Chapter 5

Results and Discussion

The dbPedigree database (Database of Pedigree Analysis), a relational database, having a collection of heritable genetic defect; its causable gene; associated mutations and polymorphisms. The counts of disease, gene, pedigree, mutation, SNP, citations are shown in Table 5.1.

DISEASE	360
GENE	450
PEDIGREE	2850
MUTATION	1170
SNPs	408
CITATION	1050

Table 5.1 Shows the counts of disease, gene, pedigree, mutation, SNP, citations.

5.1 The dbPedigree

DiseaseName

The disease name is a broad category of disease. The database has around 360 diseases, therefore 360 Disease-ID.

SubDiseaseName

Represented as SDName in the maintable of the database contains the different categories of the same disease. The Figure 5.1 illustrations Subclasses of Androgen Insensitivity Syndrome namely Androgen Insensitivity Complete and Androgen Insensitivity Partial.

AI15	Androgen Insensitivity Complete	X-Linked		AR	1-BP INS; 179A		Xq12		http://jcem.endojournals.org/content/
AI15	Androgen Insensitivity Complete	X-Linked		AR	2-BP DEL; 180GC		Xq12		http://jcem.endojournals.org/content/
AI15	Androgen Insensitivity Complete	X-Linked		AR	INS/DEL; EX5		Xq12	Danish	http://www.nature.com/jhg/journal/v48
AI15	Androgen Insensitivity Complete	X-Linked		AR	PARTIAL DEL		Xq12		http://www.ncbi.nlm.nih.gov/pmc/artic
AI15	Androgen Insensitivity Complete	X-Linked		AR	TRP717TER		Xq12		http://www.ncbi.nlm.nih.gov/pmc/artic
AI15	Androgen Insensitivity Complete	X-Linked		AR	5-KB DEL; EX E		Xq12		http://www.ncbi.nlm.nih.gov/pmc/artic
AI15	Androgen Insensitivity Partial	X-Linked		AR	TYR761CYS		Xq12		http://www.ncbi.nlm.nih.gov/pmc/artic
AI15	Androgen Insensitivity Partial	X-Linked		AR	TYR761CYS		Xq12		http://www.ncbi.nlm.nih.gov/pmc/artic
AI15	Androgen Insensitivity Partial	X-Linked		AR	ARG839HIS		Xq12		http://www.ncbi.nlm.nih.gov/pmc/artic
AI15	Androgen Insensitivity Partial	X-Linked		AR	ARG839CYS		Xq12		http://www.ncbi.nlm.nih.gov/pmc/artic
AI15	Androgen Insensitivity Partial	X-Linked		AR	GLU2LYS		Xq12		http://www.ncbi.nlm.nih.gov/pmc/artic
AI15	Androgen	X-Linked		AR	ARG846HIS		Xq12		http://jcem.endojournals.org/content/

Figure 5.1 Illustrations Subclasses of Androgen Insensitivity Syndrome namely Androgen Insensitivity Complete and Androgen Insensitivity Partial.

DiseaseType

The number of records mentioned under each category autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive, Y- linked or mitochondrial mutation as reported from literature have been shown in Table 4.2.

DISEASE TYPE	COUNT
AUTOSOMAL DOMINANT :	**1176**
AUTOSOMAL RECESSIVE :	**934**
X-LINKED DOMINANT :	64
X-LINKED RECESSIVE :	11
X-LINKED (UNCLASSIFIED) :	227
TOTAL X-LINKED :	**302**
MITOCHONDRIAL :	**116**

Table 5.2 Shows the number of records mentioned under each category of disease type.

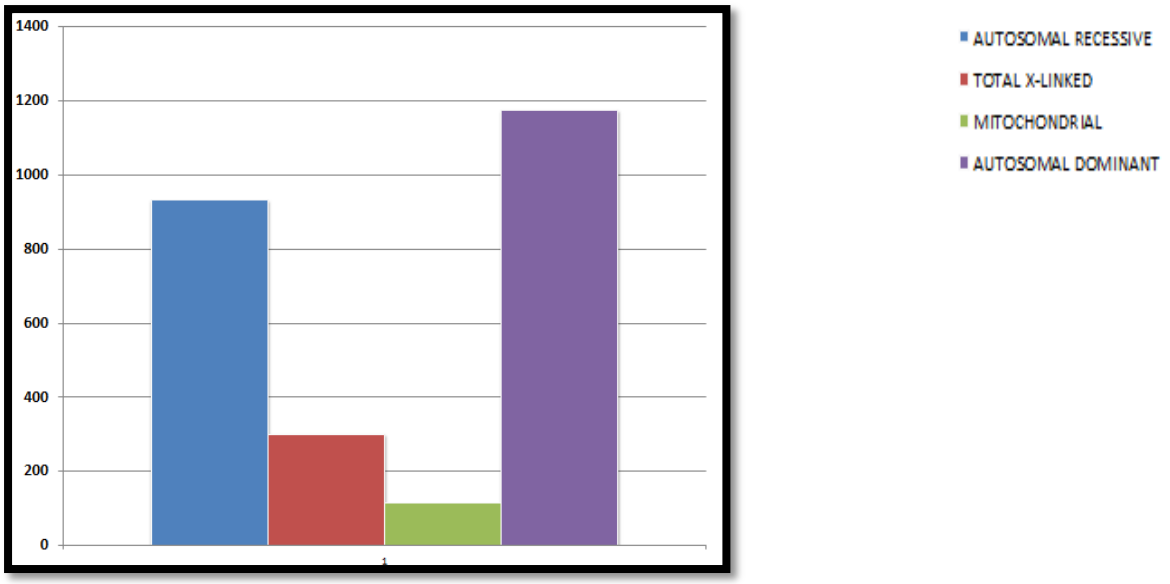


Figure 5.2 Shows the bar chart of data recorded under different modes of inheritance.

OtherDisease

Represented as ODisease in the maintable of the database shows the name of any other disease which could have been present in the proband's family. The Figure 5.3 shows that while studying the pedigree for Li-Fraumeni-Like Syndromes, Sarcomas; premenopausal breast cancers were also considered.

DID	SDName	DType	ODisease	Gene	Locus	Mutation	SNP	Ethnicity	PNo	Link	PName
LI198	Li-Fraumeni-Like Syndromes	Autosomal Dominant	Sarcomas; premenopausal breast cancers	TP53	17p12	ARG248TRP exon 7			PID1	http://www.nejm.org/action/showImage?doi=10.1056%2...	family 6
LI198	Li-Fraumeni-Like Syndromes	Autosomal Dominant	Sarcomas; premenopausal breast cancers	TP53	17p13	ARG248TRP exon 7			PID1947	http://www.nejm.org/action/showImage?doi=10.1056%2...	family 6
NH249			Colon cancer	TP53	17p13	GLY325VAL exon 9			PID10	http://www.nejm.org/action/showImage?doi=10.1056%2...	family 49
OL253				LEP	7q32.1	homozygosity for the delta-133G		PAKISTANI	PID100	http://jcem.endojournals.org/content/89/10/4821/F1...	
EB118		Autosomal		COL7A1	3p21.31	GLY2569ARG			PID1000	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC191467...	family 5

Figure 5.2 Shows the data included in the OtherDisease column.

