U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: Cooper P, Morris R. NCBI News, July 2012. 2012 Jul 12 [Updated 2012 Jul 12]. In: NCBI News [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 1991-2012.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/

NCBI News
National Center for Biotechnology Information
National Library of Medicine | National Institutes of Health
Department of Health and Human Services

NCBI News, July 2012

Peter Cooper, Ph.D.⌧[1] and Rana Morris, Ph.D.[2]

Created: July 12, 2012; Updated: July 12, 2012.

Registration now open for NCBI Discovery Workshops September 4-5 at NLM

Registration is now open for the two-day Discovery Workshops to be offered on September 4 -5, on the NIH campus in Bethesda, Maryland. The course is free and is open to anyone interested in NCBI resources. The workshops provide hands-on experience exploring practical examples using tools and databases on the NCBI website. The four workshops are Sequences, Genomes, and Maps; Proteins, Domains and Structures; NCBI BLAST Services; and Human Variation and Disease Genes. For more information see the Discovery Workshops page, which also includes a registration link.

1000 Genomes Dataset Browser

The new 1000 Genomes Browser shows variants, genotypes, and supporting sequence read alignments produced by the 1000 Genomes project. The genotype data are based on the Phase 1, March 2012 set and the variation (NCBI SNP) data are from SNP build 135. The 1000 Genomes browser is accessible from the new NCBI Variation page (Figure 1) that also provides links to other NCBI variation resources including SNP, dbVar, dbGaP, the Variation Reporter, Clinical Remap, and the Phenotype Genotype Integrator. The graphical portion of the genome browser is based on the NCBI graphical sequence viewer (GSV) and offers the familiar features and functions of the GSV. A table of genotypes organized by 1000 Genome populations is shown under the sequence viewer. A summary genotype frequency shows in the table for each population. The population sections can be expanded to show individual level genotypes.

The browser initially opens with an expanded view of human chromosome1 (Figure 1, *Top panel*). The search function on the right-hand side of the browser allows searches for RefSNP accession numbers, gene names, and chromosome positions. The search results show a list of matching items. Clicking on one of these items jumps the display to that position in the genome browser. The browser also allows scrolling either through the Scroll Region arrows on the table of genotypes or through the navigation controls in the graphical portion.

NCBI SNP records that are included in the 1000 Genomes data also link directly to the corresponding position in the browser. Figure 2 shows the region containing a SNP (rs8176058) in the Kell blood group antigen gene (KEL). For any region, the individual next-generation sequencing reads can be selected and loaded into the graphical browser through the expandable Subjects dialog box (Figure 2, *Bottom panel, left*). The Subjects dialog allows selecting individuals by population as well as filtering by the characteristics of the next-generation data. A

Author Affiliations: 1 NCBI; Email: cooper@ncbi.nlm.nih.gov. 2 NCBI; Email: morrisrc@ncbi.nlm.nih.gov.

⌧ Corresponding author.

portion of the aligned exome-sequencing reads for the heterozygous Toscan individual, NA20507, is shown on the left-hand side of the bottom panel of Figure 2.

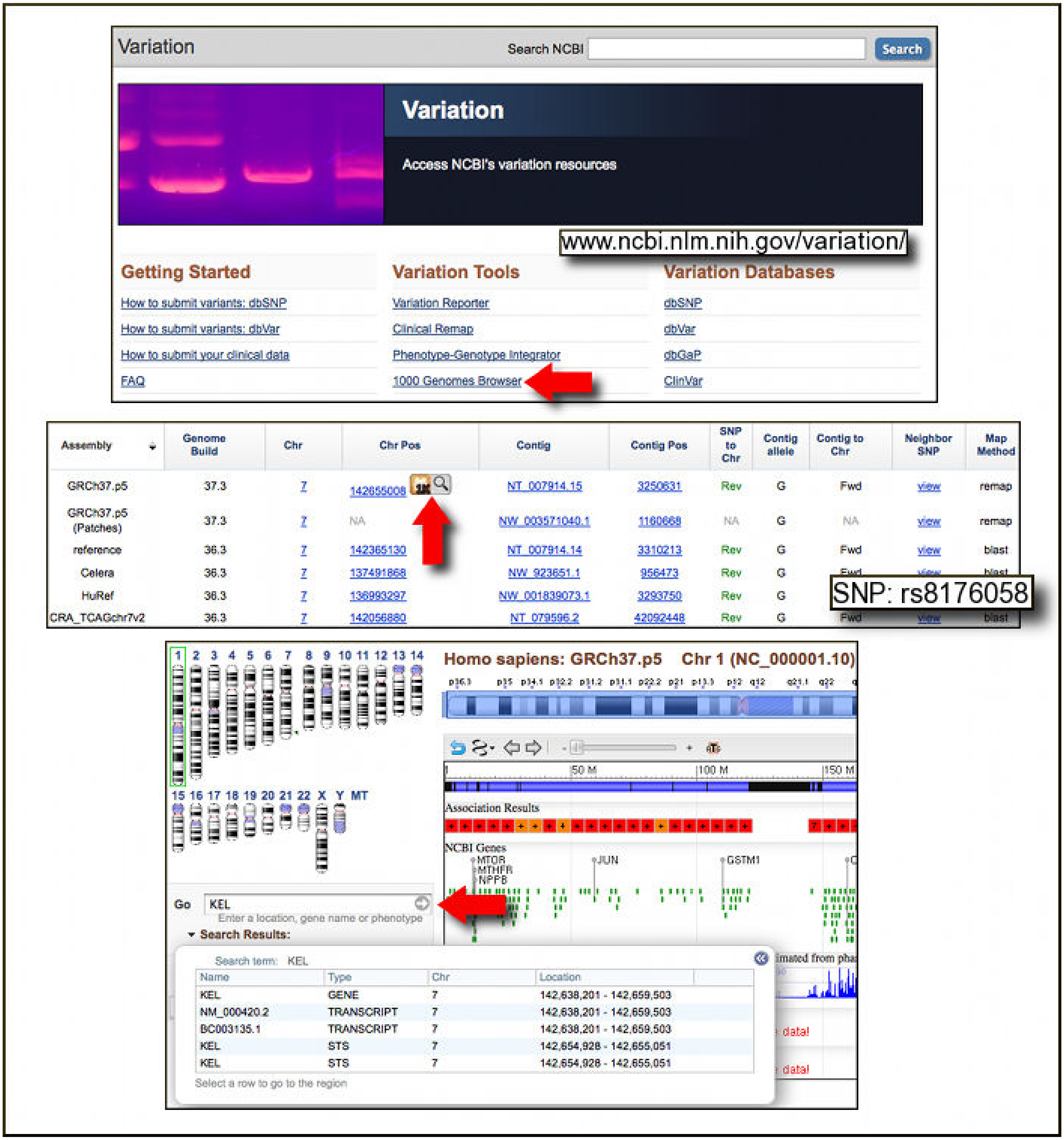
The data from the 1000 genomes project are representative of the increasing importance and presence of “big data” at the NCBI. Currently these data and associated metadata are stored in many different databases at the NCBI including the Sequence Read Archive (SRA), SNP, BioSample, and BioProject. The 1000 Genomes Browser provides a simple and powerful single interface to complex and very large sets of data and metadata that comprise the 1000 Genomes project.

