Topics

* Pubmed, the database of biomedical scientific literature

* How to travel between
    * Pubmed and sequence databases

* For what kind scientific and practical questions is database information exploited

* Microbe genomes covered by database information

Example: **You want to know, how SNP database information has recently been exploited in scientific literature**

Go to Pubmed, for example google ´Pubmed´
Pubmed –query: type ´SNP database´

You´ll get a list of papers, the first is the most recent one etc.
423 scientific papers were captured by ´SNP database´, the most recent one has been published two weeks ago (Jan 8)

1 - 20 of 423
of 22Next

1:
Analysis of the MTHFD1 promoter and risk of neural tube defects.
Carroll N, Pangilinan F, Molloy AM, Troendle J, Mills JL, Kirke PN, Brody LC, Scott JM, Parle-McDermott A.
Hum Genet. 2009 Jan 8. [Epub ahead of print]
PMID: 19130090 [PubMed - as supplied by publisher]
Related Articles

2:
An integrated database-pipeline system for studying single nucleotide polymorphisms and diseases.
Yang JO, Hwang S, Oh J, Bhak J, Sohn TK.
BMC Bioinformatics. 2008 Dec 12;9 Suppl 12:S19.
PMID: 19091018 [PubMed - in process]
Related Articles

3:
OpenADAM: an open source genome-wide association data management system for Affymetrix SNP arrays.
Yeung JM, Sham PC, Chan AS, Cherny SS.
BMC Genomics. 2008 Dec 31;9(1):636. [Epub ahead of print]
PMID: 19117518 [PubMed - as supplied by publisher]
Related Articles

4:
MedRefSNP: a database of medically investigated SNPs.
Rhee H, Lee JS.
Hum Mutat. 2008 Dec 22. [Epub ahead of print]
PMID: 19105187 [PubMed - as supplied by publisher]
Related Articles

5:
SNPnexus: A web database for functional annotation of newly discovered and public domain Single Nucleotide Polymorphisms.
Chelala C, Khan A, Lemoine NR.
Bioinformatics. 2008 Dec 19. [Epub ahead of print]
PMID: 19098027 [PubMed - as supplied by publisher]
Related Articles

An Analysis Pipeline for Genome-wide Association Studies.
Stefanov S, Lautenberger J, Gold B.
Cancer Inform. 2008 Sep 24;6:455-461.
PMID: 19096721 [PubMed]
Related Articles Free article in PMC

7:
The Pig Genome Database (PiGenome): an integrated database for pig genome research.
Lim D, Cho YM, Lee KT, Kang Y, Sung S, Nam J, Park EW, Oh SJ, Im SK, Kim H.
Mamm Genome. 2008 Dec 10. [Epub ahead of print]
PMID: 19082661 [PubMed - as supplied by publisher]
Related Articles

8:
Association of serum interleukin-33 level and the interleukin-33 genetic variant with Japanese cedar pollinosis.
Sakashita M, Yoshimoto T, Hirota T, Harada M, Okubo K, Osawa Y, Fujieda S, Nakamura Y, Yasuda K, Nakanishi K, Tamari M. etc. (423 items)
Clin Exp Allergy. 2008 Dec;38(12):1875-81.

You can open the papers by clicking them.....

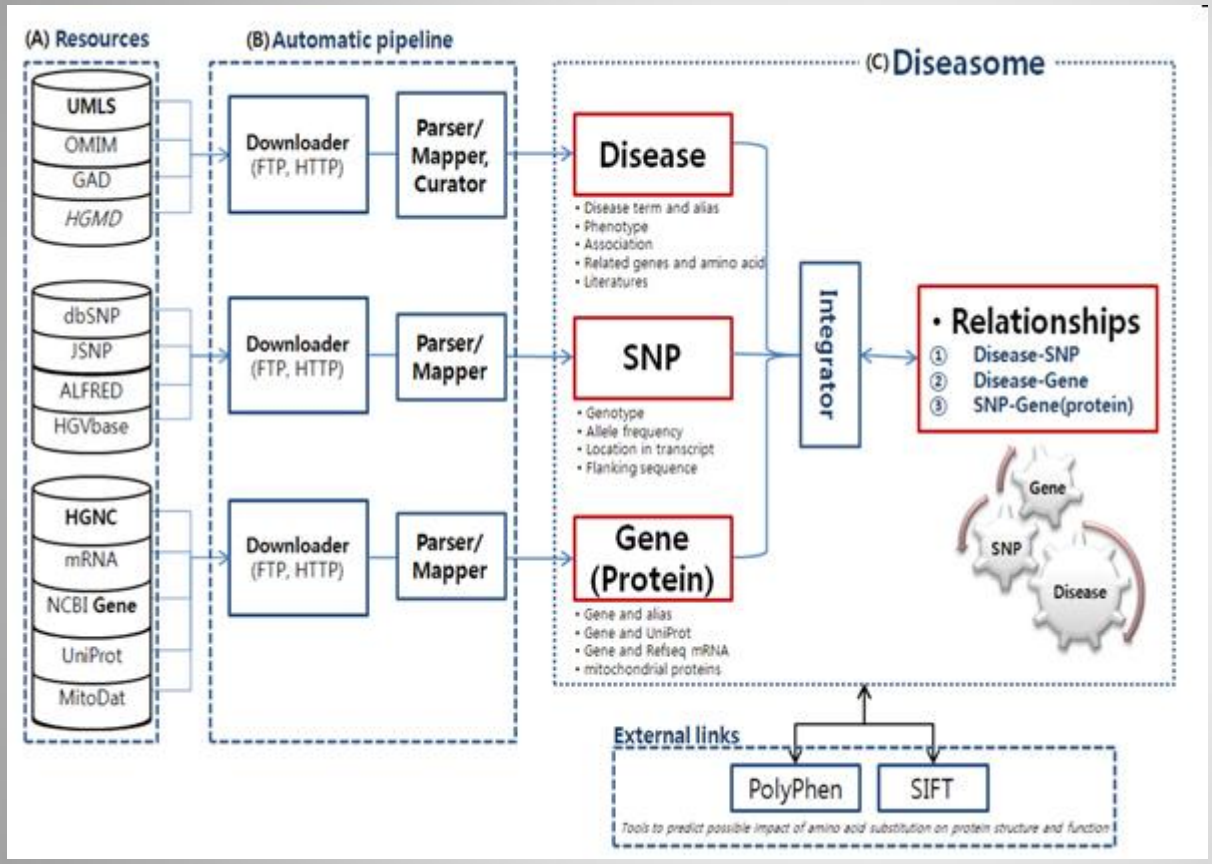
Some examples:

An integrated database-pipeline system for studying single nucleotide polymorphisms and diseases
Jin Ok Yang* et al.
BMC Bioinformatics 2008, **9**(Suppl 12)**:**S19 doi:10.1186/1471-2105-9-S12-S19

Background
Studies on the relationship between disease and genetic variations such as single nucleotide polymorphisms (SNPs) are important. Genetic variations can cause disease by influencing important biological regulation processes. Despite the needs for analyzing SNP and disease correlation, most existing databases provide information only on functional variants at specific locations on the genome, or deal with only a few genes associated with disease. There is no combined resource to widely support gene-, SNP-, and disease-related information, and to capture relationships among such data. Therefore, we developed an integrated database-pipeline system for studying SNPs and diseases.
Results
To implement the pipeline system for the integrated database, we first unified complicated and redundant disease terms and gene names using the Unified Medical Language System (UMLS) for classification and noun modification, and the HUGO Gene Nomenclature Committee (HGNC) and NCBI gene databases. Next, we collected and integrated representative databases for three categories of information. For genes and proteins, we examined the NCBI mRNA, UniProt, UCSC Table Track and MitoDat databases. For genetic variants we used the dbSNP, JSNP, ALFRED, and HGVbase databases. For disease, we employed OMIM, GAD, and HGMD databases. The database-pipeline system provides a disease thesaurus, including genes and SNPs associated with disease. The search results for these categories are available on the web page http://diseasome.kobic.re.kr/ webcite, and a genome browser is also available to highlight findings, as well as to permit the convenient review of potentially deleterious SNPs among genes strongly associated with specific diseases and clinical phenotypes.

Conclusion

Our system is designed to capture the relationships between SNPs associated with disease and disease-causing genes. The integrated database-pipeline provides a list of candidate genes and SNP markers for evaluation in both epidemiological and molecular biological approaches to diseases-gene association studies. Furthermore, researchers then can decide semi-automatically the data set for association studies while considering the relationships between genetic variation and diseases. The database can also be economical for disease-association studies, as well as to facilitate an understanding of the processes which cause disease. Currently, the database contains 14,674 SNP records and 109,715 gene records associated with human diseases and it is updated at regular intervals.



Overview of the integrated database-pipeline system. Rectangles represent computational applications, and are three in number. The Resource (A) contains gene-, SNP-, and disease-related primary resources and constructs a primary information database. The Automatic pipeline (B) retrieves information from primary databases and extracts essential gene-, SNP-, and disease-related data. We mapped disease terms and aliases, or gene names and aliases, based on the UMLS and HGNC databases. Also, disease terms were corrected for noun modification, stop word, and suffix. SNP effects were investigated by amino acid substitution; locations are available. The Diseasome (C) is a database including three categories of information (gene, SNP, and disease), and relationships among the three categories.

Gene Information

Gene Symbol	BRCA1
Gene Aliases	BRCAI, BRCC1, IRIS, PSCP, RNF53
Gene Name	breast cancer 1, early onset
Gene ID	672
Cytogenetic Band	17q21
Gene Type	protein-coding
mitoDat ID	-

Disease Information

Disease Hit	9
Disease Name	HIJATZ TUMORE GAIZTOA BRCA1 Syndrome Ovarian Carcinoma Malignant Neoplasms Endometrial Carcinoma Colorectal Cancer Germ-Line Mutation Endarterectomy, Carotid Fallopian Tube Carcinoma
Others	no cui diseases

Transcription Information

No	mRNA Accession	Chromosome Position	Strand	Exon Count	Promoter SNP Count	5'UTR SNP Count	CDS SNP Count	3'UTR SNP Count	Intron SNP Count
1	NM_007294	chr17:38449840-38530994	-	23	22	3	71	11	440
2	NM_007295	chr17:38449840-38530657	-	23	23	1	71	11	438
3	NM_007296	chr17:38449840-38530994	-	23	22	3	71	11	440
4	NM_007297	chr17:38449840-38530994	-	15	22	6	62	11	446
5	NM_007298	chr17:38449840-38530994	-	20	22	3	29	11	482
6	NM_007299	chr17:38449840-38530994	-	19	22	3	59	11	452
7	NM_007300	chr17:38449840-38530994	-	18	22	3	58	11	453
8	NM_007302	chr17:38449840-38530994	-	21	22	3	70	11	441
9	NM_007303	chr17:38449840-38530994	-	22	22	3	30	11	481
10	NM_007304	chr17:38449840-38530994	-	23	22	3	32	11	479
11	NM_007305	chr17:38449840-38530994	-	21	22	3	31	11	480

SNP Information

SNP ID	Chromosome	Chromosome Position	Strand	Allele	Fuction Class
rs799905	17	38530713	+	C/G	Intron, Promoter
rs34191881	17	38530897..38530898	+	-/A	5UTR, Promoter
rs35436937	17	38530928..38530929	+	-/T	5UTR, Promoter
rs8176075	17	38530940	-	-/T	5UTR, Promoter
rs8176074	17	38531291	-	A/G	Promoter
rs8176073	17	38531359	-	A/G	Promoter
rs8176072	17	38531470	-	A/T	Promoter
rs3092986	17	38531522	-	A/G	Promoter
rs34095552	17	38531530..38531531	+	-/GT	Promoter
rs8176071	17	38531531..38531532	-	-/ACA	Promoter
rs799906	17	38531642	+	C/T	Promoter
rs11655505	17	38531903	+	A/G	Promoter
rs799907	17	38532251	+	C/G	Promoter
rs799908	17	38532442	+	A/G	Promoter
rs799909	17	38532753	+	A/G	Promoter