Gene

Genes hold the information to build and maintain an organism's cells and pass genetic traits to offspring. Any change in the sequence of the gene may give rise to a defective protein or may silent the gene itself. In both the cases the individual is affected. This has been considered as an important feature for searching. The search can be carried out using Gene HGNC Symbol. The HGNC symbol is unique. It is necessary to provide a unify symbol for each gene so that it can be used globally and this also facilitates electronic data retrieval from publications and databases. The DbPedigree database has around 450 Gene HGNC symbols.

GeneLocus

The specific place on a chromosome where a gene is located is referred as locus. The Figure 5.4 shows the number of gene locus located on each chromosome analysed during making of dbPedigree. The maximum number of diseases lies on chromosome 2 and X while the chromosome 22 is least related.

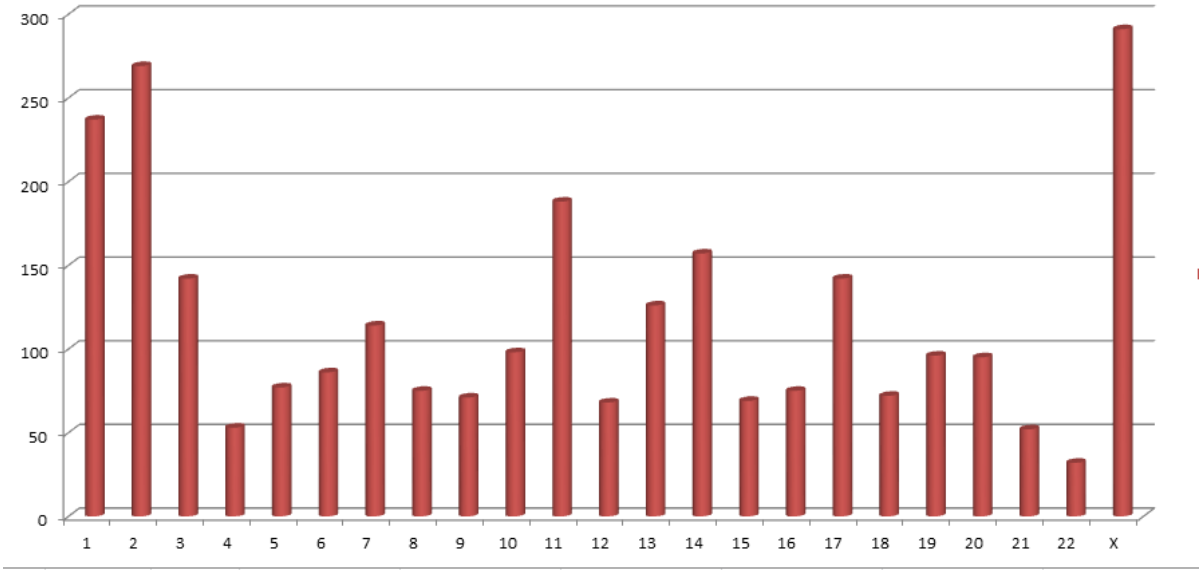


Figure 5.4 Shows the number of gene locus located on each chromosome.

Associated Mutation

The mutation can be insertion, deletion, duplication, single nucleotide change, Translocation etc. The single nucleotide change around 1200 entries of mutational data, rest 1000 entries are all others. The Fig. depicts contribution of each type of mutation.

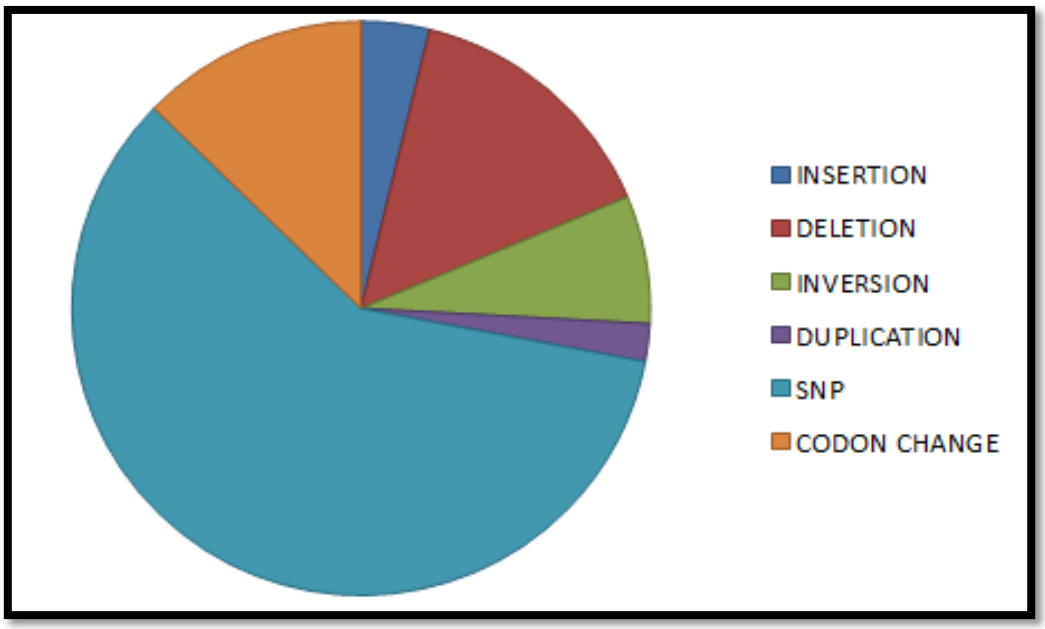


Figure 5.5 Depicts contribution of each type of mutation in building of dbPedigree.