PubMed News

PubMed Send to Citation Manager and Favorites

PubMed now offers the ability to download citations for use in citation manager software such as Endnote, RefWorks or other bibliography program through the "Send to" menu. The PubMed Technical Bulletin has more details on using this feature.

Abstracts in PubMed also now include a "Save items" section that will provide easy way to add items of interest to a My NCBI collection. If you are signed in to My NCBI clicking the "Favorite" button adds the citation to a new My NCBI collection, Favorites. You can add multiple items to My Collections, including Favorites, in My NCBI through the "Send to" menu in the upper right of search result displays. For more information on My NCBI and My Collections please visit My NCBI Help on the NCBI Bookshelf.



Variation
Search NCBI
Search
Variation
Access NCBI's variation resources
www.ncbi.nlm.nih.gov/variation/

Getting Started	Variation Tools	Variation Databases
How to submit variants: dbSNP	Variation Reporter	dbSNP
How to submit variants: dbVar	Clinical Remap	dbVar
How to submit your clinical data	Phenotype-Genotype Integrator	dbGaP
FAQ	1000 Genomes Browser	ClinVar

Assembly	Genome Build	Chr	Chr Pos	Contig	Contig Pos	SNP to Chr	Contig allele	Contig to Chr	Neighbor SNP	Map Method
GRCh37.p5	37.3	7	142655008	NT_007914.15	3250631	Rev	G	Fwd	view	remap
GRCh37.p5 (Patches)	37.3	7	NA	NW_003571040.1	1160668	NA	G	NA	view	remap
reference	36.3	7	142365130	NT_007914.14	3310213	Rev	G	Fwd	view	blast
Celera	36.3	7	137491868	NW_923651.1	956473	Rev	G	Fwd	view	blast
HuRef	36.3	7	136993297	NW_001839073.1	3293750	Rev	G	[illegible]	[illegible]	[illegible]
CRA_TCAGchr7v2	36.3	7	142056880	NT_079596.2	42092448	Rev	G	Fwd	view	blast

SNP: rs8176058
1 2 3 4 5 6 7 8 9 10 11 12 13 14
15 16 17 18 19 20 21 22 X Y MT
Homo sapiens: GRCh37.p5 Chr 1 (NC_000001.10)
p36.1 p35 p34.1 p32.2 p31.2 p31.1 p22.2 p21 p13.3 p12 q12 q21.1 q22
1 50 M 100 M 150 M
Association Results
NCBI Genes
MTOR
MTHFR
NPPB
JUN
GSTM1
imated from pha
data!
data!

Go KEL

Enter a location, gene name or phenotype

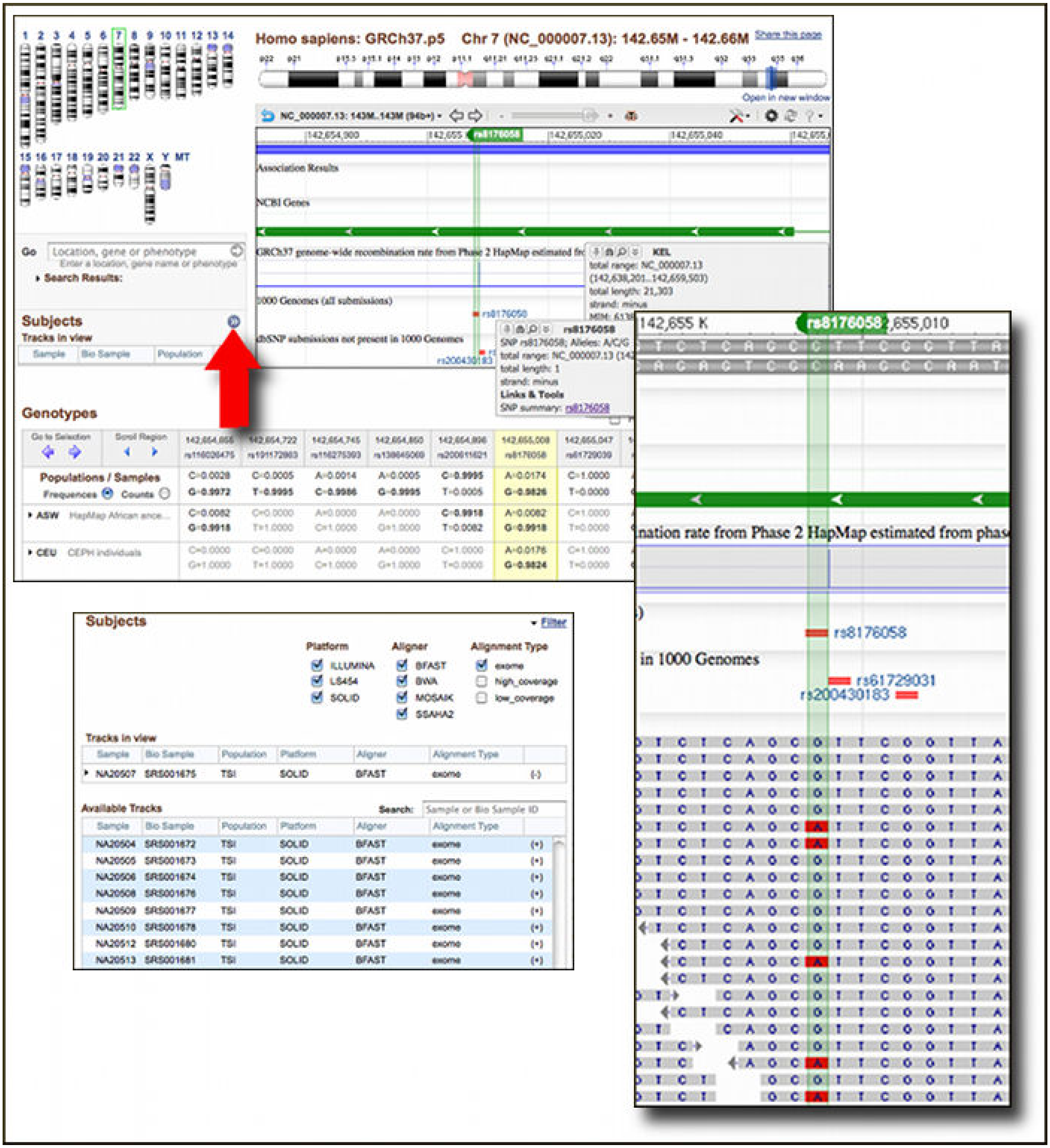
Search Results:

