCBI Entrez Structure database (33) and the Protein Data Bank (PDB; 34). Each structural entry is manually inspected to ensure that the solved structure contains the homeodomain region of the protein. Also noted is the experimental technique used to determine its structure (either X-ray diffraction or NMR spectroscopy). Information on solved 3D structures of both homeodomain proteins and protein–DNA complexes is available in a concise, columnar format. Protein name, PDB and MMDB accessions and the source organism are given for each entry, and the table can be sorted, as needed. For each entry, a link is also provided to a detailed view of that structure record, providing additional information such as experimental technique, PDB title and its primary PubMed reference. From this detailed view, users can follow links to the source Entrez Structure and PDB records, where one can view still images of the structure and download the 3D coordinates of a structure of interest. The detailed view also provides a link to the protein annotation within the Homeodomain Resource itself, as well as to the PubMed abstract corresponding to the primary literature citation listed in PDB.

Protein–protein interactions involving homeodomain proteins

The Homeodomain Resource contains a systematically and thoroughly curated catalogue of experimentally determined protein–protein interaction data for the homeodomain protein family. To the best of our knowledge, this collection represents the most comprehensive collection of protein–protein interaction annotations specific to the homeodomain family. Interaction data were collected through manual literature searches; essential information about the nature of the specific protein–protein interactions was then extracted from the experimental data presented in these manuscripts. The identification of articles containing relevant biological information from PubMed required the use of discriminatory MeSH terms, from specific to more general keyword search combinations. PubMed titles, abstracts and full text were searched for keywords that would be indicative of relevant protein–protein interactions (e.g. 'DNA-independent interaction'). Interacting proteins were annotated and cross-linked to their corresponding protein entry within the Homeodomain Resource.

Protein–protein interaction data can be searched by publication information, interaction description and keyword data associated with their corresponding protein entries. Interaction data are returned in columnar format, listing the interacting proteins, the primary citation from the literature, the corresponding Biomolecular Interaction Network Database (BIND; 35) identifier and a link to a detailed view of the interaction. The detailed view provides additional information describing the homeodomain protein interacting regions, interacting residue locations and a description of the mechanism of interaction derived from the primary publication, as well as internal links to details on each of the interacting proteins within the HDR.

A new feature of this release is the cross-referencing of homeodomain protein–protein interaction data to their respective BIND interaction entries. BIND was queried for previously unreported homeodomain protein–protein interactions in parallel with the aforementioned PubMed literature searches, using general (e.g. 'homeobox OR homeodomain AND interaction') to more specific (e.g. 'homeobox OR homeodomain AND interaction_object_type = protein AND NOT = DNA') search criteria. Following

Database, Vol. 2009, Article ID bap004, doi:10.1093/database/bap004

a manual extraction of false positives, interactions from BIND were extracted and deposited in the Homeodomain Resource. All protein–protein interaction data derived from manual curation of PubMed have also been deposited into the BIND database. Each interaction derived from the Homeodomain Resource has been assigned a unique BIND accession number and is hyperlinked from BIND back to the Homeodomain Resource (Figure 3).

Homeodomain DNA-binding sites

DNA binding sites for homeodomain proteins have been obtained through extensive review of the published literature, citations in Online Mendelian Inheritance in Man (OMIM; 36,37) and entries for DNA-bound homeodomain structures from PDB. As with the interaction data described above, binding site data can be searched by publication information and by keyword data associated with its corresponding protein entry. Binding site data are returned in columnar format; the columns include homeodomain names, their respective DNA-binding sequences and references to the primary citation from which the information was retrieved. The core regions of each of the DNA binding sites are shown in bold type. A detailed view of the binding site record displays the consensus DNA sequence, the corresponding PubMed reference and a link to details about the protein; the protein details includes the Protein HDR identifier, the common name of the protein, the gene symbol listed in Entrez Gene and the UniProt protein accession.

Human genetic and genomic disorders linked to homeodomain proteins

Information on human genetic and genomic disorders linked to homeodomain proteins has been compiled from manual searches of both OMIM and the Human Gene Mutation Database (HGMD; 38). Any false positives resulting from the OMIM and HGMD searches were manually removed from the dataset. Manually derived entries from the previous Homeodomain Resource release were automatically compared and updated, while new automated entries were manually verified.