Single Nucleotide Polymorphism

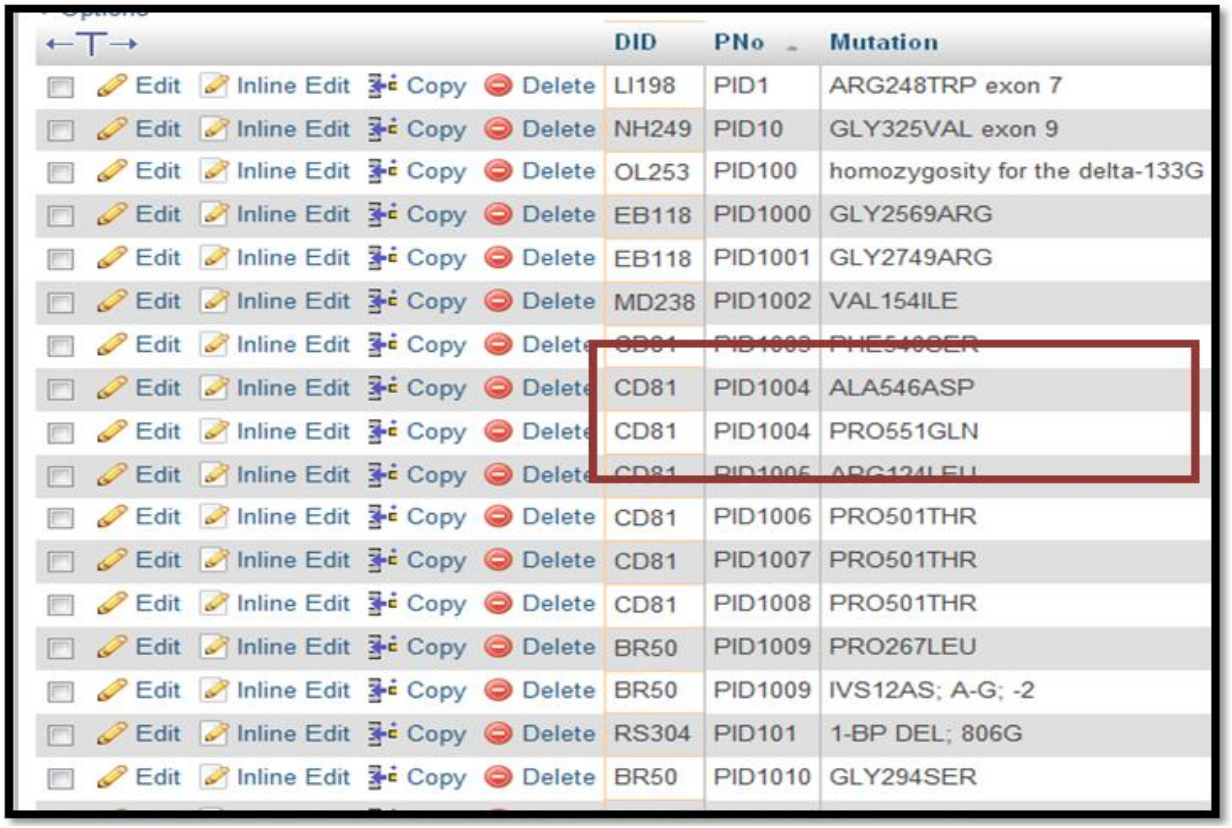
This field contains only those SNP which have been reported in dbSNP. The database contains 408 rs_id's.

Ethnicity

The field contains the region to which the proband belongs. Most of the proband's belonging to African, European, French, United States, and Chinese ethnicity respectively.

PedigreeNumber

This field was introduced so that the problem of a single pedigree involving more than one mutation can be sorted. The Figure 5.6 shows a single pedigree PID1004 showing 2 mutations.



	DID	PNo	Mutation
Edit Inline Edit Copy Delete	LI198	PID1	ARG248TRP exon 7
Edit Inline Edit Copy Delete	NH249	PID10	GLY325VAL exon 9
Edit Inline Edit Copy Delete	OL253	PID100	homozygosity for the delta-133G
Edit Inline Edit Copy Delete	EB118	PID1000	GLY2569ARG
Edit Inline Edit Copy Delete	EB118	PID1001	GLY2749ARG
Edit Inline Edit Copy Delete	MD238	PID1002	VAL154ILE
Edit Inline Edit Copy Delete	CD81	PID1004	ALA546ASP
Edit Inline Edit Copy Delete	CD81	PID1004	PRO551GLN
Edit Inline Edit Copy Delete	CD81	PID1006	PRO501THR
Edit Inline Edit Copy Delete	CD81	PID1007	PRO501THR
Edit Inline Edit Copy Delete	CD81	PID1008	PRO501THR
Edit Inline Edit Copy Delete	BR50	PID1009	PRO267LEU
Edit Inline Edit Copy Delete	BR50	PID1009	IVS12AS; A-G; -2
Edit Inline Edit Copy Delete	RS304	PID101	1-BP DEL; 806G
Edit Inline Edit Copy Delete	BR50	PID1010	GLY294SER

Figure 5.6 Shows a single pedigree PID1004 showing 2 mutations.

Link

The pedigree has been reported in the form of hyperlinks to the images in the research article.

Pedigree Name

If the pedigree link contains more than 1 pedigree, then this field will report the family name or pedigree reference name.

p-value

A measure of how much evidence there is against the null hypothesis. The smaller the p-value, the more evidence exists against. Traditionally, researchers will reject the null hypothesis if the p-value is less than 0.05. A small p-value is evidence against the null hypothesis while a large p-value means little or no evidence against the null hypothesis.

LOD

It is the LOD score, logarithm (base 10) of odds, is a statistical test often used for linkage analysis. It compares the likelihood of obtaining the test data if the two loci are indeed linked,

to the likelihood of observing the same data purely by chance. Positive LOD scores favour the presence of linkage, whereas negative LOD scores indicate that linkage is less likely.

Citation

The PubMed ID or the PubMed Central ID of every literature associated with the pedigree has been reported.

It is a manually curated database expected to reporting 2850 pedigrees. It is an open access database, and user will be able to access it from DTU server.

5.2 The Webpages

A web-based graphical user interface has been designed using HTML. The software used for designing the interface is Macromedia Dreamweaver 8. The dbPedigree has 5 main pages.

The Homepage dbPedigree
About Database
Search Page
Glossary
Contacts

5.2.1 The "HOMEPAGE" dbPedigree

When the user visits the dbPedigree website, the following homepage will be displayed in the browser. It will also have links to types of disease.

Figure 5.3 The dbPedigree homepage.

5.2.2 The "ABOUT DATABASE" Page

The "About Database" gives the brief introduction about the different modes of inheritance of genetic disorder.