Search term: KEL

Name	Type	Chr	Location
KEL	GENE	7	142,638,201 - 142,659,503
NM_000420.2	TRANSCRIPT	7	142,638,201 - 142,659,503
BC003135.1	TRANSCRIPT	7	142,638,201 - 142,659,503
KEL	STS	7	142,654,928 - 142,655,051
KEL	STS	7	142,654,928 - 142,655,051

Select a row to go to the region

Figure 1. Access to the 1000 Genomes browser. *Top panel*: The Variation Portal that serves as a gateway to a variety of NCBI variation resources including the 1000 Genomes Browser, red arrow. *Center panel*: NCBI SNP records for Reference SNPs included in the 1000 Genomes data link directly to the location in the 1000 Genomes Browser. *Bottom panel:* The 1000 Genomes Browser search interface. Useful search terms include gene names, SNP accessions, and chromosomal positions.


Homo sapiens: GRCh37.p5 Chr 7 (NC_000007.13): 142.65M - 142.66M
Share this page
Open in new window
NC_000007.13: 143M..143M (94bt) ▾
rs8176058
Association Results
NCBI Genes
GRCh37 genome-wide recombination rate from Phase 2 HapMap estimated fr...
1000 Genomes (all submissions)
dbSNP submissions not present in 1000 Genomes
rs8176058
rs200430183
rs61729031
Go
Location, gene or phenotype
Enter a location, gene name or phenotype
Search Results:
142,655 K
655,010
nation rate from Phase 2 HapMap estimated from phas
in 1000 Genomes

KEL
total range: NC_000007.13
(142,638,201..142,659,503)
total length: 21,303
strand: minus

rs8176058
SNP rs8176058; Alleles: A/C/G
total range: NC_000007.13 (142...
total length: 1
strand: minus
Links & Tools
SNP summary: rs8176058

Subjects
Tracks in view

Sample	Bio Sample	Population

Genotypes

Go to Selection / Scroll Region	142,654,658 rs116026475	142,654,722 rs191172863	142,654,745 rs116275393	142,654,860 rs138645069	142,654,896 rs200611621	142,655,008 rs8176058	142,655,047 rs61729039
Populations / Samples Frequences ● Counts ○	C=0.0028 **G=0.9972**	C=0.0005 **T=0.9995**	A=0.0014 **C=0.9986**	A=0.0005 **G=0.9995**	**C=0.9995** T=0.0005	A=0.0174 **G=0.9826**	C=1.0000 T=0.0000
▸ **ASW** HapMap African ance...	C=0.0082 **G=0.9918**	C=0.0000 T=1.0000	A=0.0000 C=1.0000	A=0.0000 G=1.0000	**C=0.9918** T=0.0082	A=0.0082 **G=0.9918**	C=1.0000 T=0.0000
▸ **CEU** CEPH individuals	C=0.0000 G=1.0000	C=0.0000 T=1.0000	A=0.0000 C=1.0000	A=0.0000 G=1.0000	C=1.0000 T=0.0000	A=0.0176 **G=0.9824**	C=1.0000 T=0.0000

Subjects ▾ Filter

Platform	Aligner	Alignment Type
☑ ILLUMINA	☑ BFAST	☑ exome
☑ LS454	☑ BWA	□ high_coverage
☑ SOLID	☑ MOSAIK	□ low_coverage
	☑ SSAHA2	