Each entry in the Disorders and Mutations dataset represents a single homeobox gene associated with one or more disease(s) or disorder(s). For each, the corresponding OMIM nucleotide- (e.g. 1-BP DEL, 504T) and/or protein-level (e.g. GLN140TER) mutations are shown. This dataset can be queried using any of the aforementioned fields, and the results can be sorted by clicking on the appropriate column field heading. Gene symbols are hyperlinked to the corresponding entry in the proteins table as well as to entries in HGMD (registration required).

Technical improvements

In addition to an overhaul of the interface, a number of back-end technical modifications have been made to improve data collection, storage and automation. A number of new Perl scripts have been developed for this release which facilitate the automation and updating of external annotation sources linked to the database, thereby eliminating a number of manual steps previously required for these processes. For example, a new set of Perl scripts uses a list of existing gene symbols obtained from Swiss-Prot to automatically search Entrez Gene, pairing protein-centric annotation of existing homeodomain entries with their gene-centric equivalent. A second set of Perl scripts parses Entrez data via E-utilities, mapping a homeodomain entry to its corresponding Disease and Disorders entry at OMIM. Each of the new entries is examined manually and either added to the database or designated as false positive. The search and update functions are executed quarterly to update the disorders and mutations annotation. Another Perl script was developed to parse the output of HMMsearch, retrieve sequence and annotation information from Entrez, and insert unique hits into the Homeodomain Resource. This approach results in a relatively simple pipeline for adding new sequence entries, thereby keeping this database current.

Future considerations

With these new tools in hand for importing complete sets of homeodomain sequences from fully sequenced genomes, we intend to continue to add sequence data from already-sequenced species. We also intend to include additional homeodomain sequence data from newly sequenced genomes, fully anticipating a new wave of such data becoming available with the advent of new, next-generation sequencing technologies.

It is becoming increasingly evident that homeodomain transcription factors have played and continue to play key roles in the evolution of eukaryotic species. Likewise, research in this area continually shows that disruptions in the wild-type function of this class of proteins underlie a significant number of devastating human disorders, as evidenced by the extensive list of genetic and genomic disorders catalogued in the Disorders and Mutations section of the Homeodomain Resource Web site. As a result, the amount of homeodomain-related data being generated—and the need for biologists to be able to process and consider these data—will be critical to the advancement of our understanding of these proteins. It is our intention to continue to maintain and update the Homeodomain Resource in the future, so as to provide a solid discovery framework for biologists and clinicians studying this important class of proteins.