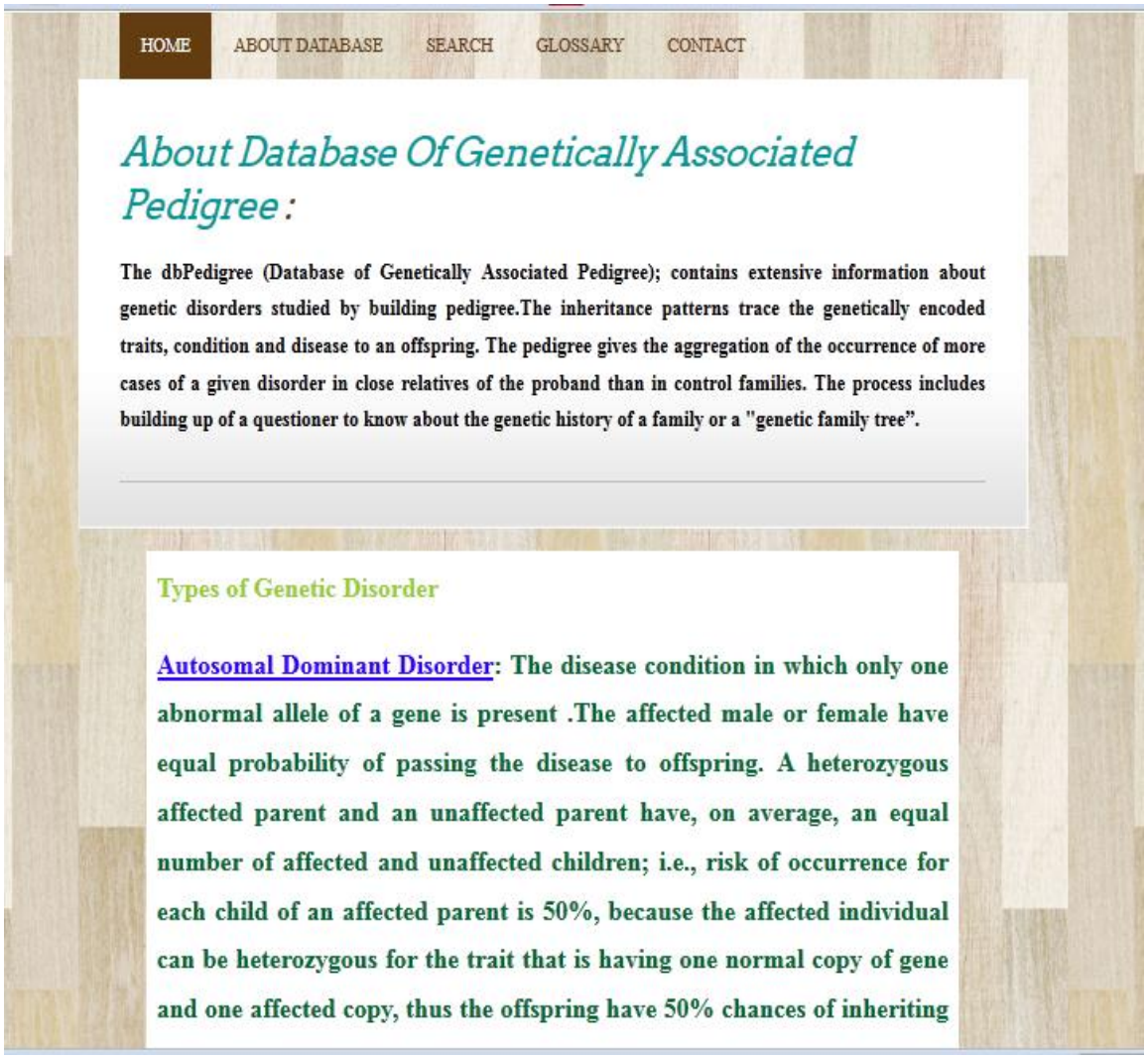


Figure 5.4 The About Database page.

5.2.3 The " SEARCH" Page

The search page allows user to search the database using disease name or the gene name. The page also contains links to list of genetic disorders and list of associated gene.

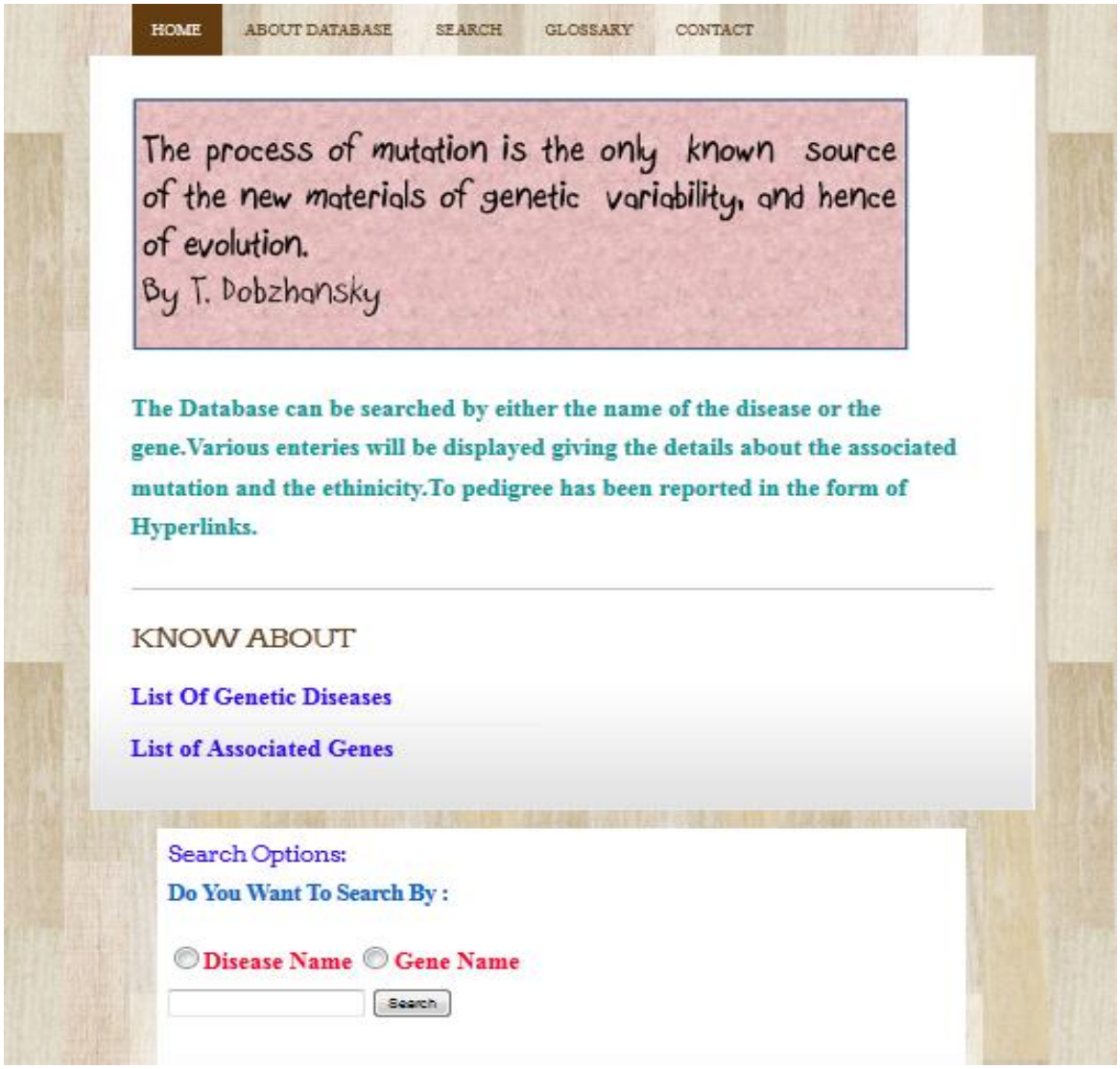


Figure 5.5 The search page.

Search page has links to list of disease and links of genes pages.

5.2.3.1 The "List of Disease" Page

The list of disease contains 360 disorders whose pedigree has been recorded in dbPedigree.