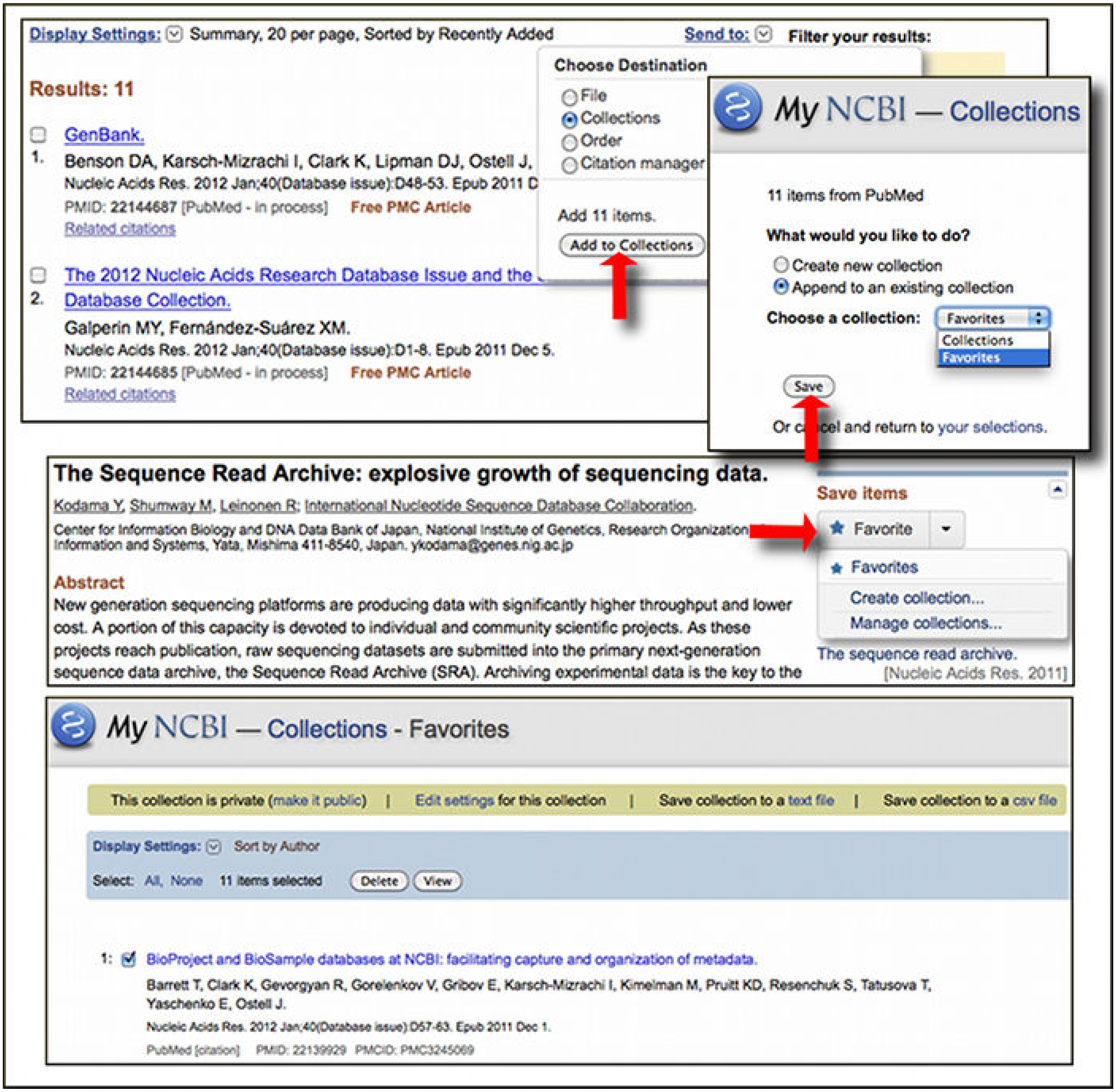
Tracks in view

Sample	Bio Sample	Population	Platform	Aligner	Alignment Type	
▸ NA20507	SRS001675	TSI	SOLID	BFAST	exome	(-)

Available Tracks Search: Sample or Bio Sample ID

Sample	Bio Sample	Population	Platform	Aligner	Alignment Type	
NA20504	SRS001672	TSI	SOLID	BFAST	exome	(+)
NA20505	SRS001673	TSI	SOLID	BFAST	exome	(+)
NA20506	SRS001674	TSI	SOLID	BFAST	exome	(+)
NA20508	SRS001676	TSI	SOLID	BFAST	exome	(+)
NA20509	SRS001677	TSI	SOLID	BFAST	exome	(+)
NA20510	SRS001678	TSI	SOLID	BFAST	exome	(+)
NA20512	SRS001680	TSI	SOLID	BFAST	exome	(+)
NA20513	SRS001681	TSI	SOLID	BFAST	exome	(+)

Figure 2. The 1000 Genomes Browser showing views for the SNP rs8176058, a polymorphism in the Kell blood group antigen protein, the product of the KEL gene. *Top panel*: Initial overview on genotypes for populations showing overall frequencies. The major allele is in bold font. Sections expand to show individual level genotypes. Clicking the red arrow opens ad dialog box (*lower left panel*) that allows selecting and loading next generation sequence read alignments from individuals into the browser. The lower right panel shows some of the exome sequencing reads from a heterozygous individual (NA20507) from the Toscan population (TSI) aligned at the position of rs8176058.



Display Settings: ⌄ Summary, 20 per page, Sorted by Recently Added

Send to: ⌄

Filter your results:

Results: 11

□ 1. GenBank.
Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J,
Nucleic Acids Res. 2012 Jan;40(Database issue):D48-53. Epub 2011 D
PMID: 22144687 [PubMed - in process] Free PMC Article
Related citations

□ 2. The 2012 Nucleic Acids Research Database Issue and the
Database Collection.
Galperin MY, Fernández-Suárez XM.
Nucleic Acids Res. 2012 Jan;40(Database issue):D1-8. Epub 2011 Dec 5.
PMID: 22144685 [PubMed - in process] Free PMC Article
Related citations

Choose Destination
○ File
● Collections
○ Order
○ Citation manager

Add 11 items.

Add to Collections

My NCBI — Collections

11 items from PubMed

What would you like to do?
○ Create new collection
● Append to an existing collection

Choose a collection: Favorites
Collections
Favorites

Save

Or cancel and return to your selections.

The Sequence Read Archive: explosive growth of sequencing data.

Kodama Y, Shumway M, Leinonen R; International Nucleotide Sequence Database Collaboration.

Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization
Information and Systems, Yata, Mishima 411-8540, Japan. ykodama@genes.nig.ac.jp

Abstract

New generation sequencing platforms are producing data with significantly higher throughput and lower cost. A portion of this capacity is devoted to individual and community scientific projects. As these projects reach publication, raw sequencing datasets are submitted into the primary next-generation sequence data archive, the Sequence Read Archive (SRA). Archiving experimental data is the key to the

Save items

★ Favorite

★ Favorites
Create collection...
Manage collections...

The sequence read archive.
[Nucleic Acids Res. 2011]

My NCBI — Collections - Favorites

This collection is private (make it public) | Edit settings for this collection | Save collection to a text file | Save collection to a csv file