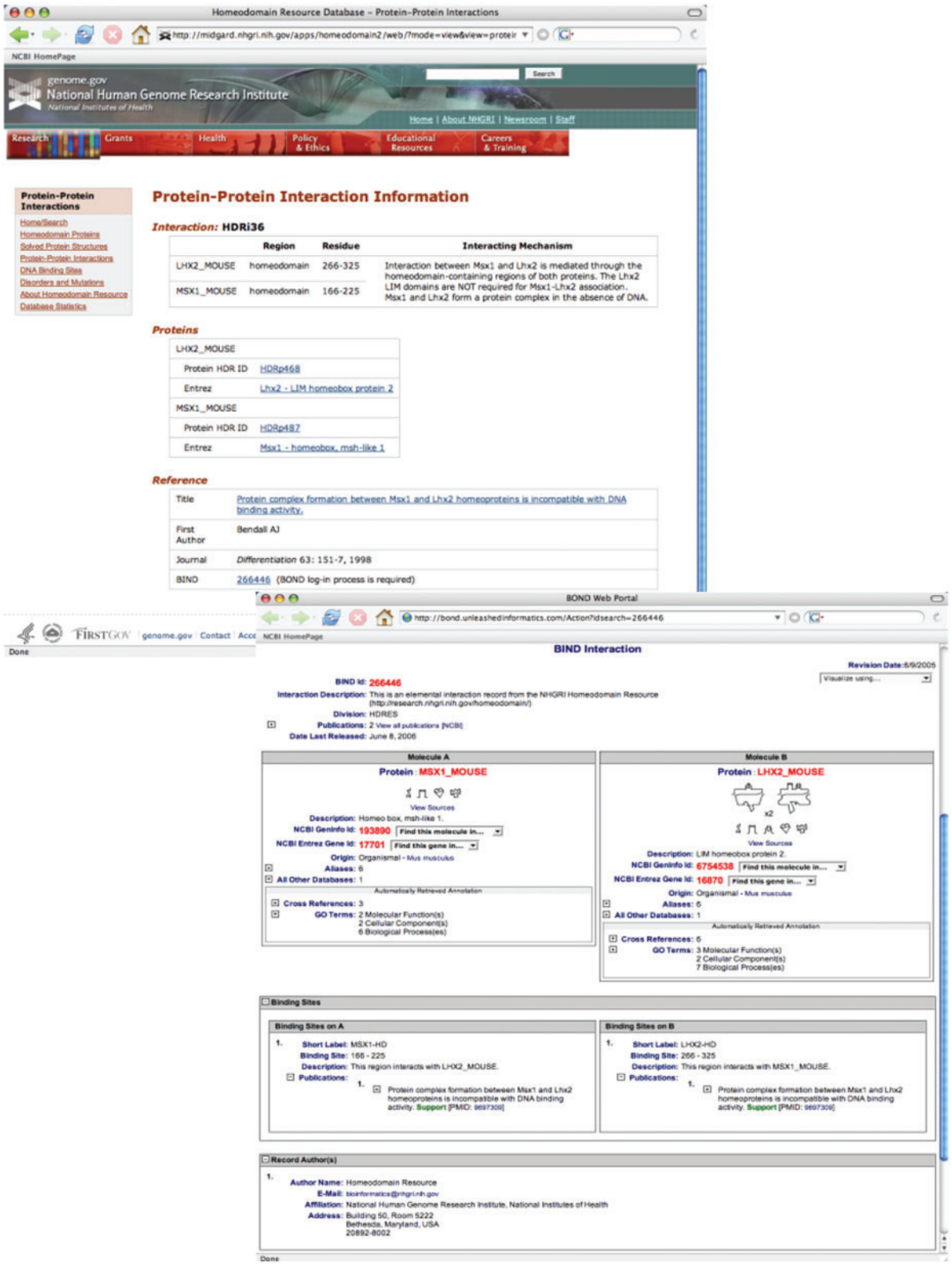

Figure 3. Search results from a query of protein–protein interactions data for the interaction of homeodomain proteins Lhx2 and Msx1 (SEARCH 'Protein-Protein Interactions' FOR 'MSX1') (top). Each protein–protein interaction entry within the Homeodomain Resource is hyperlinked to the corresponding entry in BIND, which provides additional details on the mechanism(s) of interaction (bottom). See text for additional details.

Funding

This research was supported by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health.

Conflict of interest. None declared.