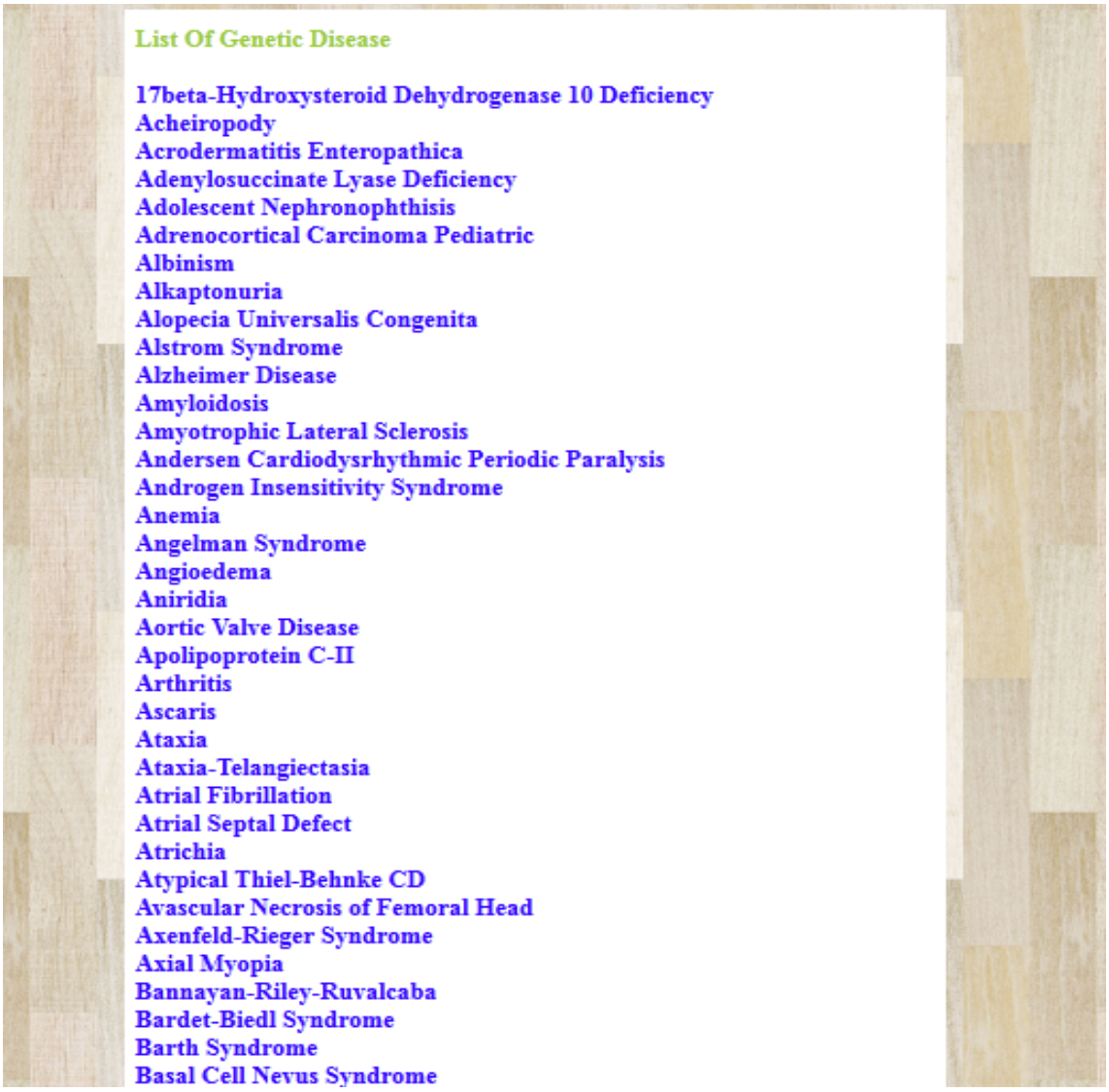


Figure 5.6 The list of genetic disease.

5.2.3.2 The "List of genes" Page

The list of genes contains 450 HGNC symbols of genes associated with genetic disorders that have been recorded in dbPedigree. The list contains 4 column and 113 rows.