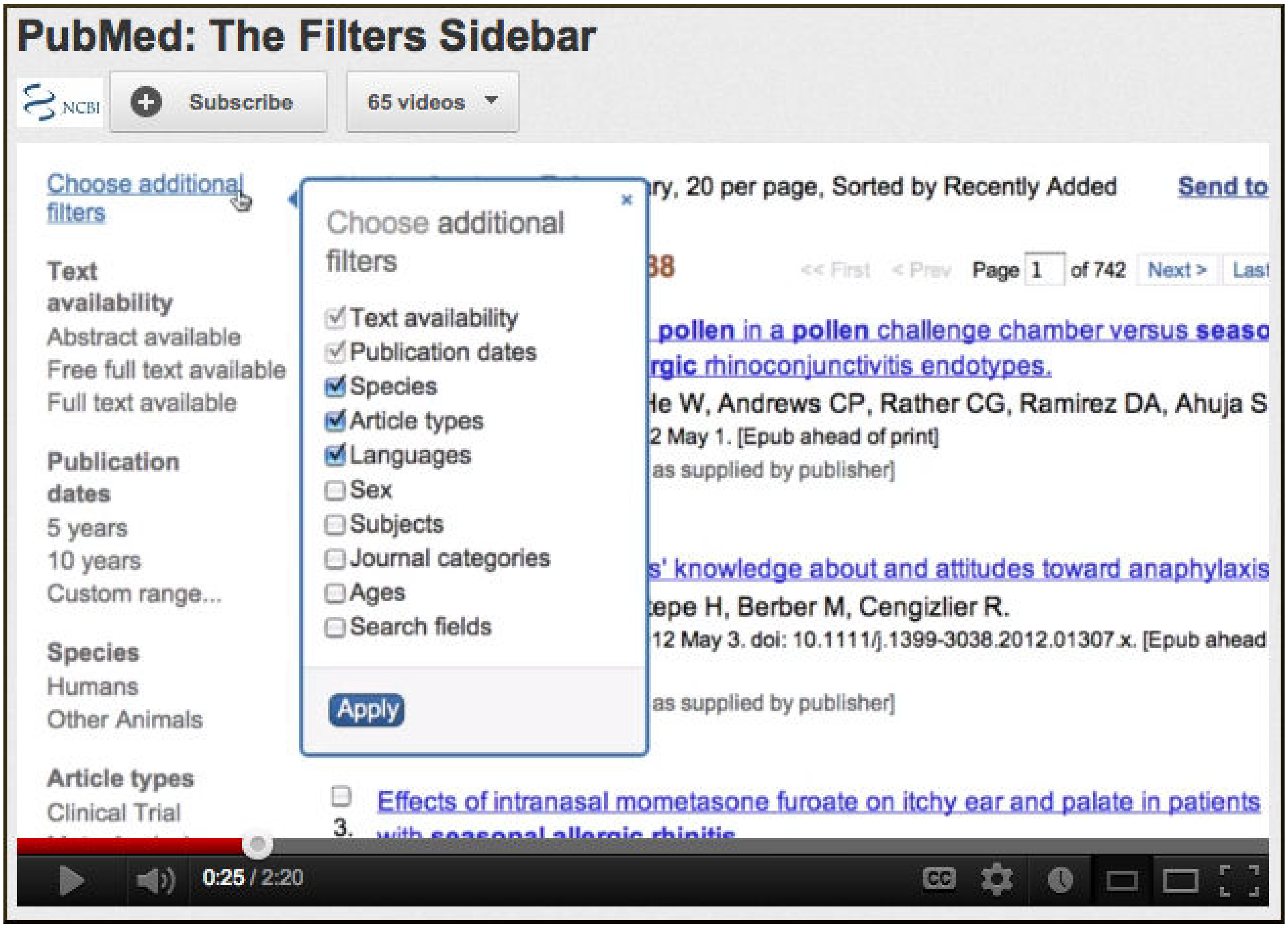
Display Settings: ⌄ Sort by Author

Select: All, None 11 items selected Delete View

1: ☑ BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata.
Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E, Karsch-Mizrachi I, Kimelman M, Pruitt KD, Resenchuk S, Tatusova T, Yaschenko E, Ostell J.
Nucleic Acids Res. 2012 Jan;40(Database issue):D57-63. Epub 2011 Dec 1.
PubMed [citation] PMID: 22139929 PMCID: PMC3245069

PubMed Filter Sidebar

PubMed now has a Filter Sidebar in the PubMed results. The useful features of the popular Limits page have been made more visible by placing them in this Filter Sidebar and should make it easier to refine PubMed search results. For more information, please see the NLM Technical Bulletin. A new video on NCBI's YouTube Channel also demonstrates this useful new addition to PubMed searching.



PubMed: The Filters Sidebar

Choose additional filters

- ☑ Text availability
- ☑ Publication dates
- ☑ Species
- ☑ Article types
- ☑ Languages
- □ Sex
- □ Subjects
- □ Journal categories
- □ Ages
- □ Search fields

Apply

pollen in a pollen challenge chamber versus seaso
rgic rhinoconjunctivitis endotypes.
He W, Andrews CP, Rather CG, Ramirez DA, Ahuja S
2 May 1. [Epub ahead of print]
as supplied by publisher]

s' knowledge about and attitudes toward anaphylaxis
epe H, Berber M, Cengizler R.
12 May 3. doi: 10.1111/j.1399-3038.2012.01307.x. [Epub ahead
as supplied by publisher]

□ 3. Effects of intranasal mometasone furoate on itchy ear and palate in patients with seasonal allergic rhinitis

BLAST News

New Microbial Genomes BLAST Service

A new microbial BLAST service is now live. The service is easier to use and has the familiar format and features of the standard BLAST services at NCBI including the ability to select of taxonomic categories using an auto-complete "Organism" box and to include or exclude multiple taxonomic categories. Other standard features of the BLAST pages such as "Edit and Resubmit" and the ability to optimize for a specific search are also included. For nucleotide databases the search sets have also been divided into Complete and Draft genomes.

Article on Primer-BLAST Published

An article describing Primer-BLAST, NCBI's PCR primer designing service, is now available in BMC Bioinformatics.

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden T. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012 Jun 18;13(1):134. PubMed PMID: 22708584.

BLAST Programming Interface: End of OLD BLAST=true option

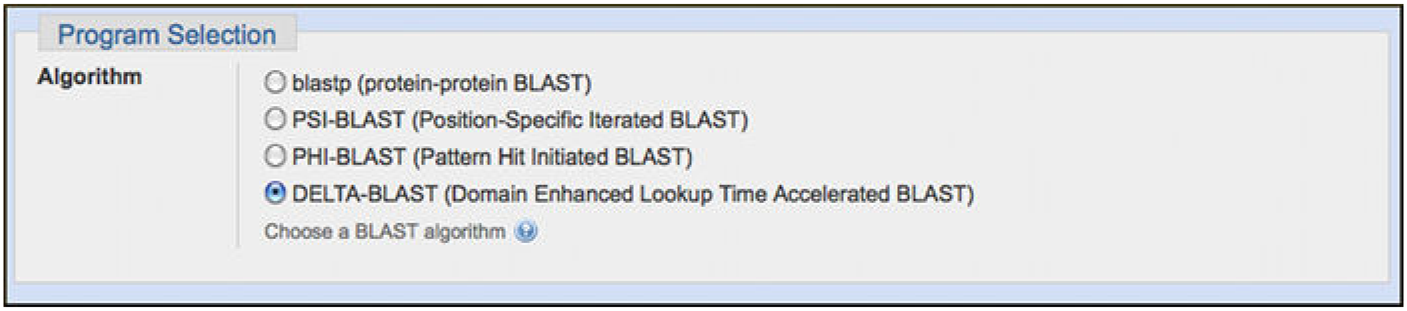
Beginning Sept. 10, 2012, the BLAST service will ignore the OLD_BLAST parameter in posted URLs. We are removing this old and little-used option to prepare for upcoming enhancements to the BLAST service later this year. Setting OLD_BLAST=true produces an older version of the BLAST HTML results that a few people have used for automated processing (parsing) of results. NCBI BLAST supports a number of different and more stable parsable formats. These include XML, tabular reports and ASN.1. For more details, please see BLAST Developer Information and links on that page.