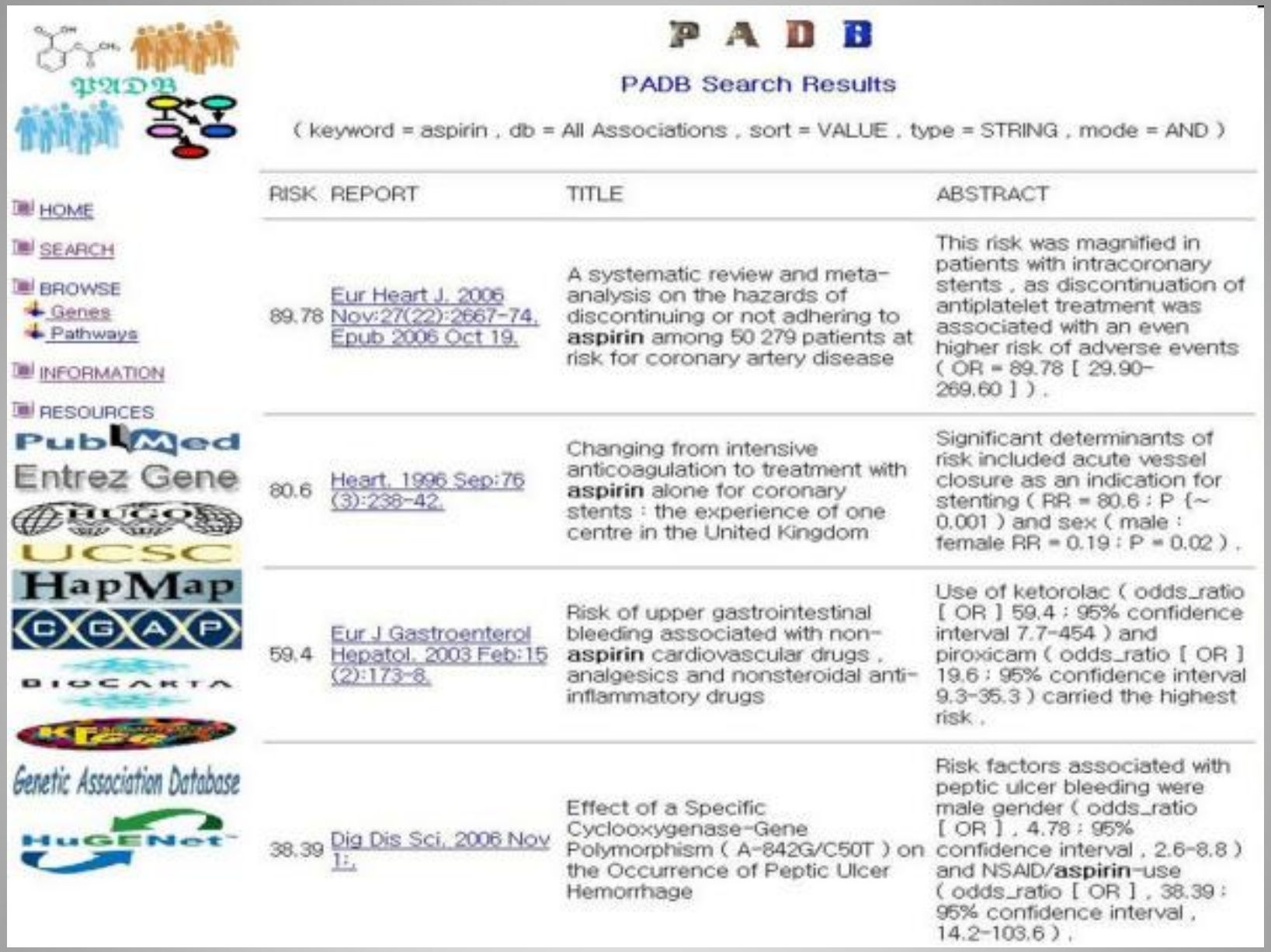
Query table results and graphic viewer. The retrieval page of the integrated gene, SNP, and diseases database. The information on diseases, genes, and SNP markers found as result of a query (e.g., BRCA1) are shown. When a user queries a gene symbol, the system retrieves the Gene Information table, which shows various gene annotations, disease information related to the queried gene, transcript information including the number of SNPs located in each transcript, and SNP information associated with the queried gene. In addition, the user can explore the data on gene-related transcripts, SNPs, and disease information, using the genome browser. If a user requires more specific information on any item, the user can click on a disease term, a gene ID, or a genetic variation number (SNP rs number).

When you open a paper, Pubmed also gives you additional information (right window) about related papers, about the scientists, about papers which have referenced to the paper you are looking

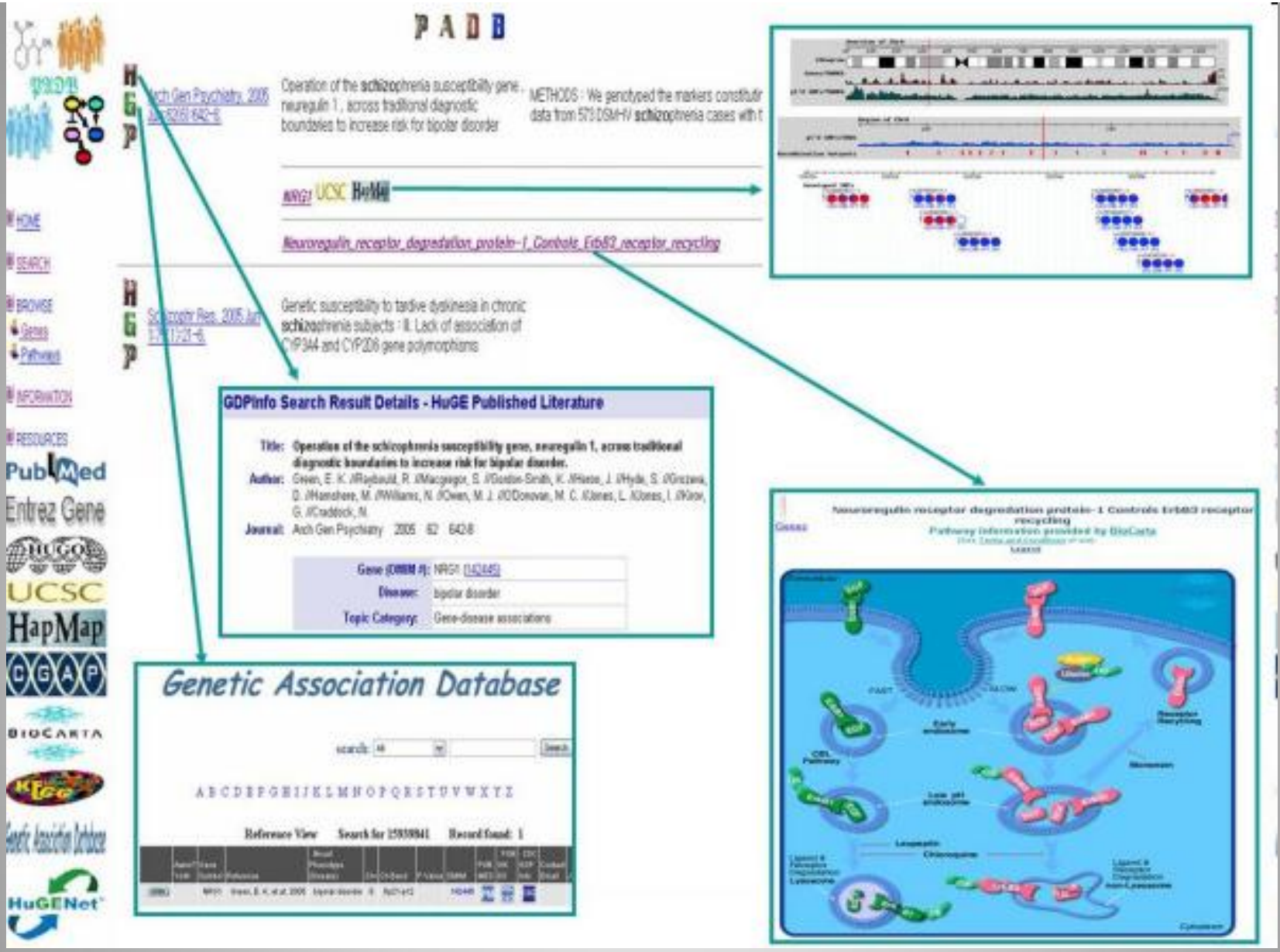
For example, when you opened the previous paper by Yang et al., you also received a hint from Pubmed about the the paper:

PADB : Published Association Database
Hwanseok Rhee and **Jin-Sung Lee**
BMC Bioinformatics 2007, ***8:****348 doi:10.1186/1471-2105-8-348*

Background
Although molecular pathway information and the International HapMap Project data can help biomedical researchers to investigate the aetiology of complex diseases more effectively, such information is missing or insufficient in current genetic association databases. In addition, only a few of the environmental risk factors are included as gene-environment interactions, and the risk measures of associations are not indexed in any association databases.
Description
We have developed a published association database (PADB; http://www.medclue.com/padb webcite) that includes both the genetic associations and the environmental risk factors available in PubMed database. Each genetic risk factor is linked to a molecular pathway database and the HapMap database through human gene symbols identified in the abstracts. And the risk measures such as odds ratios or hazard ratios are extracted automatically from the abstracts when available. Thus, users can review the association data sorted by the risk measures, and genetic associations can be grouped by human genes or molecular pathways. The search results can also be saved to tab-delimited text files for further sorting or analysis. Currently, PADB indexes more than 1,500,000 PubMed abstracts that include 3442 human genes, 461 molecular pathways and about 190,000 risk measures ranging from 0.00001 to 4878.9.
Conclusion
PADB is a unique online database of published associations that will serve as a novel and powerful resource for reviewing and interpreting huge association data of complex human diseases.



Sorting associations by risk measures. PADB automatically extracts the odds ratio, hazard ratio, risk ratio and relative risk data if they are available in sentences. When multiple associations are reported in a single sentence, those multiple association data are indexed as separate records.

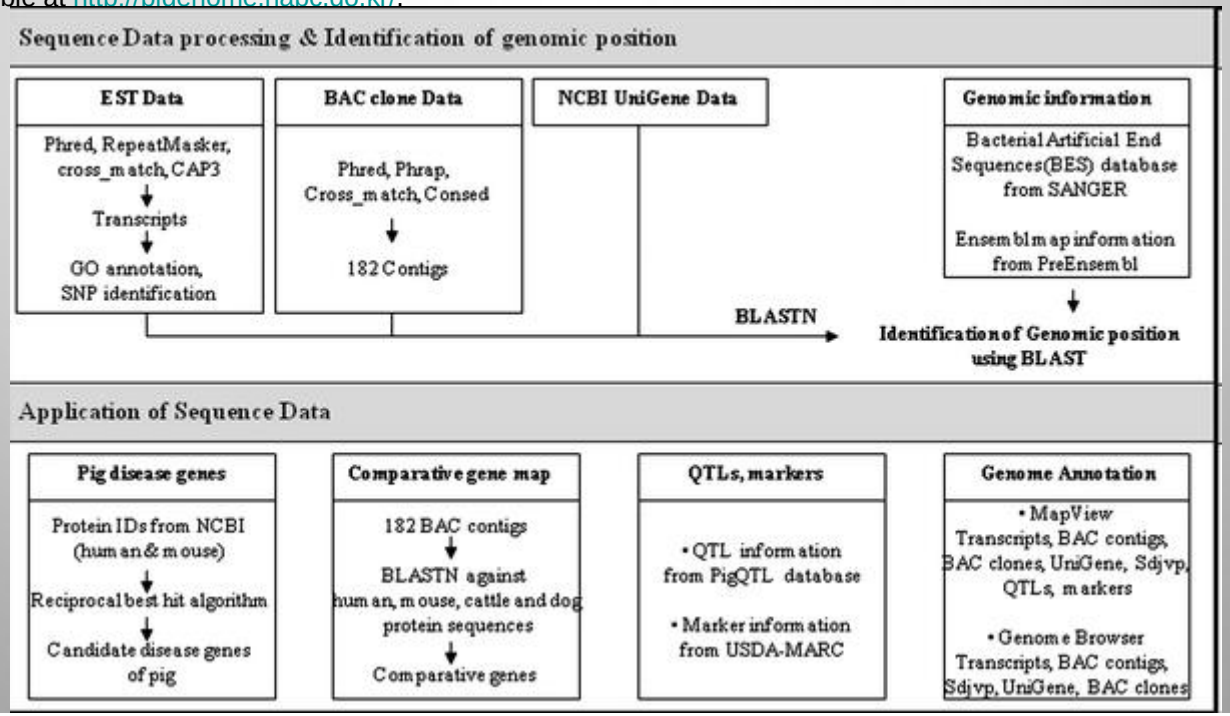


Linking genetic risks to molecular pathway and HapMap information. PADB can help biomedical researchers to review and interpret genetic risk factors more effectively along with molecular pathway and HapMap information.