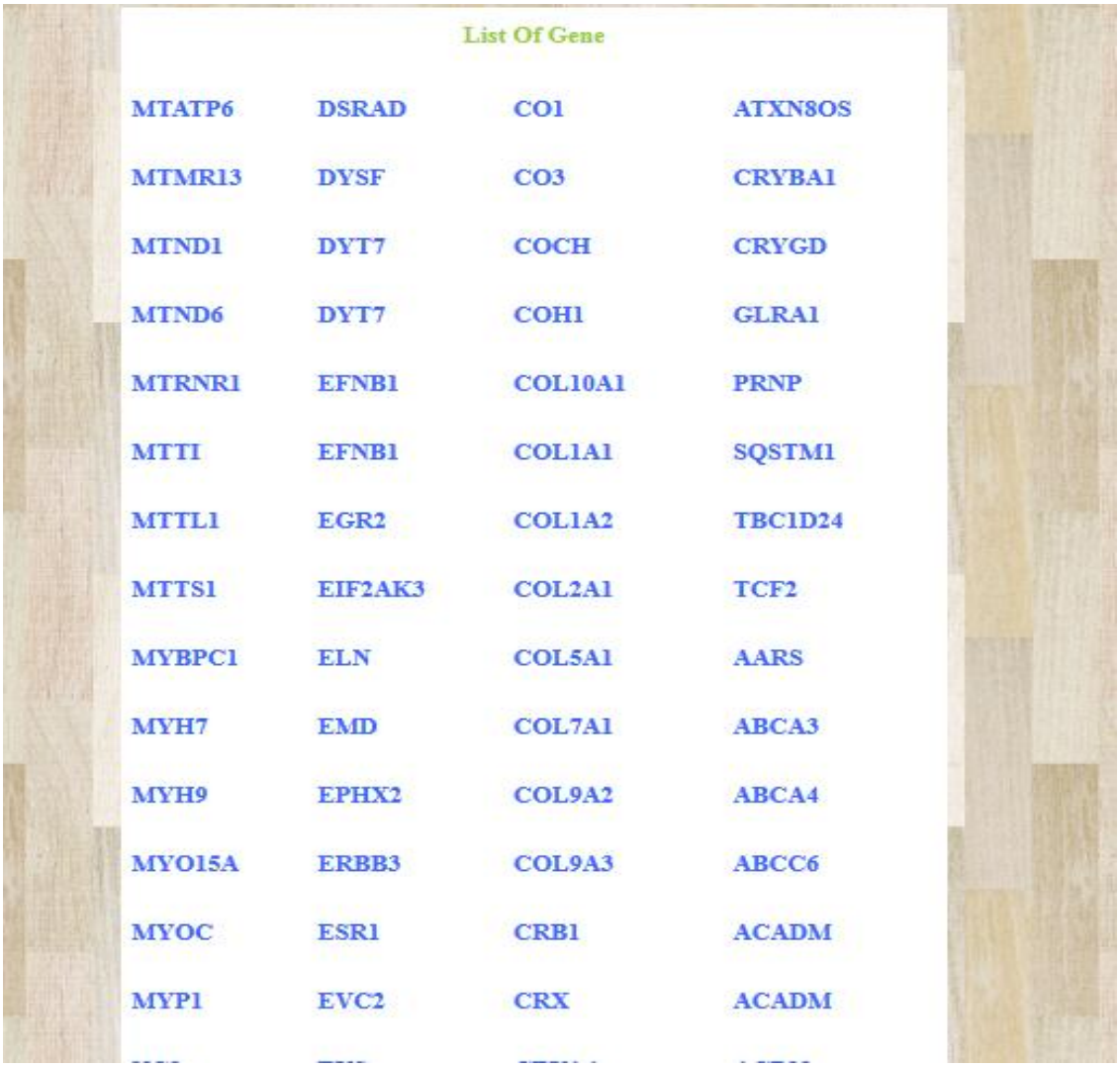
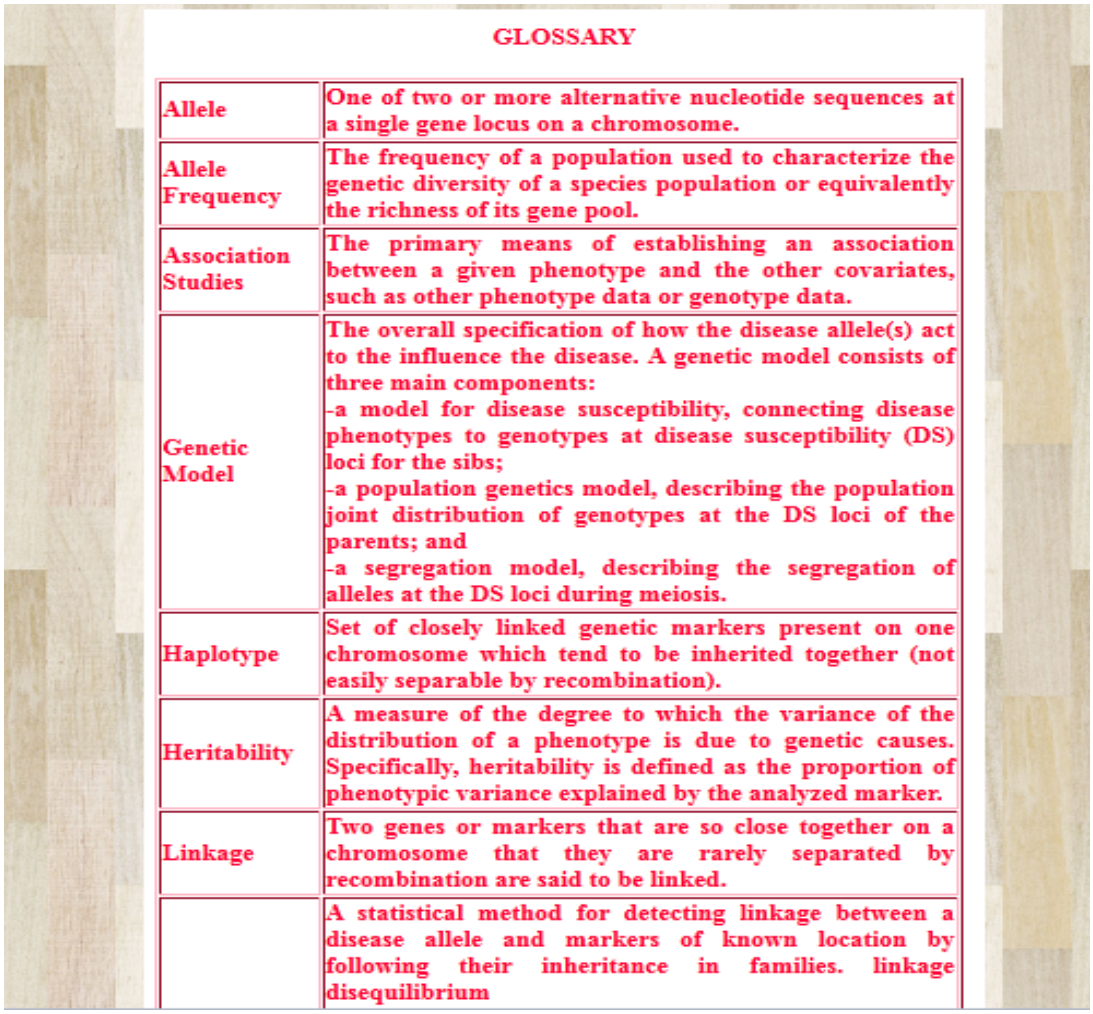


Figure 5.7 The list of genes.

5.2.4 The Glossary page

The Glossary page has also been designed to make users familiar to the terms that they can come across while going through the literature of human genetic disorders. The page contains 14 such terms.



GLOSSARY

Allele	One of two or more alternative nucleotide sequences at a single gene locus on a chromosome.
Allele Frequency	The frequency of a population used to characterize the genetic diversity of a species population or equivalently the richness of its gene pool.
Association Studies	The primary means of establishing an association between a given phenotype and the other covariates, such as other phenotype data or genotype data.
Genetic Model	The overall specification of how the disease allele(s) act to the influence the disease. A genetic model consists of three main components: -a model for disease susceptibility, connecting disease phenotypes to genotypes at disease susceptibility (DS) loci for the sibs; -a population genetics model, describing the population joint distribution of genotypes at the DS loci of the parents; and -a segregation model, describing the segregation of alleles at the DS loci during meiosis.
Haplotype	Set of closely linked genetic markers present on one chromosome which tend to be inherited together (not easily separable by recombination).
Heritability	A measure of the degree to which the variance of the distribution of a phenotype is due to genetic causes. Specifically, heritability is defined as the proportion of phenotypic variance explained by the analyzed marker.
Linkage	Two genes or markers that are so close together on a chromosome that they are rarely separated by recombination are said to be linked.
	A statistical method for detecting linkage between a disease allele and markers of known location by following their inheritance in families. linkage disequilibrium

Figure 5.8 The Glossary Page.

5.2.5 The Contacts Page

For any further queries the users can mail at the specified E-mail ID given on the contact page.