DELTA-BLAST Service and Article

As described in the April 2012 NCBI News Domain Enhanced Lookup Time Accelerated BLAST (DELTA-BLAST) included in the BLAST 2.2.26+ release, offers a more sensitive protein-protein BLAST search by performing a position specific score matrix search using results from an initial conserved domain search. A paper in Biology Direct describes the DELTA-BLAST algorithm and discusses its enhanced sensitivity compared to other methods.

Boratyn GM, Schaffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. *Biol Direct.* 2012 Apr 17;**7**(1):12. PubMed PMID: 22510480.

The protein BLAST web service offers DELTA-BLAST as a Protein BLAST program selection option on the Basic Protein BLAST service.



Program Selection

Algorithm

○ blastp (protein-protein BLAST)
○ PSI-BLAST (Position-Specific Iterated BLAST)
○ PHI-BLAST (Pattern Hit Initiated BLAST)
● DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

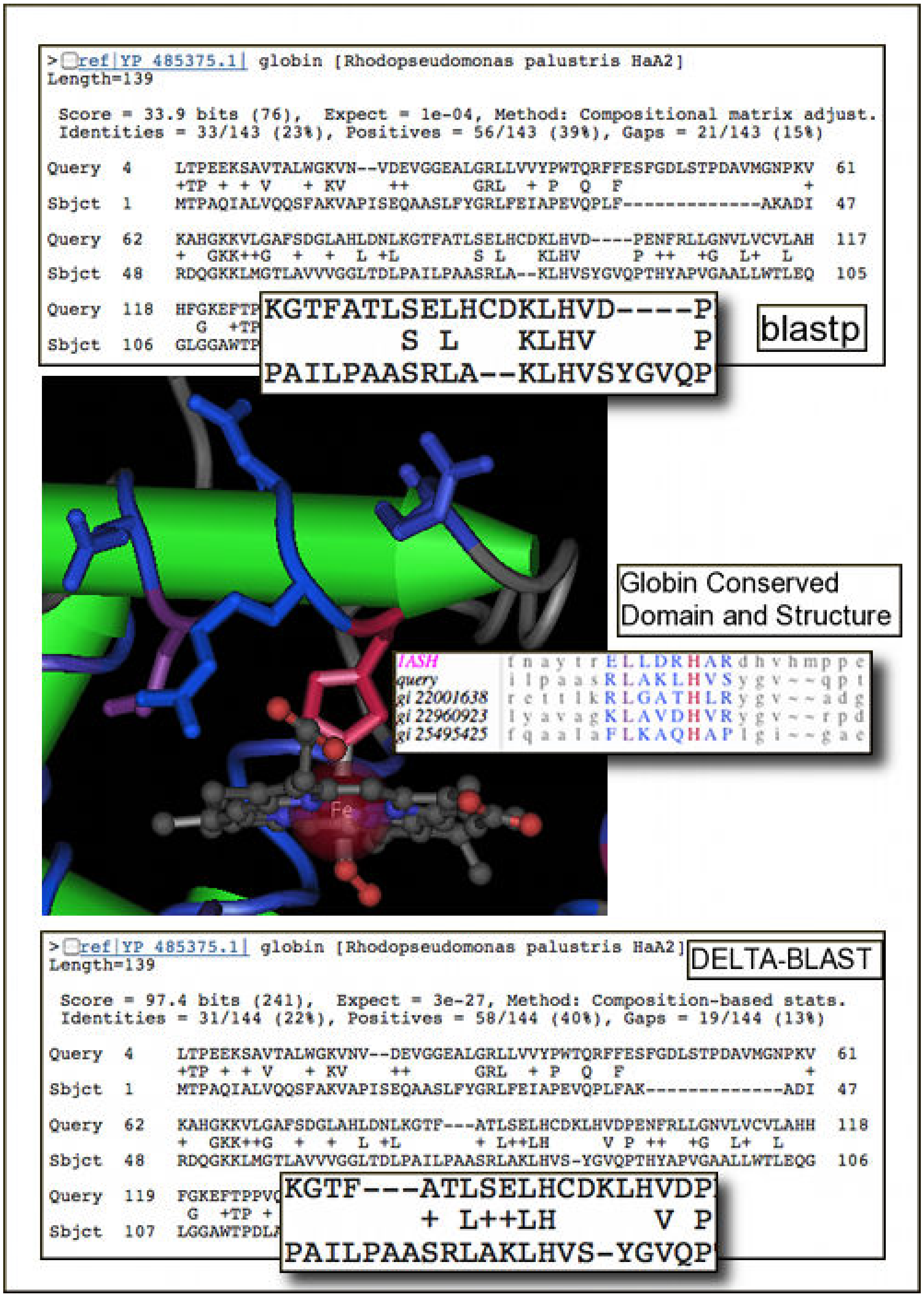
Choose a BLAST algorithm

DELTA-BLAST improves the sensitivity and selectivity of most protein searches that have strong conserved domain results. Figure 3 shows the differing alignments, scores, and expect values for the same match with standard protein blast, blastp (RID: 0E4F72U7013, Figure 3,*Top panel*) and DELTA-BLAST (RID: 0E5RMA6X016, Figure 3, *Bottom panel*). Both of these searches use the human hemoglobin subunit beta protein (NP_000509) as a query against bacterial sequences from the NCBI RefSeq protein database. Standard protein BLAST finds a globin protein (YP_485375) from the purple non-sulfur bacterium *Rhodopseudomonas palustris HaA2* with an expect value of 1 X 10^{-4}. In these results the same expect value is found for some non-globin sequences including an aspartate kinase, an amino peptidase, and a succinate-semialdehyde dehydrogenase. In addition the blastp alignment does not match the conserved histidine (position 93 in NP_000509) that is part of the heme wbinding site in the human hemoglobin domain and structure (Figure 3, *Middle panel*). The blastp alignment also inserts a gap in the conserved alpha helix in this region. In contrast, DELTA-BLAST finds the same protein but with a much better expect value of 3 X 10^{-27}, thus easily segregating the hit from the non-globin proteins found in the blastp search. Moreover, the alignment now corresponds to the globin conserved domain, matching the conserved histidine and preserving the secondary structure block.