References

1. Gehring,W.J., Affolter,M. and Burglin,T. (1994) Homeodomain proteins. *Annu. Rev. Biochem.*, **63**, 487–526.
2. Ceska,T.A., Lamers,M., Monaci,P. *et al*. (1993) The X-ray structure of an atypical homeodomain present in the rat liver transcription factor LFB1/HNF1 and implications for DNA binding. *EMBO J.*, **12**, 1805–1810.
3. Dekker,N., Cox,M., Boelens,R. *et al*. (1993) Solution structure of the POU-specific DNA-binding domain of Oct-1. *Nature*, **362**, 852–855.
4. Endo,T., Ohta,K., Saito,T. *et al*. (1994) Structure of the rat thyroid transcription factor-1 (TTF-1) gene. *Biochem. Biophys. Res. Commun.*, **204**, 1358–1363.
5. Kissinger,C.R., Liu,B.S., Martin-Blanco,E. *et al*. (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. *Cell*, **63**, 579–590.
6. Wolberger,C., Vershon,A.K., Liu,B. *et al*. (1991) Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell*, **67**, 517–528.
7. Chi,Y.I. (2005) Homeodomain revisited: a lesson from disease-causing mutations. *Hum. Genet.*, **116**, 433–444.
8. D'Elia,A.V., Tell,G., Paron,I. *et al*. (2001) Missense mutations of human homeoboxes: a review. *Hum. Mutat.*, **18**, 361–374.
9. Bürglin,T.R. (2005) In Meyers,R.A. (ed), *Encyclopedia or Molecular Cell Biology and Molecular Medicine*, 2nd edn. Wiley-VCH Verlag, Weinheim.
10. Valentine,J.W. and Jablonski,D. (2003) Morphological and developmental macroevolution: a paleontological perspective. *Int. J. Dev. Biol.*, **47**, 517–522.
11. McGinnis,W., Levine,M.S., Hafen,E. *et al*. (1984) A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature*, **308**, 428–433.
12. Lewis,E.B. (1978) A gene complex controlling segmentation in Drosophila. *Nature*, **276**, 565–570.
13. Banerjee-Basu,S., Moreland,T., Hsu,B.J. *et al*. (2003) The Homeodomain Resource: 2003 update. *Nucleic Acids Res.*, **31**, 304–306.
14. Banerjee-Basu,S., Sink,D.W. and Baxevanis,A.D. (2001) The Homeodomain Resource: sequences, structures, DNA binding sites and genomic information. *Nucleic Acids Res.*, **29**, 291–293.
15. Berger,M.F., Badis,G., Gehrke,A.R. *et al*. (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*, **133**, 1266–1276.
16. Gloor,G.B., Martin,L.C., Wahl,L.M. *et al*. (2005) Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions&#x2020. *Biochemistry*, **44**, 7156–7165.
17. Simon,M.D., Sato,K., Weiss,G.A. *et al*. (2004) A phage display selection of engrailed homeodomain mutants and the importance of residue Q50. *Nucleic Acids Res.*, **32**, 3623–3631.
18. Jauch,R., Ng,C.K., Saikatendu,K.S. *et al*. (2008) Crystal structure and DNA binding of the homeodomain of the stem cell transcription factor Nanog. *J. Mol. Biol.*, **376**, 758–770.
19. Baird-Titus,J.M., Clark-Baldwin,K., Dave,V. *et al*. (2006) The solution structure of the native K50 Bicoid homeodomain bound to the consensus TAATCC DNA-binding site. *J. Mol. Biol.*, **356**, 1137–1151.
20. Jorge,A.A., Souza,S.C., Nishi,M.Y. *et al*. (2007) SHOX mutations in idiopathic short stature and Leri-Weill dyschondrosteosis: frequency and phenotypic variability. *Clin. Endocrinol.*, **66**, 130–135.
21. Johnson,D., Kan,S.H., Oldridge,M. *et al*. (2003) Missense mutations in the homeodomain of HOXD13 are associated with brachydactyly types D and E. *Am. J. Hum. Genet.*, **72**, 984–997.
22. Holland,P.W., Booth,H.A. and Bruford,E.A. (2007) Classification and nomenclature of all human homeobox genes. *BMC Biol.*, **5**, 47.
23. Nam,J. and Nei,M. (2005) Evolutionary change of the numbers of homeobox genes in bilateral animals. *Mol. Biol. Evol.*, **22**, 2386–2394.
24. Ryan,J.F., Burton,P.M., Mazza,M.E. *et al*. (2006) The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: evidence from the starlet sea anemone, Nematostella vectensis. *Genome Biol.*, **7**, R64.
25. Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.
26. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
27. Nicholas,K.B., Nicholas,H.B. Jr and Deerfield,D.W. II. (1997) GeneDoc: analysis and visualization of genetic variation. *EMBNEW.NEWS*, **4**, 14.
28. Maglott,D., Ostell,J., Pruitt,K.D. *et al*. (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **35**, D26–D31.
29. Kersey,P.J., Duarte,J., Williams,A. *et al*. (2004) The International Protein Index: an integrated database for proteomics experiments. *Proteomics*, **4**, 1985–1988.
30. Bult,C.J., Eppig,J.T., Kadin,J.A. *et al*. (2008) The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.*, **36**, D724–D728.
31. Sprague,J., Bayraktaroglu,L., Bradford,Y. *et al*. (2008) The Zebrafish Information Network: the zebrafish model organism database provides expanded support for genotypes and phenotypes. *Nucleic Acids Res.*, **36**, D768–D772.
32. Zhong,Y.F., Butts,T. and Holland,P.W. (2008) HomeoDB: a database of homeobox gene diversity. *Evol. Dev.*, **10**, 516–518.
33. Wang,Y., Addess,K.J., Chen,J. *et al*. (2007) MMDB: annotating protein sequences with Entrez's 3D-structure database. *Nucleic Acids Res.*, **35**, D298–D300.
34. Berman,H.M., Westbrook,J., Feng,Z. *et al*. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
35. Alfarano,C., Andrade,C.E., Anthony,K. *et al*. (2005) The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Res.*, **33**, D418–D424.
36. Baxevanis,A.D. (2003) Searching Online Mendelian Inheritance in Man (OMIM) for information for genetic loci involved in human disease. *Curr. Protoc. Hum. Genet.* Chapter 9, Unit913.
37. Hamosh,A., Scott,A.F., Amberger,J.S. *et al*. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
38. Stenson,P.D., Ball,E.V., Mort,M. *et al*. (2003) Human Gene Mutation Database (HGMD): 2003 update. *Hum. Mutat.*, **21**, 577–581.