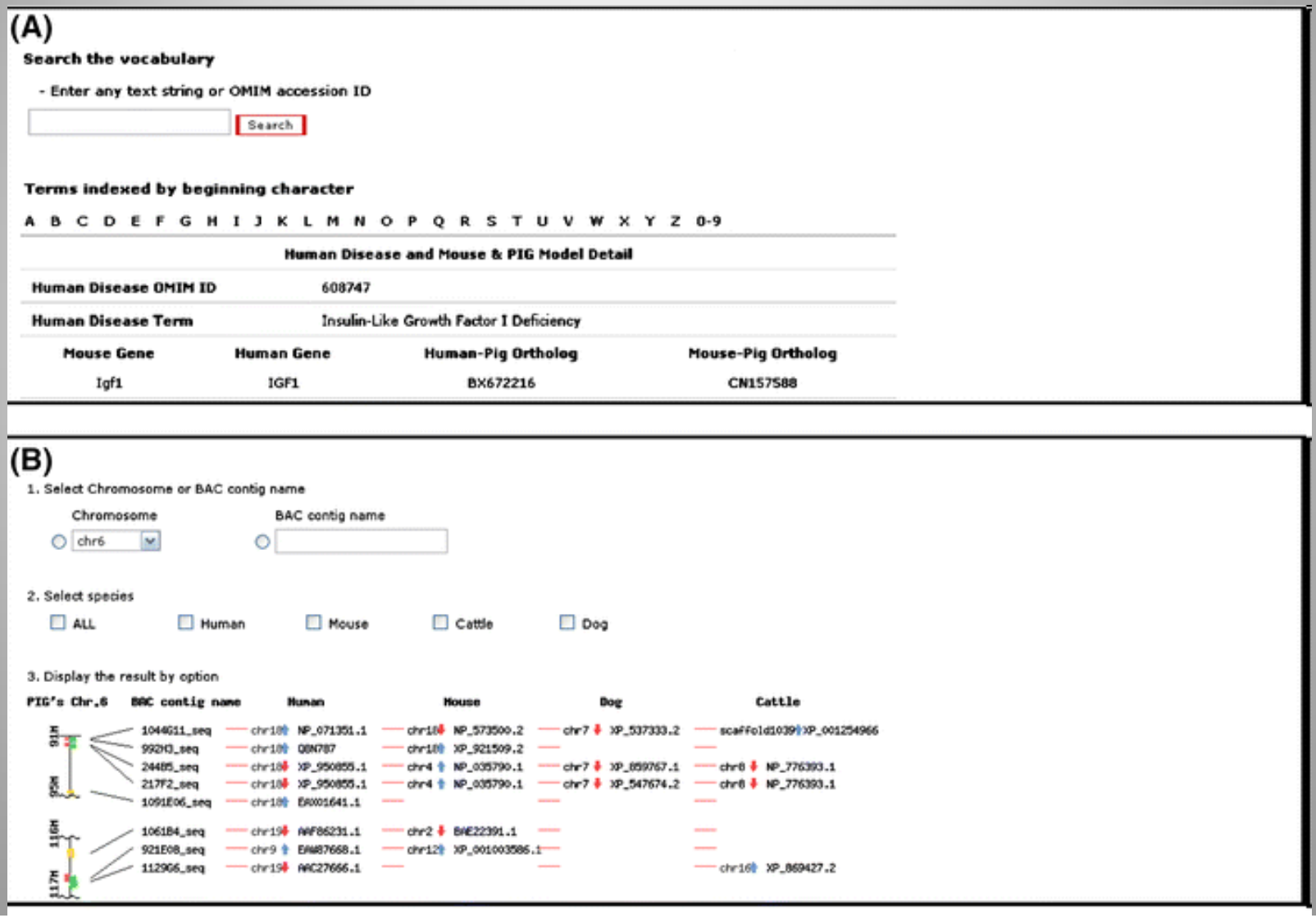
From the original list (page 2) you´ll find that SNP databases are not solely human databases, in fact SNP databases exist for many organisms, for example economically important animals:

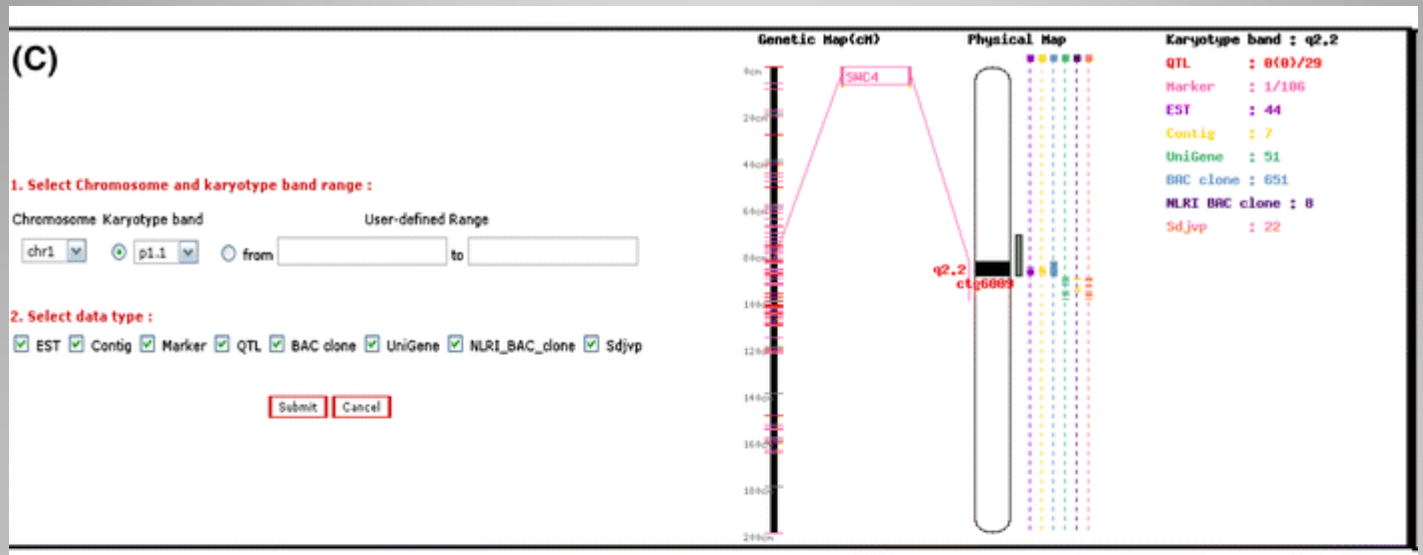
The Pig Genome Database (PiGenome): an integrated database for pig genome research
Dajeong Lim et al.
Mammalian Genome 2009,

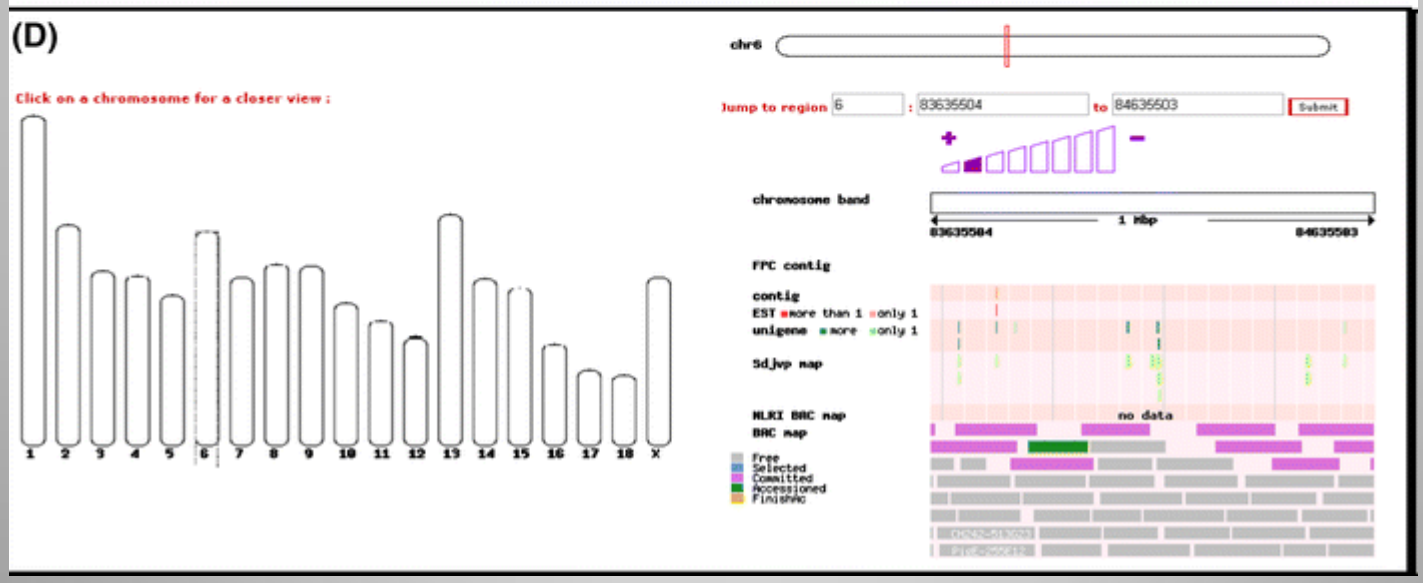
We established the Pig Genome Database (PiGenome) for pig genome research. The PiGenome integrates and analyzes all publicly available genome-wide data on pigs, including UniGenes, sequence tagged sites (STS) markers, quantitative trait loci (QTLs) data, and bacterial artificial chromosome (BAC) contigs. In addition, we produced 69,545 expressed sequence tags (ESTs) from the full-length enriched cDNA libraries of six tissues and 182 BAC contig sequences, which are also included in the database. QTLs, genetic markers, and BAC end-sequencing information were collected from public databases. The full-length enriched EST data were clustered and assembled into unique sequences, contigs, and singletons. The PiGenome provides functional annotation, identification of transcripts, mapping of coding sequences, and SNP information. It also provides an advanced search interface, a disease browser, alternative-splicing events, and a comparative gene map of the pig. A graphical map view and genome browser can map ESTs, contigs, BAC contigs (from the National Institute of Animal Science), Sino-Danish Pig Genome Project transcripts, and UniGene onto pig genome sequences which include our 182 BAC contigs and publically available BAC sequences of the Wellcome Trust Sanger Institute. The PiGenome is accessible at http://pigenome.nabc.go.kr/.



Animal and human SNP databases are, to some extent, synergistic; the pig-paper (previous page) introduced, how human disease OMIM database can be exploited to give information about pigs




(C)
1. Select Chromosome and karyotype band range :
Chromosome
Karyotype band
User-defined Range
chr1
p1.1
from
to
2. Select data type :
EST
Contig
Marker
QTL
BAC clone
UniGene
NLRI_BAC_clone
Sdjvp
Submit
Cancel
Genetic Map(cM)
Physical Map
SMC4
q2.2
ctg6003
Karyotype band : q2.2
QTL : 0(0)/29
Marker : 1/106
EST : 44
Contig : 7
UniGene : 51
BAC clone : 651
NLRI BAC clone : 8
Sdjvp : 22

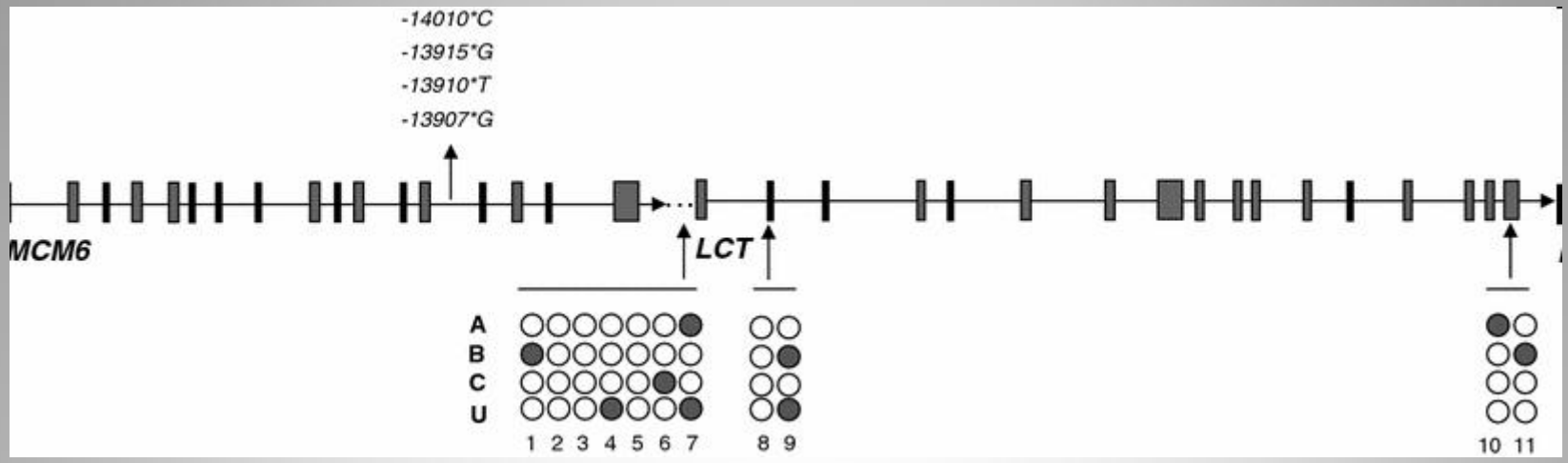

(D)
Click on a chromosome for a closer view :
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 X
chr6
Jump to region
6
83635504
to
84635503
Submit
chromosone band
1 Mbp
83635504
84635503
FPC contig
contig
EST more than 1 only 1
unigene more only 1
Sdjvp map
NLRI BAC map
no data
BAC map
Free
Selected
Committed
Accessioned
Finished