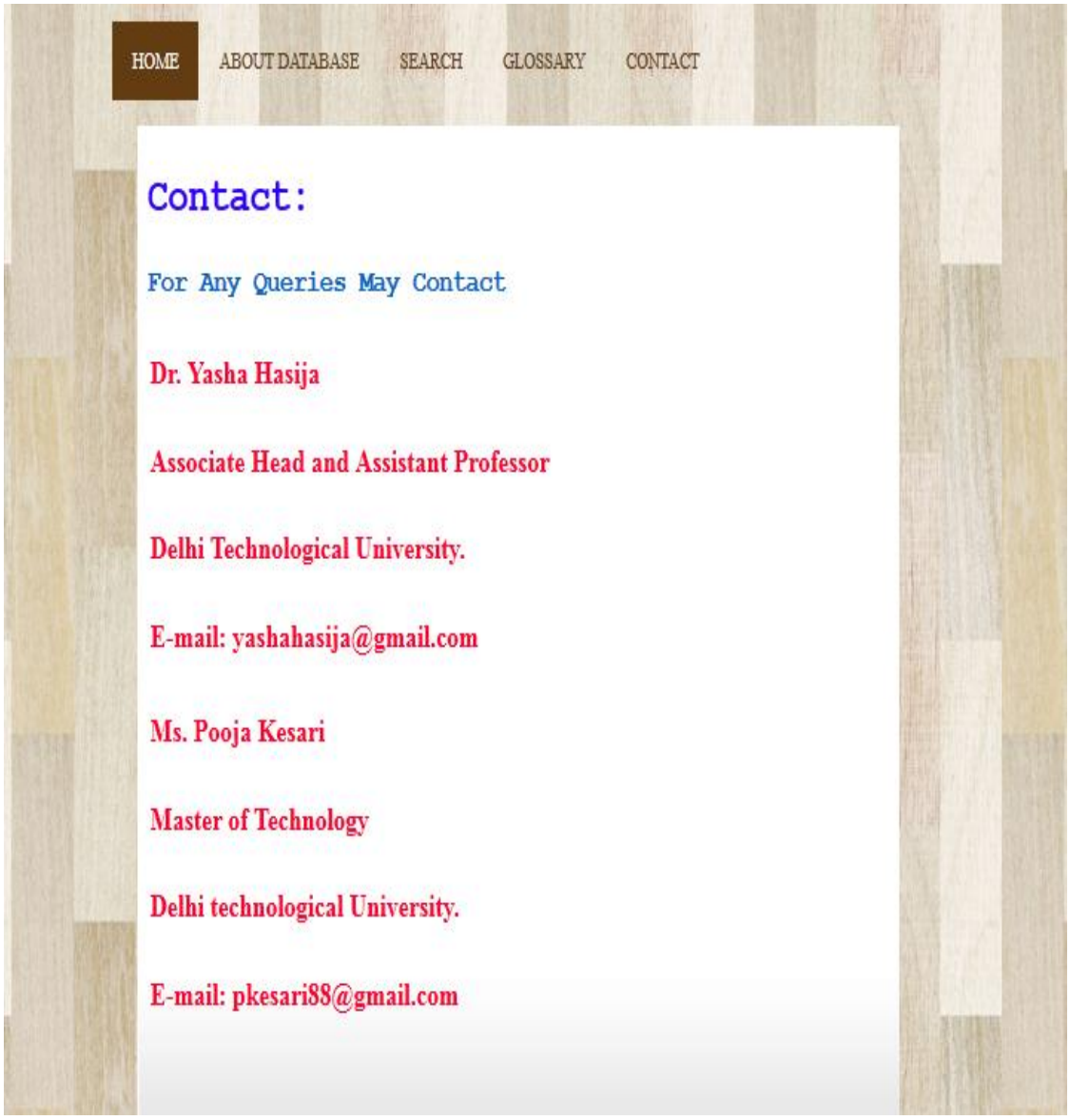


Figure 5.9 The Contact Page.

5.3 The steps for querying dbPedigree

- **Disease-Wise Search**

1. Select the **disease** radio button.

Figure 5.10 Disease radio button.

2. Specify the name of the disease and click on the **search** button.

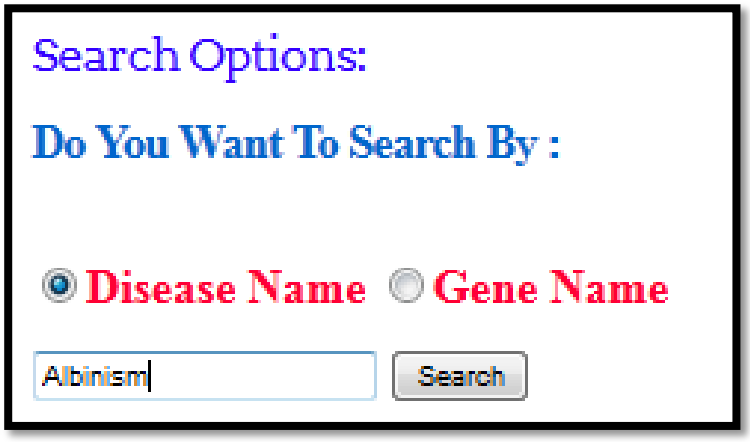


Figure 5.11 Searching "Albinism" in the database.

3. The result would be shown in a tabular format containing different column like sub-disease category (if any),type of disorder, other disease included during the pedigree study, causative gene along with its locus, associated mutation ,SNP, ethnicity of the family ,pedigree number and name, p-value, LOD score and the reference of the literature .The result contains 23 records against "Albinism".

Sub Disease Name	Disease Type	Disease Include	Gene Name	Gene locus	Associated Mutation	rs_id	Ethnicity	pedigree number	Link to pedigree	Pedigree Name	P-value	LOD Score	Citation
Ocular Albinism Type I	X-Linked		GPR143	Xp22.2	14-BP DEL; NT816			PID1578	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1724103/figure/F1/				11520764
Albinism Oculocutaneous Type IB	Autosomal Recessive		TYR	11q14.3	PRO406LEU	rs104894313	FRENCH;USA;Swiss	PID1642	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1706616/?page=2				5516239
Albinism Oculocutaneous Type IB	Autosomal Recessive	Waardenburg Syndrome 2	TYR	11q14.3	ARG402GLN	rs1126809		PID1643	http://hmg.oxfordjournals.org/content/6/5/659.long				9158138
Albinism Oculocutaneous Type IA	Autosomal Recessive		TYR	11q14.3	ARG59GLN	rs61753185		PID1644	http://www.jbc.org/content/265/29/17792.long				2120217
Temperature-Sensitive Oculocutaneous Albinism	Autosomal Recessive		TYR	11q14.3	ARG402GLN	rs1126809	Caucasian	PID1645	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC329899/?page=2				1900307
Ocular Albinism Type 1	X-Linked		GPR143	Xp22.2	CYS116TRP		Canadian	PID2350	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1468396/figure/F1/	pedigree 1I			PMC1468396
Ocular Albinism Type 1	X-Linked		GPR143	Xp22.2	THR166ASN		FRENCH	PID2351	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1468396/figure/F1/	pedigree 1L			PMC1468396
Ocular Albinism Type 1	X-Linked		GPR143	Xp22.2	163_170dup		FRENCH	PID2352	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1468396/figure/F2/	pedigree 2B			PMC1468396

Figure 5.12 Results of search using keyword "Albinism".

4. The pedigree result is sorted by the pedigree number. The pedigree image can be opened by right clicking the link and opening it in a new window.