DELTA-BLAST is an important new addition that extends the capabilities of the NCBI BLAST service and produces more accurate alignments and more discriminating statistics by using conserved domain information in the initial search.

New HomoloGene Build: Rhesus macaque now included

HomoloGene, the NCBI resource that identifies and clusters homologous genes, transcripts and proteins for selected eukaryotes, has a new build (Build 66). With this build, HomoloGene for the first time includes genes and sequences for the Rhesus monkey (*Macaca mulatta*). The new build also includes updated annotations for



```
>ref|YP_485375.1| globin [Rhodopseudomonas palustris HaA2]
Length=139

 Score = 33.9 bits (76),  Expect = 1e-04, Method: Compositional matrix adjust.
 Identities = 33/143 (23%), Positives = 56/143 (39%), Gaps = 21/143 (15%)

Query  4    LTPEEKSAVTALWGKVN--VDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV  61
            +TP + + V    + KV    ++         GRL  + P  Q  F              +
Sbjct  1    MTPAQIALVQQSFAKVAPISEQAASLFYGRLFEIAPEVQPLF-------------AKADI  47

Query  62   KAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD----PENFRLLGNVLVCVLAH  117
            +   GKK++G  +  +  L +L          S L    KLHV     P ++  +G  L+  L
Sbjct  48   RDQGKKLMGTLAVVVGGLTDLPAILPAASRLA--KLHVSYGVQPTHYAPVGAALLWTLEQ  105

Query  118  HFGKEFTP
              G  +TP
Sbjct  106  GLGGAWTP
```

```
KGTFATLSELHCDKLHVD----P
       S L   KLHV     P
PAILPAASRLA--KLHVSYGVQP
```

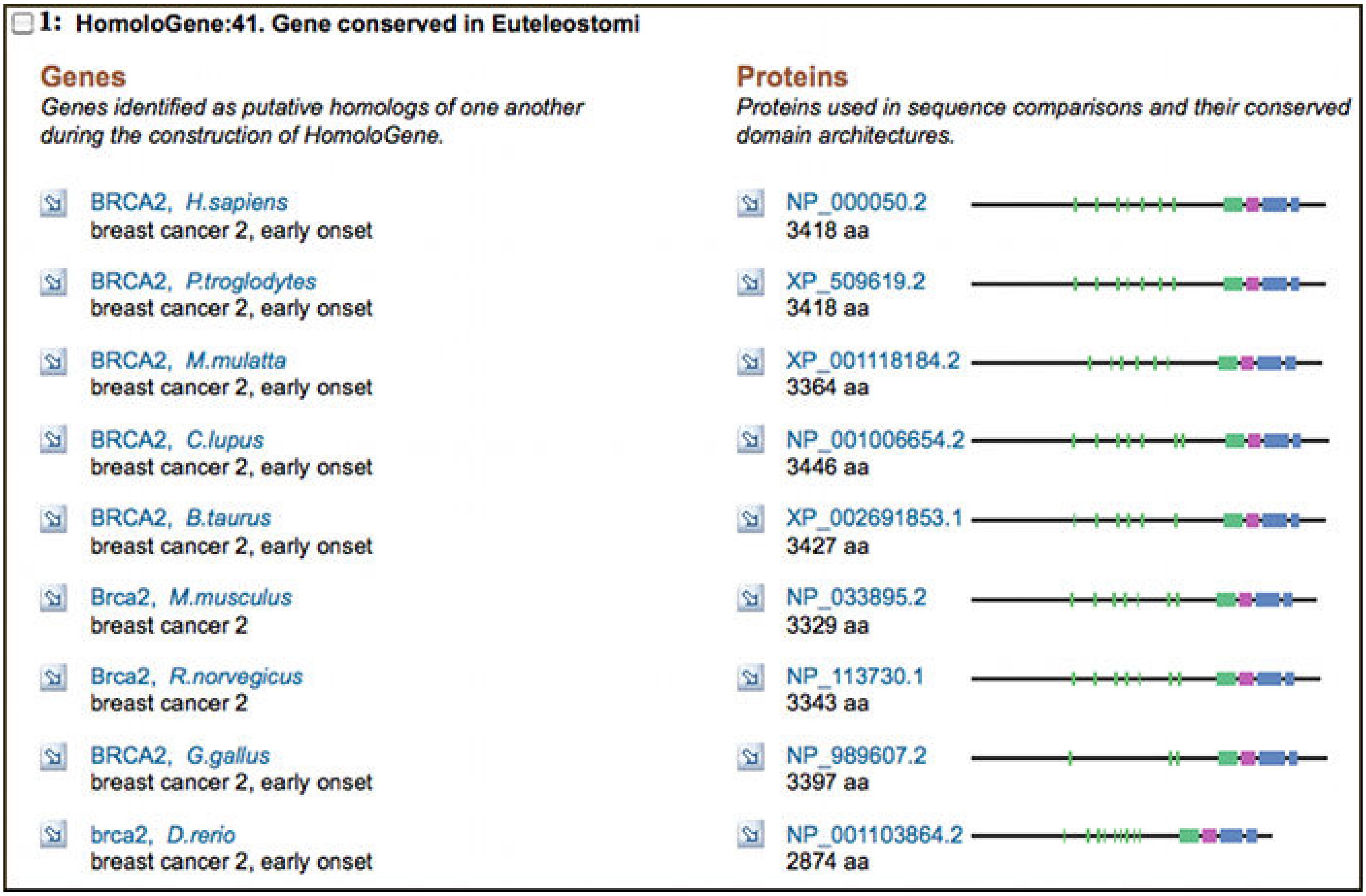
```
IASH        f n a y t r E L L D R H A R d h v h m p p e
query       i l p a a s R L A K L H V S y g v ~ ~ q p t
gi 22001638 r e t t l k R L G A T H L R y g v ~ ~ a d g
gi 22960923 l y a v a g K L A V D H V R y g v ~ ~ r p d
gi 25495425 f q a a l a F L K A Q H A P l g i ~ ~ g a e
```

```
>ref|YP_485375.1| globin [Rhodopseudomonas palustris HaA2]
Length=139

 Score = 97.4 bits (241),  Expect = 3e-27, Method: Composition-based stats.
 Identities = 31/144 (22%), Positives = 58/144 (40%), Gaps = 19/144 (13%)

Query  4    LTPEEKSAVTALWGKVNV--DEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV  61
            +TP + + V    + KV    ++         GRL  + P  Q  F              +
Sbjct  1    MTPAQIALVQQSFAKVAPISEQAASLFYGRLFEIAPEVQPLFAK-------------ADI  47

Query  62   KAHGKKVLGAFSDGLAHLDNLKGTF---ATLSELHCDKLHVDPENFRLLGNVLVCVLAHH  118
            +   GKK++G  +  +  L +L           + L++LH      V P ++  +G  L+  L
Sbjct  48   RDQGKKLMGTLAVVVGGLTDLPAILPAASRLAKLHVS-YGVQPTHYAPVGAALLWTLEQG  106

Query  119  FGKEFTPPVQ
              G  +TP +
Sbjct  107  LGGAWTPDLA
```

```
KGTF---ATLSELHCDKLHVDP
        + L++LH      V P
PAILPAASRLAKLHVS-YGVQP
```