So, ´SNP database´ query gives information about recent progress in database design (previous examples), but also other kind of scientific stories, for example:

Lactose digestion and the evolutionary genetics of lactase persistence

Catherine J. E. Ingram et al.
Human Genetics (2009) 124:579-591

It has been known for some 40 years that lactase production persists into adult life in some people but not in others. However, the mechanism and evolutionary significance of this variation have proved more elusive, and continue to excite the interest of investigators from different disciplines. This genetically determined trait differs in frequency worldwide and is due to *cis*-acting polymorphism of regulation of lactase gene expression. A single nucleotide polymorphism located 13.9 kb upstream from the lactase gene (*C-13910 > T*) was proposed to be the cause, and the *−13910*T* allele, which is widespread in Europe was found to be located on a very extended haplotype of 500 kb or more. The long region of haplotype conservation reflects a recent origin, and this, together with high frequencies, is evidence of positive selection, but also means that *−13910*T* might be an associated marker, rather than being causal of lactase persistence itself. Doubt about function was increased when it was shown that the original SNP did not account for lactase persistence in most African populations. However, the recent discovery that there are several other SNPs associated with lactase persistence in close proximity (within 100 bp), and that they all reside in a piece of sequence that has enhancer function in vitro, does suggest that they may each be functional, and their occurrence on different haplotype backgrounds shows that several independent mutations led to lactase persistence. Here we provide access to a database of worldwide distributions of lactase persistence and of the *C-13910*T* allele, as well as reviewing lactase molecular and population genetics and the role of selection in determining present day distributions of the lactase persistence phenotype.

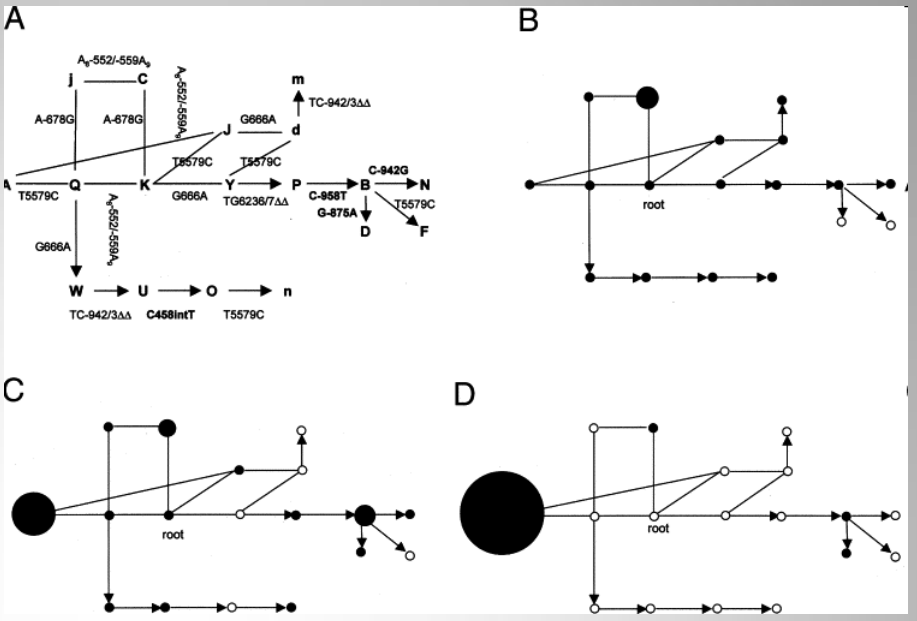


Diagrammatic representation of the genes *MCM6* and *LCT.* The *arrow* indicates the location of *−13910*T*, and the other alleles shown more recently to be associated with lactase persistence. Locations of SNPs used for *LCT* core haplotype analysis are shown, with the possible allelic combinations of the four common worldwide 11 SNP haplotypes described in Hollox et al. (*2001*). The *open circles* indicate an ancestral allele and *filled circles* denote the derived allele at a locus. SNPs used for assessing haplotype background of the lactase persistence associated variants in our own studies are 4, 6, 9 and 10

Reading this recent lactose paper, you can notice, for example, that they refer to Hollox et al. 2001 (of course, they refer to many other papers, too, but this is an example....), probably important background……. have a look…. click Hollox et al paper from the reference list (which is at the end of the Ingram et al. paper), you get the following paper, from which you can find what was known already 8 years ago etc.

Lactase Haplotype Diversity in the Old World
Edward J. Hollox et al.
Amer.J. Human Genet. 2001

Lactase persistence, the genetic trait in which intestinal lactase activity persists at childhood levels into adulthood, varies in frequency in different human populations, being most frequent in northern Europeans and certain African and Arabian nomadic tribes, who have a history of drinking fresh milk. Selection is likely to have played an important role in establishing these different frequencies since the development of agricultural pastoralism 9,000 years ago. We have previously shown that the element responsible for the lactase persistence/nonpersistence polymorphism in humans is *cis*-acting to the lactase gene and that lactase persistence is associated, in Europeans, with the most common 70-kb lactase haplotype, A. We report here a study of the 11-site haplotype in 1,338 chromosomes from 11 populations that differ in lactase persistence frequency. Our data show that haplotype diversity was generated both by point mutations and recombinations. The four globally common haplotypes (A, B, C, and U) are not closely related and have different distributions; the A haplotype is at high frequencies only in northern Europeans, where lactase persistence is common; and the U haplotype is virtually absent from Indo-European populations. Much more diversity is seen in sub-Saharan Africans than in non-Africans, consistent with an "Out of Africa" model for peopling of the Old World. Analysis of recent recombinant haplotypes by allele-specific PCR, along with deduction of the root haplotype from chimpanzee sequence, allowed construction of a haplotype network that assisted in evaluation of the relative roles of drift and selection in establishing the haplotype frequencies in the different populations. We suggest that genetic drift was important in shaping the general pattern of non-African haplotype diversity, with recent directional selection in northern Europeans for the haplotype associated with lactase persistence.



Lactase haplotype networks. *A,* Haplotype network showing probable phylogeny of the four common haplotypes (A, B, C, and U). Each line is annotated with its corresponding mutational change, and an arrow is shown where the directionality of the mutation is known. Mutational changes shown in bold are changes that occur only once in the network. *B,* Haplotype network, based on the framework of *A,* with circle size corresponding to the frequency of the haplotype in the population. An unblackened circle shows that none of that haplotype was observed in the population, and the smallest blackened circle represents frequencies of ≤.1. The sub-Saharan African populations are grouped and shown here, with 79% of total haplotype diversity represented in the diagram. *C,* As *B,* with non-African populations showing 92% of total non-African haplotype diversity represented in the diagram. Non-African excludes northern European. *D,* As *B,* with northern European populations showing 98% of total northern European haplotype diversity represented in the diagram.

What about the lactase gene sequence?

- *go to* NCBI
- select ´nucleotide´
  type ´human lactase´ :

This search in Gene shows 19 results, including:
LCT (*Homo sapiens*): lactase
LCTL (*Homo sapiens*): lactase-like
MCM6 (*Homo sapiens*): minichromosome maintenance complex component 6
1: NM_000155
Reports
Order cDNA clone, LinksHomo sapiens galactose-1-phosphate uridylyltransferase (GALT), mRNA
gi|22165415|ref|NM_000155.2|[22165415]

2: NG_008104
Reports
LinksHomo sapiens lactase (LCT) on chromosome 2
gi|193211369|ref|NG_008104.1|[193211369]

3: NM_000388
Reports
LinksHomo sapiens calcium-sensing receptor (CASR), mRNA
gi|189409146|ref|NM_000388.3|[189409146]

4: NM_014212
Reports
Order cDNA clone, LinksHomo sapiens homeobox C11 (HOXC11), mRNA
gi|84043954|ref|NM_014212.3|[84043954]

5: NM_002299 **TAKE THIS, CLICK THE ACCESSION NUMBER**
Reports
LinksHomo sapiens lactase (LCT), mRNA
gi|32481205|ref|NM_002299.2|[32481205]

- Comment
- Features
- Sequence

LOCUS NM_002299 6274 bp mRNA linear PRI 25-JAN-2009 DEFINITION Homo sapiens lactase (LCT), mRNA. ACCESSION NM_002299 VERSION NM_002299.2 GI:32481205 KEYWORDS . SOURCE Homo sapiens (human) ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo. REFERENCE 1 (bases 1 to 6274) AUTHORS Ingram,C.J., Mulcare,C.A., Itan,Y., Thomas,M.G. and Swallow,D.M. TITLE Lactose digestion and the evolutionary genetics of lactase persistence JOURNAL Hum. Genet. 124 (6), 579-591 (2009)

.....and other information concerning the occurrence of this particular sequence, NM_002299, in the literature
.....and after this, description of the gene structure (only the coding parts are here, this is from mRNA)

You are interested in the (known) relationships of this sequence to other animals,

- take the sequence in FASTA format (from the Display-window)

>gi|32481205|ref|NM_002299.2| Homo sapiens lactase (LCT), mRNA
GTTCCTAGAAAATGGAGCTGTCTTGGCATGTAGTCTTTATTGCCCTGCTAAGTTTTTCATGCTGGGGGTC
AGACTGGGAGTCTGATAGAAATTTCATTTCCACCGCTGGTCCTCTAACCAATGACTTGCTGCACAACCTG
AGTGGTCTCCTGGGAGACCAGAGTTCTAACTTTGTAGCAGGGGACAAAGACATGTATGTTTGTCACCAGC
CACTGCCCACTTTCCTGCCAGAATACTTCAGCAGTCTCCATGCCAGTCAGATCACCCATTATAAGGTATT……..

- copy-paste the sequence, and open the BLAST facilities:

Basic BLAST
Choose a BLAST program to run.
nucleotide blast Search a **nucleotide** database using a **nucleotide** query **Continue from here……..**
Algorithms: **blastn, megablast, discontiguous megablast**
protein blast Search **protein** database using a **protein** query
Algorithms: blastp, psi-blast, phi-blast
blastx Search **protein** database using a **translated nucleotide** query
tblastn Search **translated nucleotide** database using a **protein** query
tblastx Search **translated nucleotide** database using a **translated nucleotide** query

- Enter the sequence to the query window, and..... (orientate yourself by using the help-facilities....)

Sequences producing significant alignments: (Click headers to sort columns)
NM_002299.2Homo sapiens lactase (LCT), mRNA
X07994.1Human mRNA for lactase-phlorizin hydrolase LPH (EC 3.2.1.23-62)

.... (and other human sequences).

XM_001096426.1Macaca mulatta similar to lactase-phlorizin hydrolase preproprotein (LOC707761),
XR_024199.1Pan troglodytes similar to lactase-phlorizin hydrolase preproprotein (LCT), mRNA
XM_001915472.1Equus caballus similar to lactase phlorizinhydrolase (LOC100055369), mRNA
Z27166.1O.cuniculus (BL20) mRNA for-lactase-phlorizin hydrolase

XM_592166.3Bos taurus similar to lactase-phlorizin hydrolase preproprotein (LCT), mRNA
X07995.1Rabbit mRNA for lactase-phlorizin hydrolase LPH (EC 3.2.1.23-62)
AY191611.1Homo sapiens lactase-phlorizin hydrolase-1 (LCT) mRNA, partial cds
NM_001081078.1Mus musculus lactase (Lct), mRNA
XM_341115.3 Rattus norvegicus lactase (Lct), mRNA
XM_541018.2 Canis familiaris similar to lactase-phlorizin hydrolase preproprotein (LOC483898), mRNA
XM_001055600.1 Rattus norvegicus lactase, transcript variant 1 (Lct), mRNA
XM_001055660.1 Rattus norvegicus lactase, transcript variant 2 (Lct), mRNA
X56747.1Rat mRNA for fetal intestinal lactase-phlorizin hydrolase precursor, partial
AK158042.1Mus musculus adult inner ear cDNA, RIKEN full-length enriched library, clone:F930020G04
NM_001111346.1Gallus gallus lactase (LCT), mRNA

.....