Sub Disease Name	Disease Type	Disease Include	Gene Name	Gene locus	Associated Mutation	rs_id	Ethnicity	pedigree number	Link to pedigree	Pedigree Name	P-value	LOD Score
Ocular Albinism Type I	X-Linked		GPR143	Xp22.2	14-BP DEL; NT816			PID1578	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1724103/figure/F1/			
Albinism Oculocutaneous Type IB	Autosomal Recessive		TYR	11q14.3	PRO406LEU	rs104894313	FRENCH;USA;Swiss	PID1642	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1706616/?page=2			
Albinism Oculocutaneous Type IB	Autosomal Recessive	Waardenburg Syndrome 2	TYR	11q14.3	ARG402GLN	rs1126809		PID1643	http://hm/content/6			
Albinism Oculocutaneous Type IA	Autosomal Recessive		TYR	11q14.3	ARG59GLN	rs61753185		PID1644	http://ww/265/29/1			
Temperature-Sensitive Oculocutaneous Albinism	Autosomal Recessive		TYR	11q14.3	ARG402GLN	rs1126809	Caucasian	PID1645	http://ww/pmc/arti/?page=2			
Ocular Albinism Type 1	X-Linked		GPR143	Xp22.2	CYS116TRP		Canadian	PID2350	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1468396/figure/F1/	pedigree 1I		

Figure 5.13 Right clicking the link to open the "Albinism"pedigree.

5. **Pedigree Image**

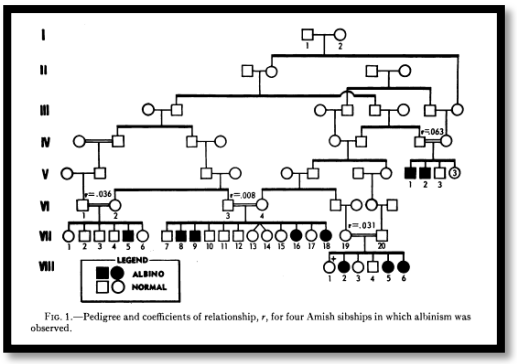


Figure 5.14 The pedigree image of Albinism.

- **Gene-Wise Search**

1 Select the **Gene** radio button.

.

Figure 5.15 Gene radio button.

2 Specify the name of the disease and click on the **search** button.

.

Figure 5.16 Searching "ALK" gene in the database.

3. The result would be shown in a tabular format containing different column like disease category, type of disorder, other disease included during the pedigree study, causative gene along with its locus, associated mutation, SNP, ethnicity of the family, pedigree number and name, p-value, LOD score and the reference of the literature .The result contains 8 records against "ALK" gene.

Sub Disease Name	Disease Type	Disease Include	Gene Name	Gene locus	Associated Mutation	rs_id	Ethnicity	pedigree number	Link to pedigree	Pedigree Name	P-value	LOD Score	Citation
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	GLY1128ALA c.3383G>C	rs113994088		PID422	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2672043/figure/F1/	FBN52			PMC2672043
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	ARG1192PRO c.3575G>C	rs113994089		PID423	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2672043/figure/F1/	FNB2			PMC2672043
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	ARG1275GLN c.3824G>A	rs113994087		PID424	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2672043/figure/F1/	FNB6			PMC2672043
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	ARG1275GLN c.3824G>A	rs113994087		PID425	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2672043/figure/F1/	FBN11			PMC2672043
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	ARG1275GLN c.3824G>A	rs113994087		PID426	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2672043/figure/F1/	FBN12			PMC2672043
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	ARG1275GLN c.3824G>A	rs113994087		PID427	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2672043/figure/F1/	FBN13			PMC2672043
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	ARG1275GLN c.3824G>A	rs113994087		PID428	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2672043/figure/F1/	FBN56			PMC2672043
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	ARG1192PRO c.3575G>C	rs113994089		PID429	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2672043/figure/F1/	FBN32			PMC2672043

Figure 5.17 Results of search using keyword "ALK".

4. The pedigree result is sorted by the pedigree number. The pedigree image can be opened by right clicking the link and opening it in a new window.

Sub Disease Name	Disease Type	Disease Include	Gene Name	Gene locus	Associated Mutation	rs_id	Ethnicity	pedigree number	Link to pedigree	Pedigree Name	P-value	LOD Score	Citation
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	GLY1128ALA c.3383G>C	rs113994088		PID422	http://www.ncbi.nlm.nih.gov /pmc/articles/PMC2672043 /figure/F1/	FBN52			PMC267204
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	ARG1192PRO c.3575G>C	rs113994089		PID423	http://www.ncbi.nlm.nih.gov /pmc/art /figure/F				PMC267204
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	ARG1275GLN c.3824G>A	rs113994087		PID424	http://w /pmc/art /figure/F				PMC267204
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	ARG1275GLN c.3824G>A	rs113994087		PID425	http://w /pmc/art /figure/F				PMC267204
Neuroblastoma Type 3	Autosomal Dominant		ALK	2p23.2-p23.1	ARG1275GLN c.3824G>A	rs113994087		PID426	http://w /pmc/articles/PMC2672043 /figure/F1/	FBN12			PMC267204

Figure 5. 18 Right clicking the link to open the "ALK" pedigree.

5 Pedigree Image

.

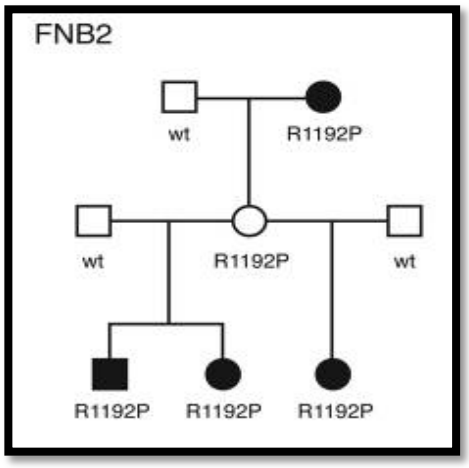


Figure 5.19 The pedigree image of "ALK" gene.

In case the disease or the gene name does not exist in the database. An error message would be shown.

Figure 5.20 The Error message.

5.4 The Working of PHP Script

The PHP script receives the query form from the browser and retrieves the records from the database after filtering them as required by the user. The selections made by the user are then obtained from the filtered records and sent to the server which then sends them to the browser. The detailed procedure is as follows:

1. Connection is made with the database using command **mysql_connect**. This requires the database name, the hostname and port as well as the username and password that allows one to access the localhost.
2. Connection is made with the database using command **mysql_select_db**. This requires the database name, the hostname and port as well as the username and password that allow one to access the database.
3. A variable will point to the selection made among the two radio button.
4. The SQL command is run, where the query will be executed for whichever keyword specified in the text box.

   In case query is created using disease name, first it would be searched in DiseaseTable and the diseaseID would be recorded. Then records would be selected from maintable against that diseaseID.

   Similar is the process for Gene-wise search.
5. The selected records would be displayed in the HTML table.