Figure 3. Comparison of standard blastp and DELTA-BLAST statistics and alignments. The match found between the human hemoglobin beta (NP_000509) a globin (YP_485375) from the purple non-sulfur bacterium *Rhodopseudomonas palustris HaA2* is shown. *Top panel*: Protein blast results (RID: 0E4F72U7013). In the blastp alignment the conserved histidine at position 92 in the human protein is not aligned with a corresponding histidine in the bacterial sequence, and gaps are inserted into the conserved alpha helix in this region. *Middle panel*: Partial conserved domain alignment and structure for globin (cd01040) showing the conserved histidine (red H) residue and alpha helix (colored block). *Bottom panel*: DELTA-BLAST alignment (RID: 0E5RMA6X016). The DELTA-BLAST result gives a much more significant expect value and more accurate alignment for the globin domain accurately aligning the conserved histidine and preserving the alpha helix.

human, chimpanzee, dog, cow, mouse, rat, chicken, zebrafish, fruitfly, yeast, arabidopsis, and rice. HomoloGene data are available from the NCBI FTP site.



□ **1: HomoloGene:41. Gene conserved in Euteleostomi**

Genes *Genes identified as putative homologs of one another during the construction of HomoloGene.*	**Proteins** *Proteins used in sequence comparisons and their conserved domain architectures.*
BRCA2, *H.sapiens* breast cancer 2, early onset	NP_000050.2 3418 aa
BRCA2, *P.troglodytes* breast cancer 2, early onset	XP_509619.2 3418 aa
BRCA2, *M.mulatta* breast cancer 2, early onset	XP_001118184.2 3364 aa
BRCA2, *C.lupus* breast cancer 2, early onset	NP_001006654.2 3446 aa
BRCA2, *B.taurus* breast cancer 2, early onset	XP_002691853.1 3427 aa
Brca2, *M.musculus* breast cancer 2	NP_033895.2 3329 aa
Brca2, *R.norvegicus* breast cancer 2	NP_113730.1 3343 aa
BRCA2, *G.gallus* breast cancer 2, early onset	NP_989607.2 3397 aa
brca2, *D.rerio* breast cancer 2, early onset	NP_001103864.2 2874 aa

Microbial Genomes Update

Ninety-two finished microbial (archaeal and bacterial) complete genome sequences were released for 90 microbial strains (7 archaea and 83 bacteria) from April 2012 through June 2012. These include three complete plasmid sequences and 89 chromosome sequences. The original sequence data files submitted to the International Sequence Database Collaboration (INSDC) are available in the Bacteria directory in the genomes area of the GenBank FTP site. RefSeq versions were released for a selected set of 391 of the complete INSDC microbial genome sequences for 387 microbial strains during the same period. These are available from the /genomes/Bacteria directory on the FTP site.

In addition, data from 754 microbial whole genome-shotgun (WGS) sequencing projects were added to the INSDC during this period. The original submitted files are available in the Bacteria_DRAFT directory in the GenBank genomes area. RefSeq provisional versions of 84 WGS microbial projects were released in the /genomes/Bacteria_DRAFT area of the FTP site.

All GenBank and RefSeq microbial genomes are incorporated in the NCBI integrated Entrez search and retrieval system and the BLAST sequence similarity search service.

GenBank News

GenBank release 190 is available through the NCBI web and FTP sites. The current release incorporates data available as of June 15, 2012 and, with the whole-genome shotgun portion, contains 428,920,607,871 bases from 236,206,989 sequence records. Release notes describe the current state of data and upcoming changes. The GenBank page provides more information on the database content and scope as well as submission information.

RefSeq News

RefSeq Release 54 is available through Entrez, BLAST, and from the RefSeq FTP area. The current release includes 21.9 million Reference Sequence records from 17,605 different species or strains. The RefSeq release notes provide more detailed information.

GRC Plans New Human Genome Build and Requests Input

The Genome Reference Consortium (GRC), which produces assemblies that are the basis for NCBI Reference assemblies for human, mouse, and zebrafish, is planning a new build of the human genome (GRCh38) for summer of 2013. Anyone who has questions, concerns, or input, may submit these on the GRC contact form. The GRC blog provides insights into the complexities and the process of updating, correcting, and representing the human genome.

NCBI Now Offers IPv6 Access

The NCBI website now supports the new six-byte Internet Protocol addresses (IPv6) for HTTP access as well as data downloads using FTP, Aspera, and RSync. The World IPv6 Launch site has additional information on the transition to IPv6.

Keeping Up with NCBI

Seventeen topic-specific mailing lists are available that provide email announcements about changes and updates to NCBI resources including dbGaP, BLAST, GenBank, and Sequin. The various lists are described on the Announcement List summary page. Subscribe to the NCBI Announce list to receive updates on the NCBI News.

Twenty-six RSS feeds are now available from NCBI including news on PubMed, PubMed Central, NCBI Bookshelf, LinkOut, HomoloGene, UniGene, and NCBI Announce.

NCBI's Facebook page and Twitter feed also provide updates on NCBI resources.

Send comments and questions about NCBI resources to info@ncbi.nlm.nih.gov, or call 301-496-2475 between the hours of 8:30 a.m. and 5:30 p.m. EST, Monday through Friday.