-you got sequences from a monkey (Macaca), chimpanzee (Pan), horse (Equus), rabbit , cow (Bos), mouse (Mus), rat (Rattus), dog (Canis), chicken (Gallus), etc.

-and pairwise alignments .....

Alignment between the query (human lactase) and horse:

>ref|XM_001915472.1| _ PREDICTED: Equus caballus similar to lactase phlorizinhydrolase (LOC100055369), mRNA Length=5787
GENE ID: 100055369 LOC100055369 | hypothetical LOC100055369 [Equus caballus] Score = 7193 bits (7976), Expect = 0.0 I
dentities = 5059/5769 (87%), Gaps = 9/5769 (0%) Strand=Plus/Plus

```
Query 33 GTCTTTATTGCCCTGCTAAGTTTTTCATGCTGGGGGTCAGACTGGGAGTCTGATAGAAAT 92
         |||||||| | ||| ||||||||||||||||||||| ||||||||| |||||| ||||
Sbjct 22 GTCTTTATCGTCCTCCTAAGTTTTTCATGCTGGGGGTTAGACTGGGAATCTGATCCAAAT 81

Query 93 TTCATTTCCACCGCTGGTCCTCTAACCAATGACTTGCTGCACAACCTGAGTGGTCTCCTG 152
         |||||||| | || |||||| |||| ||||||||||||||| ||||||| |||| | |
Sbjct 82 TTCATTTCAGCTGCCGGTCCTTTAACGAATGACTTGCTGCACGACCTGAGCGGTCCACCG 141

Query 153 GGAGACCAGAGTTCTAACTTTGTAGCAGGGGACAAAGACATGTATGTTTGTCACCAGCCA 212
          ||| ||| | |||||||||||||||| || ||| | || |||||||||| |||||||
Sbjct 142 GGAAACCGGGATTCTAACTTTGTAGCAGAAGATAAAAATATTTATGTTTGTCCCCAGCCA 201
```

etc….

You can compare a set of sequences by collecting them in FASTA-format,
i.e. you construct a text file for yourself and continue working with this file
by using other programs ..... (for example clustering or phylogenetic analysis)

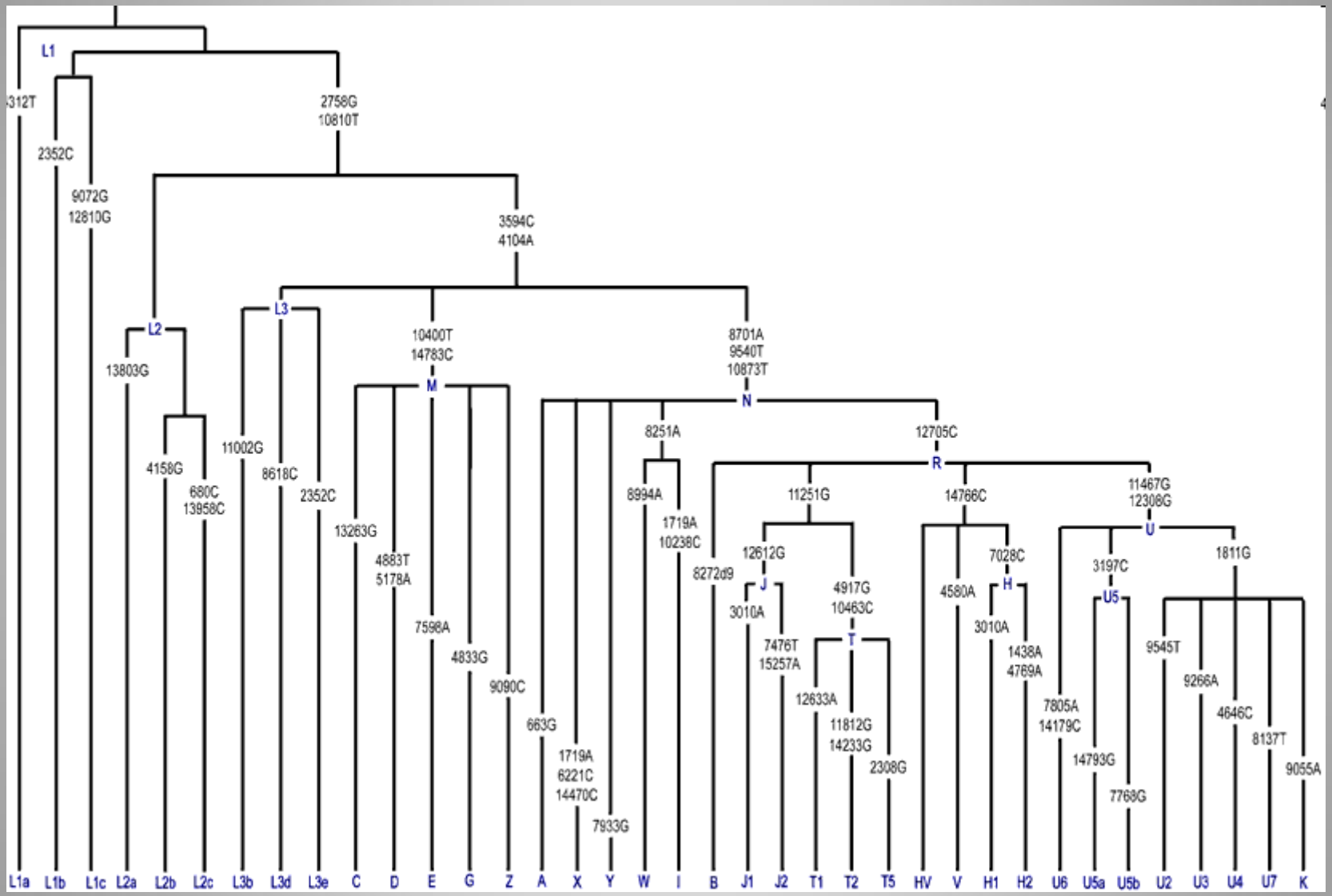
Note that you get additional information (not shown here) when you make a query,
for example links to specialized databases ....

Tracing evolutionary histories

* mitochondrial DNA (mtDNA) is of special importance

* sequence data from complete human mt-genomes in database:

http://www.genpat.uu.se/mtDB/


L1
3312T
2352C
9072G
12810G
2758G
10810T
3594C
4104A
L2
13803G
4158G
680C
13958C
L3
11002G
8618C
2352C
10400T
14783C
M
13263G
4883T
5178A
7598A
4833G
9090C
8701A
9540T
10873T
N
663G
1719A
6221C
14470C
7933G
8251A
8994A
1719A
10238C
12705C
R
8272d9
11251G
12612G
J
3010A
7476T
15257A
4917G
10463C
T
12633A
11812G
14233G
2308G
14766C
4580A
7028C
H
3010A
1438A
4769A
11467G
12308G
U
7805A
14179C
3197C
U5
14793G
7768G
1811G
9545T
9266A
4646C
8137T
9055A
L1a
L1b
L1c
L2a
L2b
L2c
L3b
L3d
L3e
C
D
E
G
Z
A
X
Y
W
I
B
J1
J2
T1
T2
T5
HV
V
H1
H2
U6
U5a
U5b
U2
U3
U4
U7
K

Exercise:

Pick up the first mt-sequence in the mtDNA database list

EF064321

NCBI Sequence Viewer v2.0 http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?tool=portal&db=nucc...

NCBI Nucleotide

My NCBI [Sign In] [Register]

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Nucleotide for EF064321 Go Clear

Display GenBank Show 20 Send to Hide: sequence all but gene, CDS and mRNA features

Range: from begin to end Reverse complemented strand Features: + Refresh

1: EF064321. Reports Homo sapiens isol...[gi:116517865] Links

Features Sequence

```
LOCUS       EF064321               16569 bp    DNA     circular PRI 15-DEC-2006
DEFINITION  Homo sapiens isolate 5_U6a1(Tor270) mitochondrion, complete genome.
ACCESSION   EF064321
VERSION     EF064321.1  GI:116517865
KEYWORDS    .
SOURCE      mitochondrion Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 16569)
  AUTHORS   Olivieri,A., Achilli,A., Pala,M., Battaglia,V., Fornarino,S.,
            Al-Zahery,N., Scozzari,R., Cruciani,F., Behar,D.M., Dugoujon,J.M.,
            Coudray,C., Santachiara-Benerecetti,A.S., Semino,O., Bandelt,H.J.
            and Torroni,A.
  TITLE     The mtDNA Legacy of the Levantine Early Upper Palaeolithic in
            Africa
  JOURNAL   Science 314 (5806), 1767-1770 (2006)
   PUBMED   17170302
REFERENCE   2  (bases 1 to 16569)
  AUTHORS   Olivieri,A., Achilli,A., pala,M., Battaglia,V., Fornarino,S.,
            Al-Zahery,N., Scozzari,R., Cruciani,F., Behar,D.M., Dugoujon,J.-M.,
            Coudray,C., Santachiara-Benerecetti,A.S., Semino,O., Bandelt,H.-J.
            and Torroni,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-2006) Dipartimento di Genetica e Microbiologia,
            Universita' di Pavia, via Ferrata, 1, Pavia, Pavia 27100, Italy
FEATURES             Location/Qualifiers
     source          1..16569
                     /organism="Homo sapiens"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /isolate="5_U6a1(Tor270)"
                     /db_xref="taxon:9606"
                     /haplotype="U6a1"
                     /country="Algeria"
     D-loop          join(16024..16569,1..577)
     tRNA            578..648
                     /product="tRNA-Phe"
     rRNA            649..1602
                     /product="12S ribosomal RNA"
     tRNA            1603..1671
                     /product="tRNA-Val"
     rRNA            1672..3229
                     /product="16S ribosomal RNA"
     tRNA            3230..3304
                     /product="tRNA-Leu"
     gene            3307..4263
                     /gene="ND1"
     CDS             3307..4263
                     /gene="ND1"
                     /codon_start=1
```

Click the Pubmed-link and pick up the scientific article in which the sequence EF064321 was published:

Science 15 December 2006: Vol. 314. no. 5806, pp. 1767 - 1770
DOI: 10.1126/science.1135566

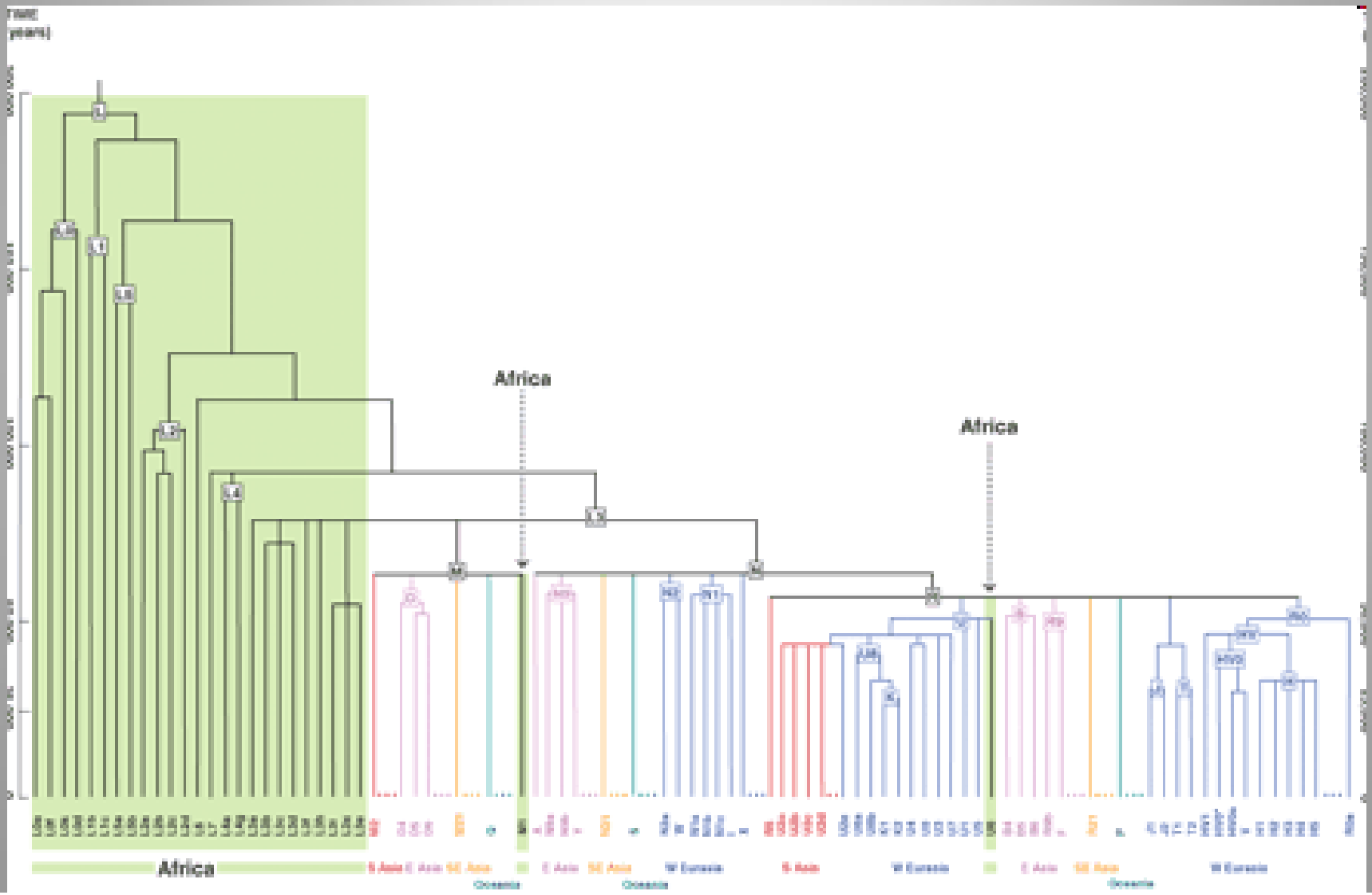
The mtDNA Legacy of the Levantine Early Upper Palaeolithic in Africa

Anna Olivieri, et al. (15 authors)

Abstract:
Sequencing of 81 entire human mitochondrial DNAs (mtDNAs) belonging to haplogroups M1 and U6 reveals that these predominantly North African clades arose in southwestern Asia and moved together to Africa about 40,000 to 45,000 years ago. Their arrival temporally overlaps with the event(s) that led to the peopling of Europe by modern humans and was most likely the result of the same change in climate conditions that allowed humans to enter the Levant, opening the way to the colonization of both Europe and North Africa. Thus, the early Upper Palaeolithic population(s) carrying M1 and U6 did not return to Africa along the southern coastal route of the "out of Africa" exit, but from the Mediterranean area; and the North African Dabban and European Aurignacian industries derived from a common Levantine source.

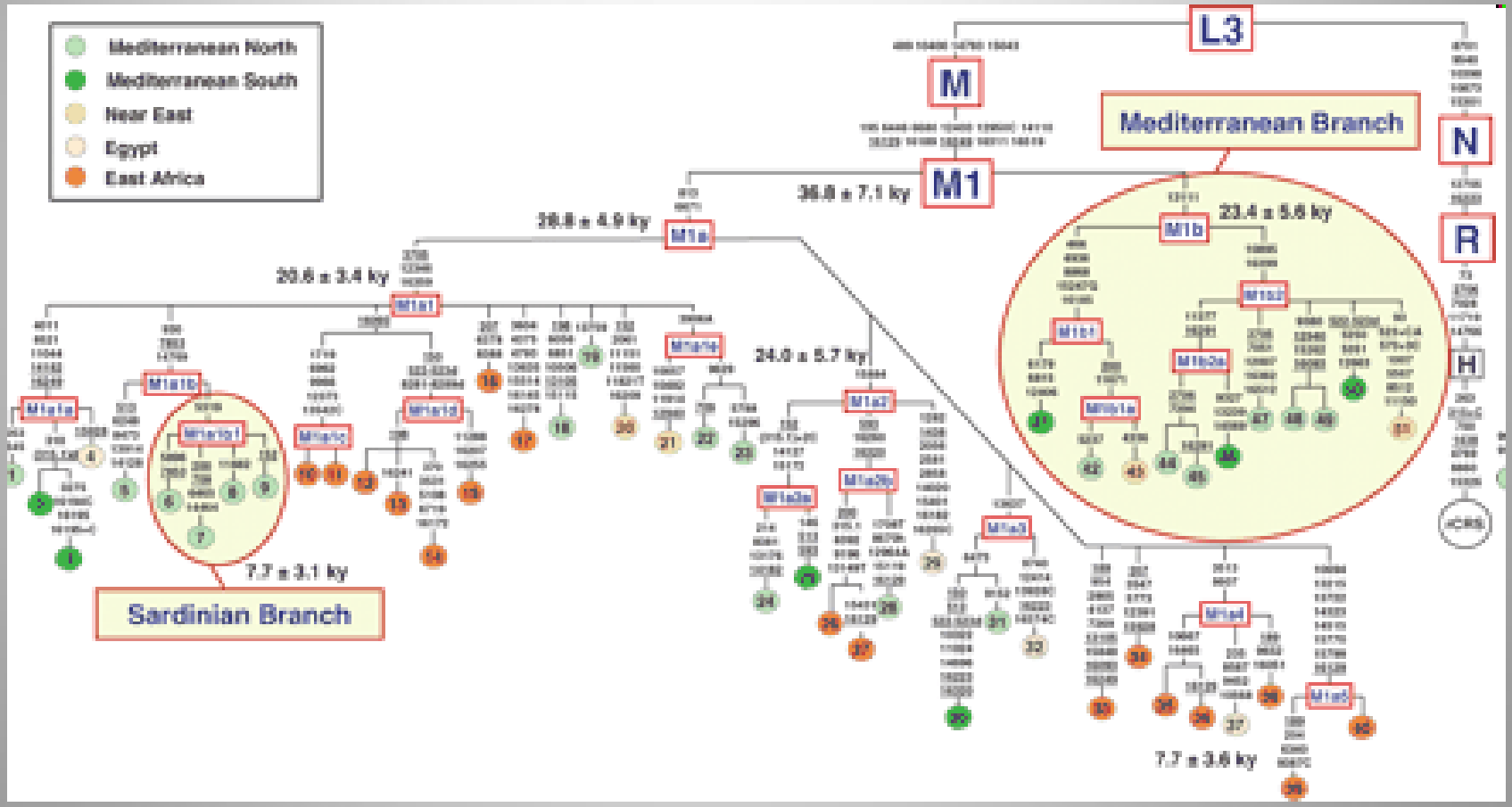
Have a look at the paper

Schematic representation of the worldwide phylogeny of human mtDNA. African haplogroups are in green and those of other geographical regions are in other colors.



Tree of 51 mtDNA sequences belonging to haplogroup M1. The tree is rooted using the reference sequence (rCRS) (27) as an outgroup. The sequencing procedure and phylogeny construction were performed as described elsewhere (4, 28, 29). mtDNAs were selected through a preliminary sequence analysis of the control region and a restriction fragment length polymorphism survey in order to include the widest possible range of internal variation of the haplogroup. All M1 sequences are new except for 17, which is the same sample as 25 in Torroni *et al*. (3). Mutations are shown on the branches; they are transitions unless a base is explicitly indicated. Suffixes indicate transversions (to A, G, C, or T), indels (+, d) or heteroplasmy (h). Recurrent mutations are underlined; pathological mutations are in italics. The ethnic or geographic origins of mtDNAs are as follows: Italy (1, 5 to 9, 23, 24, 28, 31, 42, 44, 45, and 47 to 49); Berbers of Egypt (2 and 3); Egypt (4, 29, 32, and 37); Ethiopian Jews (10 and 11); Ethiopia (12 to 17, 26, 27, 33 to 35, 38, and 40); Greece (18 and 19); Iraqi Jew (20); Druze (21); American (USA) of European ancestry (22);

Berbers of Morocco (25, 30, 46, and 50); Kenya (36); Somalia (39); Mauritania (41); Bedouin, southern Israel (43); and Iraqi (51).



Pubmed also provides links to those papers which have cited a given paper, and

Open the paper as a HTML-document, at the end you´ll find a link:

“THIS ARTICLE HAS BEEN CITED BY OTHER ARTICLES”:

Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers.

L. Quintana-Murci, H. Quach, C. Harmant, F. Luca, B. Massonnet, E. Patin, L. Sica, P. Mouguiama-Daouda, D. Comas, S. Tzur, *et al.* (2008)
PNAS **105**, 1596-1601
| Abstract » | Full Text » | PDF »

mtDNA Variation Predicts Population Size in Humans and Reveals a Major Southern Asian Chapter in Human Prehistory.

Q. D. Atkinson, R. D. Gray, and A. J. Drummond (2008)
Mol. Biol. Evol. **25**, 468-474
| Abstract » | Full Text » | PDF »

…… *and many more…..*

Open the most recent paper, PNAS 105: 1596-1601 (2008) and have a look

Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa.
Batini C, Coia V, Battaggia C, Rocha J, Pilkington MM, Spedini G, Comas D, Destro-Bisol G, Calafell F.
Mol Phylogenet Evol. 2007 May;43(2):635-44. Epub 2006 Oct 5.
PMID: 17107816 [PubMed - indexed for MEDLINE]
Related Articles

The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion.
Destro-Bisol G, Coia V, Boschi I, Verginelli F, Cagliá A, Pascali V, Spedini G, Calafell F.
Am Nat. 2004 Feb;163(2):212-26. Epub 2004 Jan 16.
PMID: 14970923 [PubMed - indexed for MEDLINE]
Related Articles

mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations.
Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC.
Am J Hum Genet. 2000 Apr;66(4):1362-83. Epub 2000 Mar 28.
PMID: 10739760 [PubMed - indexed for MEDLINE]
Related Articles Free article in PMC

Brief communication: mitochondrial DNA variation suggests extensive gene flow from Polynesian ancestors to indigenous Melanesians in the northwestern Bismarck Archipelago.
Ohashi J, Naka I, Tokunaga K, Inaoka T, Ataka Y, Nakazawa M, Matsumura Y, Ohtsuka R.
Am J Phys Anthropol. 2006 Aug;130(4):551-6.
PMID: 16425188 [PubMed - indexed for MEDLINE]
Related Articles

…….. *and many more ……*

Let´s go back to the original sequence EF064321 and start another kind of surfing in databases…

This is a human mtDNA sequence, how similar / different are mtDNA:s between human and, say,

dog, cat, cattle, chicken….?

So, a similar procedure as the one we made with lactase.....but here you make some restrictions, for example ,

choose only ´dog´

Sequences producing significant alignments: (Click headers to sort columns)

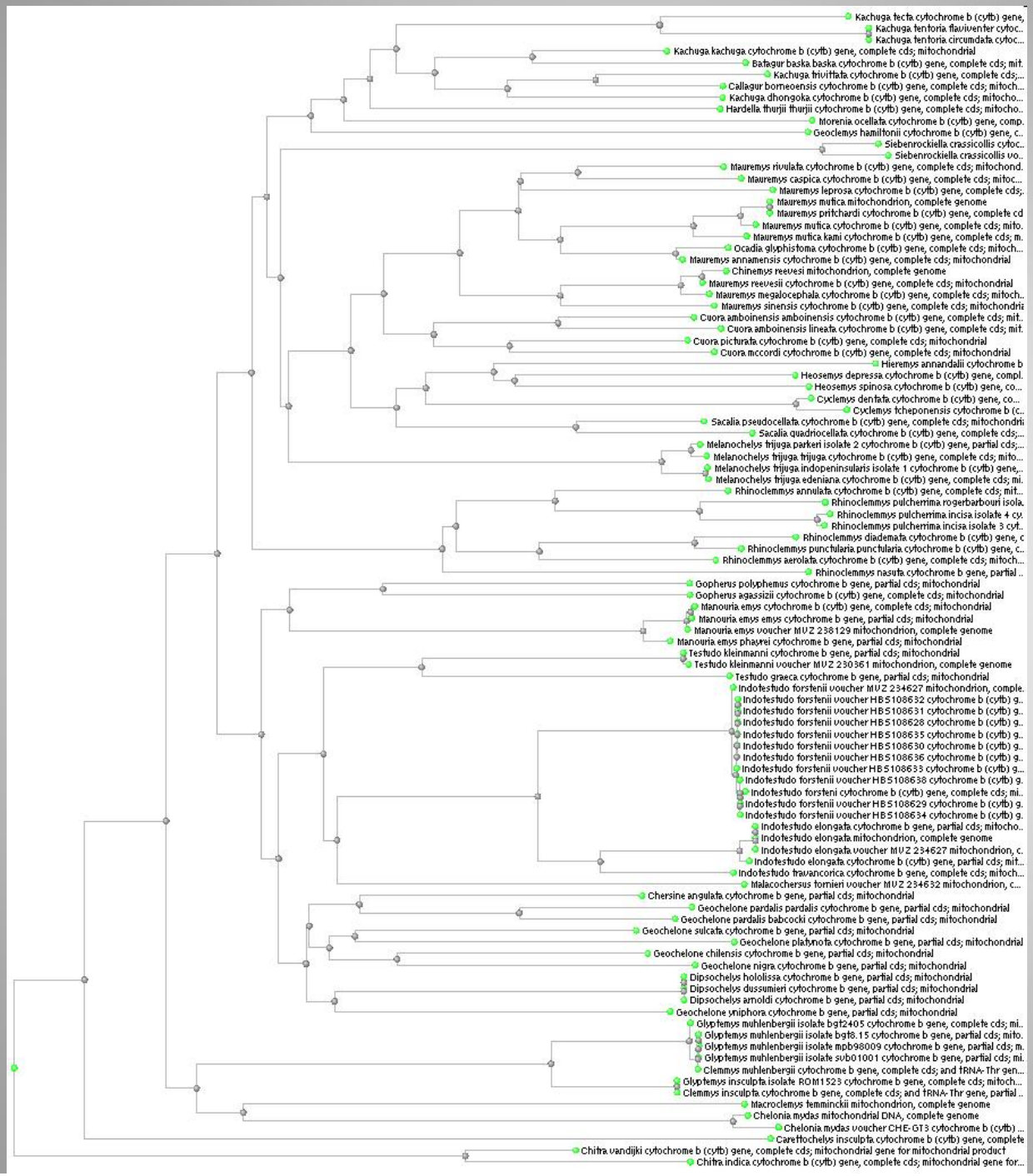
AccessionDescriptionMax scoreTotal scoreQuery coverageE valueMax identLinksDQ480500.1Canis familiaris isolate 1 breed Shetland Sheepdog mitochondrion, complete genome8960903093%0.080%DQ480494.1Canis familiaris isolate 1 breed Poodle mitochondrion, complete genome8960903093%0.080%AY656739.1Canis familiaris isolate 1 breed Poodle mitochondrion, complete genome8960903093%0.080%DQ480502.1Canis familiaris isolate 2 breed Jamthund mitochondrion, complete genome8955902493%0.080%DQ480492.1Canis familiaris isolate 1 breed Jamthund mitochondrion, complete genome8955902493%0.080%AY656740.1Canis familiaris isolate 1 breed Kerry Blue Terrier mitochondrion, complete genome8955902493%0.080%DQ480490.1Canis familiaris isolate 1 breed Flat Coated Retriever mitochondrion, complete genome8951902193%0.080%AY656752.1Canis familiaris isolate 2 breed Standard Schnauzer mitochondrion, complete genome8951902193%0.080%AY656745.1Canis familiaris isolate 2 breed English Springer Spaniel mitochondrion, complete genome8951902193%0.080%AY656743.1Canis familiaris isolate 1 breed Saint Bernard mitochondrion, complete genome8951902193%0.080%DQ480489.1Canis familiaris isolate 1 breed German Shepherd mitochondrion, complete genome8944901493%0.080%DQ480493.1Canis familiaris isolate 1 breed Black Russian Terrier mitochondrion, complete genome8940901093%0.080%DQ480501.1Canis familiaris isolate 1 breed Swedish Elkhound mitochondrion, complete genome8922899293%0.080%AY656751.1Canis familiaris isolate 1 breed Gordon Setter mitochondrion, complete genome8918899493%0.081%DQ480498.1Canis familiaris isolate 1 breed Miniature Schnauzer mitochondrion, complete genome8915899093%0.081%AY656748.1Canis familiaris isolate 1 breed Airedale Terrier mitochondrion, complete genome8915899093%0.081%AY656738.1Canis familiaris isolate 1 breed Jack Russell Terrier mitochondrion, complete genome8913898893%0.081%AY656753.1Canis familiaris isolate 2 breed Irish Setter mitochondrion, complete genome8909898593%0.081%AY656737.1Canis familiaris isolate 1 breed Basenji mitochondrion, complete genome8909898593%0.081%AY656754.1Canis familiaris isolate 1 breed Chinese Crested mitochondrion, complete genome8906898193%0.081%AY656749.1Canis familiaris isolate 2 breed Saint Bernard mitochondrion, complete genome8906898193%0.081%AY656744.1Canis familiaris isolate 1 breed English Springer Spaniel mitochondrion, complete genome8900897693%0.081%DQ480496.1Canis familiaris isolate 1 breed Irish Soft Coated Wheaten Terrier mitochondrion, complete genome88998974

>gb|DQ480500.1| _ Canis familiaris isolate 1 breed Shetland Sheepdog mitochondrion, complete genome Length=16730 Sort alignments for this subject sequence by: E value Score Percent identity Query start position Subject start position Score = 8960 bits (9936), Expect = 0.0 Identities = 11468/15613 (73%), Gaps = 320/15613 (2%) Strand=Plus/Plus

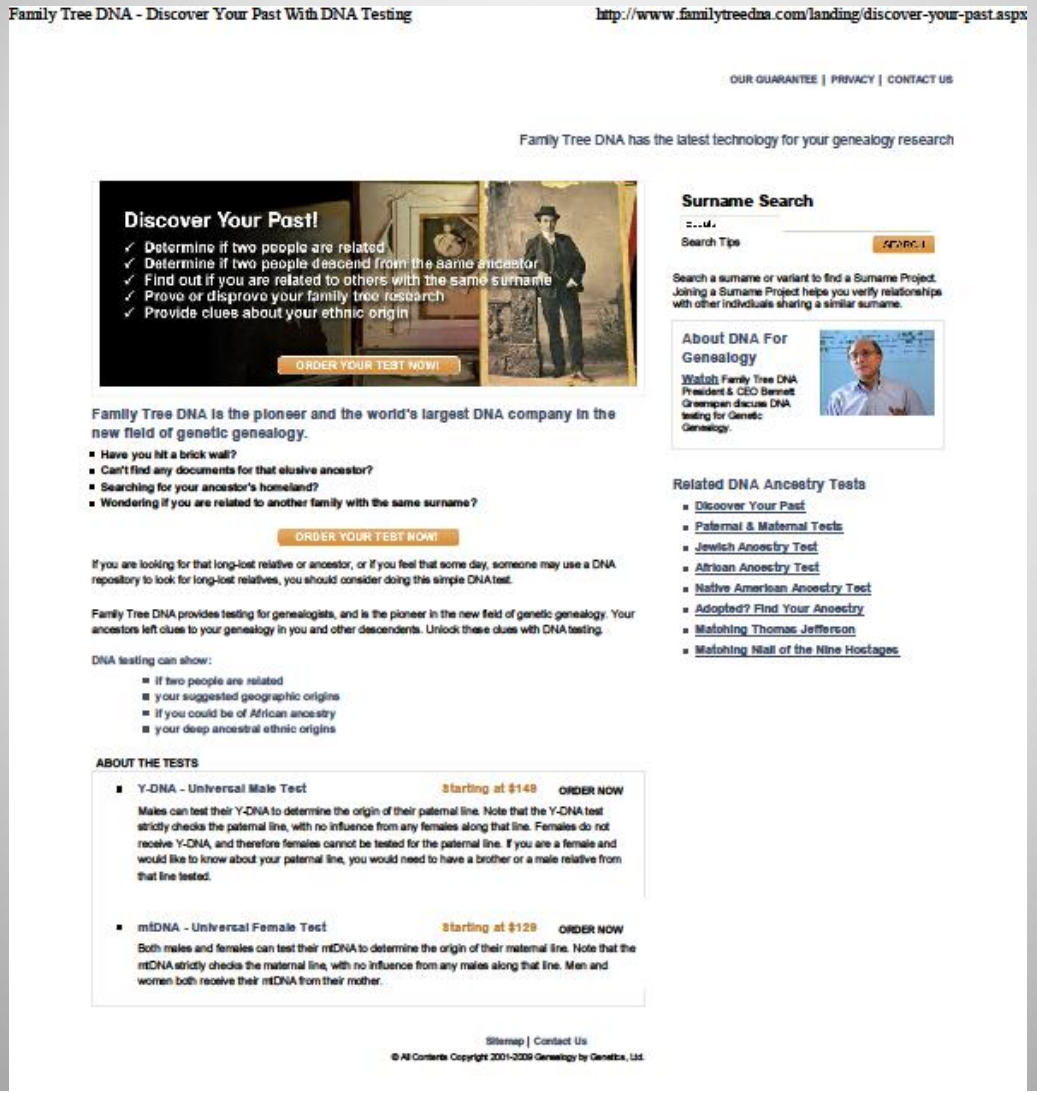
```
Query 578 GTTTATGTAGCTTACCTCCTCAAAGCAATACACTGAAAATGTTTAGACGGGCTCACATCA 637
          ||| |||||||||| || | ||||||| |||||||||| ||| | | || || |
Sbjct 1   GTTAATGTAGCTTAACTAAT-AAAGCAAGGCACTGAAAATGCCAAGATGAG-TCGCACGA 58
```

and so on .....

- human mt-sequence is 16 569bp
- 16 569bp -15 613bp = 956bp of the sequence is not alignable with dog mt-sequence
- in the alignable sequence (15 613bp) the sequences have identical nucleotides in 11 468 sites (73% identity)


Kachuga tecta cytochrome b (cytb) gene,
Kachuga tentoria flaviventer cytoc..
Kachuga tentoria circumdata cytoc...
Kachuga kachuga cytochrome b (cytb) gene, complete cds; mitochondrial
Batagur baska baska cytochrome b (cytb) gene, complete cds; mit.
Kachuga trivittata cytochrome b (cytb) gene, complete cds;...
Callagur borneoensis cytochrome b (cytb) gene, complete cds; mitoch...
Kachuga dhongoka cytochrome b (cytb) gene, complete cds; mitocho...
Hardella thurjii thurjii cytochrome b (cytb) gene, complete cds; mitocho..
Morenia ocellata cytochrome b (cytb) gene, comp.
Geoclemys hamiltonii cytochrome b (cytb) gene, c..
Siebenrockiella crassicollis cytoc...
Siebenrockiella crassicollis vo..
Mauremys rivulata cytochrome b (cytb) gene, complete cds; mitochond.
Mauremys caspica cytochrome b (cytb) gene, complete cds; mitoc...
Mauremys leprosa cytochrome b (cytb) gene, complete cds;.
Mauremys mutica mitochondrion, complete genome
Mauremys pritchardi cytochrome b (cytb) gene, complete cd
Mauremys mutica cytochrome b (cytb) gene, complete cds; mito.
Mauremys mutica kami cytochrome b (cytb) gene, complete cds; m.
Ocadia glyphistoma cytochrome b (cytb) gene, complete cds; mitoch...
Mauremys annamensis cytochrome b (cytb) gene, complete cds; mitochondrial
Chinemys reevesi mitochondrion, complete genome
Mauremys reevesii cytochrome b (cytb) gene, complete cds; mitochondrial
Mauremys megalocephala cytochrome b (cytb) gene, complete cds; mitoch..
Mauremys sinensis cytochrome b (cytb) gene, complete cds; mitochondria
Cuora amboinensis amboinensis cytochrome b (cytb) gene, complete cds; mit..
Cuora amboinensis lineata cytochrome b (cytb) gene, complete cds; mit.
Cuora picturata cytochrome b (cytb) gene, complete cds; mitochondrial
Cuora mccordi cytochrome b (cytb) gene, complete cds; mitochondrial
Hieremys annandalii cytochrome b
Heosemys depressa cytochrome b (cytb) gene, compl.
Heosemys spinosa cytochrome b (cytb) gene, co...
Cyclemys dentata cytochrome b (cytb) gene, co...
Cyclemys tcheponensis cytochrome b (c..
Sacalia pseudocellata cytochrome b (cytb) gene, complete cds; mitochondri
Sacalia quadriocellata cytochrome b (cytb) gene, complete cds;...
Melanochelys trijuga parkeri isolate 2 cytochrome b (cytb) gene, partial cds;...
Melanochelys trijuga trijuga cytochrome b (cytb) gene, complete cds; mito...
Melanochelys trijuga indopeninsularis isolate 1 cytochrome b (cytb) gene,..
Melanochelys trijuga edeniana cytochrome b (cytb) gene, complete cds; mi.
Rhinoclemmys annulata cytochrome b (cytb) gene, complete cds; mit...
Rhinoclemmys pulcherrima rogerbarbouri isola.
Rhinoclemmys pulcherrima incisa isolate 4 cy.
Rhinoclemmys pulcherrima incisa isolate 3 cyt.
Rhinoclemmys diademata cytochrome b (cytb) gene, c
Rhinoclemmys punctularia punctularia cytochrome b (cytb) gene, c..
Rhinoclemmys aerolata cytochrome b (cytb) gene, complete cds; mitoch...
Rhinoclemmys nasuta cytochrome b gene, partial ..
Gopherus polyphemus cytochrome b gene, partial cds; mitochondrial
Gopherus agassizii cytochrome b (cytb) gene, complete cds; mitochondrial
Manouria emys cytochrome b (cytb) gene, complete cds; mitochondrial
Manouria emys emys cytochrome b gene, partial cds; mitochondrial
Manouria emys voucher MVZ 238129 mitochondrion, complete genome
Manouria emys phayrei cytochrome b gene, partial cds; mitochondrial
Testudo kleinmanni cytochrome b gene, partial cds; mitochondrial
Testudo kleinmanni voucher MVZ 230361 mitochondrion, complete genome
Testudo graeca cytochrome b gene, partial cds; mitochondrial
Indotestudo forstenii voucher MVZ 234627 mitochondrion, comple.
Indotestudo forstenii voucher HBS108632 cytochrome b (cytb) g..
Indotestudo forstenii voucher HBS108631 cytochrome b (cytb) g..
Indotestudo forstenii voucher HBS108628 cytochrome b (cytb) g..
Indotestudo forstenii voucher HBS108635 cytochrome b (cytb) g..
Indotestudo forstenii voucher HBS108630 cytochrome b (cytb) g..
Indotestudo forstenii voucher HBS108636 cytochrome b (cytb) g..
Indotestudo forstenii voucher HBS108633 cytochrome b (cytb) g...
Indotestudo forstenii voucher HBS108638 cytochrome b (cytb) g.
Indotestudo forsteni cytochrome b (cytb) gene, complete cds; mi..
Indotestudo forstenii voucher HBS108629 cytochrome b (cytb) g.
Indotestudo forstenii voucher HBS108634 cytochrome b (cytb) g.
Indotestudo elongata cytochrome b gene, partial cds; mitocho..
Indotestudo elongata mitochondrion, complete genome
Indotestudo elongata voucher MVZ 234627 mitochondrion, c.
Indotestudo elongata cytochrome b (cytb) gene, partial cds; mit...
Indotestudo travancorica cytochrome b gene, complete cds; mitoch...
Malacochersus tornieri voucher MVZ 234632 mitochondrion, c...
Chersine angulata cytochrome b gene, partial cds; mitochondrial
Geochelone pardalis pardalis cytochrome b gene, partial cds; mitochondrial
Geochelone pardalis babcocki cytochrome b gene, partial cds; mitochondrial
Geochelone sulcata cytochrome b gene, partial cds; mitochondrial
Geochelone platynota cytochrome b gene, partial cds; mitochondrial
Geochelone chilensis cytochrome b gene, partial cds; mitochondrial
Geochelone nigra cytochrome b gene, partial cds; mitochondrial
Dipsochelys hololissa cytochrome b gene, partial cds; mitochondrial
Dipsochelys dussumieri cytochrome b gene, partial cds; mitochondrial
Dipsochelys arnoldi cytochrome b gene, partial cds; mitochondrial
Geochelone yniphora cytochrome b gene, partial cds; mitochondrial
Glyptemys muhlenbergii isolate bgt2405 cytochrome b gene, complete cds; mi..
Glyptemys muhlenbergii isolate bgt8.15 cytochrome b gene, partial cds; mito.
Glyptemys muhlenbergii isolate mpb98009 cytochrome b gene, partial cds; m.
Glyptemys muhlenbergii isolate svb01001 cytochrome b gene, partial cds; mi.
Clemmys muhlenbergii cytochrome b gene, complete cds; and tRNA-Thr gen...
Glyptemys insculpta isolate ROM1523 cytochrome b gene, complete cds; mitoch...
Clemmys insculpta cytochrome b gene, complete cds; and tRNA-Thr gene, partial ..
Macroclemys temminckii mitochondrion, complete genome
Chelonia mydas mitochondrial DNA, complete genome
Chelonia mydas voucher CHE-GT3 cytochrome b (cytb) ...
Carettochelys insculpta cytochrome b (cytb) gene, complete
Chitra vandijiki cytochrome b (cytb) gene, complete cds; mitochondrial gene for mitochondrial product
Chitra indica cytochrome b (cytb) gene, complete cds; mitochondrial gene for...

mtDNA-database exploitation is also commercial…….



..... and used for barcoding the life

Search

Sitemap | text size +/-

What is CBOL?

The **Consortium for the Barcode of Life** (CBOL) is an international initiative devoted to developing DNA barcoding as a global standard for the identification of biological species. DNA barcoding is a new technique that uses a short DNA sequence from a standardized and agreed-upon position in the genome as a molecular diagnostic for species-level identification. **DNA barcode sequences** are very short relative to the entire genome and they can be obtained reasonably quickly and cheaply. The "Folmer region" at the 5' end of the cytochrome c oxidase subunit 1 mitochondrial region (COI) is emerging as the standard barcode region for almost all groups of higher animals. This region is 648 nucleotide base pairs long in most groups and is flanked by regions of conserved sequences, making it relatively easy to isolate and analyze. A growing number of studies have shown that COI sequence variability is very low (generally less than 1-2%) and that the COI sequences of even closely related species differ by several percent, making it possible to identify species with high confidence. For those groups in which COI is unable to resolve species-level differences, CBOL recommends the use of an additional gene region. In some groups, COI is not an effective barcode region and a different standard region must be identified. In all cases, DNA barcoding is based on the use of a short, standard region that enables cost-effective species identification.

To learn more about DNA barcoding, see "Barcode of Life Initiative", "DNA Barcoding: A New Tool for Identifying Biological Specimens and Managing Species Diversity", "Barcoding Life: Ten Reasons" and the Barcode Blog.

CBOL has more than 160 Member Organizations from more than 50 countries including:

Natural history museums, zoos, herbaria, and botanical gardens;
University departments of biology and molecular biology;
Biodiversity and conservation organizations, NGOs;
Governmental and intergovernmental organizations; and
Private biotech companies.

CBOL's mission is to promote the exploration and development of DNA barcoding as a global standard for species identification. In pursuing this mission, CBOL promotes:

- the rapid compilation of high-quality DNA barcode records in a public library of DNA sequences,
- the development of new instruments and processes that will make barcoding cheaper, faster, and more portable,
- the participation of taxonomists and taxonomic research organizations in all regions and countries, and
- the use of DNA barcoding for the benefit of science and society.

Current Event

Taipei Conference website now available

Get Involved in CBOL!

View CBOL member locations

Find out how you can get involved

Learn about major CBOL projects

Browse Case Studies of barcoding projects

Propose a barcoding project, post a Case Study, and find partners

Submit barcode data

Examine barcode data

Join CBOL

Sponsor Links

Smithsonian

ALFRED P. SLOAN FOUNDATION

Barcode of Life Secretariat
c/o National Museum of Natural History
P.O. Box 37012 MRC 105
Washington, DC 20013-7012
CBOLinfo@si.edu


Federal Bureau of Investigation - Congressional Testimony http://www.fbi.gov/congress/congress01/dwight061201.htm

Home | Site Map | FAQs

FEDERAL BUREAU OF INVESTIGATION

SEARCH

Congressional Testimony

Contact Us
- Your Local FBI Office
- Overseas Offices
- Submit a Crime Tip
- Report Internet Crime
- More Contacts

Learn About Us
- Quick Facts
- What We Investigate
- Natl. Security Branch
- Information Technology
- Fingerprints & Training
- Laboratory Services
- Reports & Publications
- History
- More About Us

Get Our News
- Press Room
- E-mail Updates
- News Feeds

Be Crime Smart
- Wanted by the FBI
- More Protections

Use Our Resources
- For Law Enforcement
- For Communities
- For Researchers
- More Services

Visit Our Kids' Page

Apply for a Job

KRP haluaisi kansasta DNA-rekisterin

Tällä hetkellä poliisin käytössä on rikosperusteinen DNA-rekisteri, jossa on noin 30 000 nimeä.
Keskusrikospoliisin päällikön *Rauno Rannan* mielestä koko kansan kattava DNA-rekisteri olisi poliisille hyvä työkalu henki-, väkivalta- ja seksuaalirikosten tutkinnassa. Rekisteristä olisi hyötyä myös vainajien tunnistamisessa.
- Rekisterin ylläpitäjä voisi olla esimerkiksi Kansanterveyslaitos tai oikeuslääketieteen laitos. Poliisi voisi käyttää rekisteritietoja tarkoin laissa määrätyissä tapauksissa ja tarvittaessa tuomioistuimen luvalla, Ranta sanoo Savon Sanomien haastattelussa.
Rannan mukaan rekisteriä käytettäisiin vain tunnistamiseen eikä se kertoisi poliisille mitään henkilön perimästä tai perinnöllisistä sairauksista. KRP:n päällikkö muistuttaa, että DNA-käytännön muutokset ovat arkaluontoinen asia, joista lopullisen päätöksen tekee eduskunta.
Tällä hetkellä poliisin käytössä on rikosperusteinen DNA-rekisteri, jossa on noin 30 000 nimeä. Rekisteriin voi joutua, jos epäillystä rikoksesta seuraa vähintään kuuden kuukauden vankeusrangaistus.

Testimony of Dwight E. Adams, Deputy Assistant Director, Laboratory Division, FBI
Before the House Committee on Government Reform
Subcommittee on Government Efficiency, Financial Management and Intergovernmental Relations
June 12, 2001
"The FBI's DNA Program"

Mr. Chairman, members of the Subcommittee, I would like to thank the members of the Subcommittee for inviting the FBI to provide an update on our activities relating to forensic DNA analysis specifically with respect to the Combined DNA Index System or CODIS, our National DNA database and our efforts to provide this technology and assistance to state and local forensic laboratories.
The importance of collaboration between federal, state and local forensic laboratories is illustrated by that first group of federal, state and local forensic scientists that were convened by the FBI Laboratory in the 1980's to establish guidelines for the use of forensic DNA analysis in laboratories. This group, the Technical Working Group on DNA Analysis Methods or TWGDAM (now known as the Scientific Working Group on DNA Analysis Methods or SWGDAM), not only developed the guidelines which formed the basis for our national quality assurance standards but they also proposed the creation of a national DNA database for the storage and exchange of DNA profiles developed from crime scenes. This proposal formed the genesis of the development of our CODIS program - software that enables federal, state and local laboratories to store and compare DNA profiles electronically and thereby link serial crimes to each other and identify suspects by matching DNA from crime scenes to convicted offenders. The FBI Laboratory provides this CODIS software, installation, training and user support to other federal, state and local forensic laboratories at no charge. Additionally, the FBI continues to sponsor semi-annual meetings of SWGDAM for over fifty federal, state and local forensic scientists.
How does CODIS work? For example, a sexual assault is committed and an evidence kit is collected from the victim. A DNA profile of the perpetrator is developed from the sexual assault evidence kit. If there is no suspect in the case or if the suspect's DNA profile does not match that of the evidence, the laboratory will search the DNA profile against the convicted offender index. If there is a match in the convicted offender index, the laboratory will obtain the identity of the suspected perpetrator. If there is no match in the convicted offender index, the DNA profile is searched in the forensic or crime scene index. If there is a match in the forensic index, the laboratory has linked two or more crimes together and the law enforcement agencies involved in the cases are able to pool the information obtained on each of the cases.

Part 2, Microbe databases

BLAST Assembled Genomes
Choose a species genome to search, or list all genomic BLAST databases.

Human
Mouse
Rat
Arabidopsis thaliana
Oryza sativa
Bos taurus
Danio rerio
Drosophila melanogaster
Gallus gallus
Pan troglodytes
Microbes ◄
Apis mellifera

BLAST with microbial genomes (1354 bacterial/58 archaeal/239 eukaryotic genomes tree)

This lecture was cancelled because most students had something overlapping ..... microbes only entered the course during the exercise session (exercises with the influenza database)