

| PublisherInfo | | |
|----------------------|---|----------------|
| PublisherName | : | BioMed Central |
| PublisherLocation | : | London |
| PublisherImprintName | : | BioMed Central |

Variations on a gene

| ArticleInfo | | |
|-----------------------|---|--|
| ArticleID | : | 3617 |
| ArticleDOI | : | 10.1186/gb-2000-1-1-reports243 |
| ArticleCitationID | : | reports243 |
| ArticleSequenceNumber | : | 108 |
| ArticleCategory | : | Web report |
| ArticleFirstPage | : | 1 |
| ArticleLastPage | : | 4 |
| ArticleHistory | : | RegistrationDate : 2000-3-9 Received : 2000-3-9 OnlineDate : 2000-4-27 |
| ArticleCopyright | : | BioMed Central Ltd2000 |
| ArticleGrants | : | |
| ArticleContext | : | 130591111 |

Colin Semple

Abstract

Launched in September 1998, dbSNP is the central repository for data on single-nucleotide polymorphisms (SNPs), the most common form of sequence variation in human populations.

Content

Launched in September 1998, dbSNP is the central repository for data on single-nucleotide polymorphisms (SNPs), the most common form of sequence variation in human populations. It is hoped that establishing a large collection of mapped SNPs will accelerate the identification of disease genes by association studies. Such studies aim to discover statistical associations between genetic variations and heritable diseases. At the most recent update, dbSNP contained data for 26,397 SNPs but this number is expected to increase to several hundred thousand over the next few years. The database contains details of the sequences in which the SNP is found, mapping information, the populations and number of chromosomes sampled, estimated heterozygosities, and PCR assay primers and protocols. Each unique SNP is assigned a reference SNP ID, or 'rs ID'. To reduce redundancy in the database, subsequent submissions that map to locations identical to previously submitted SNPs will be linked to the existing reference SNP record. New data may be submitted to dbSNP by e-mail as explained at the site. The database may be searched by simple text queries of various database fields. For example, a SNP may be retrieved by its rs ID, the GenBank accession numbers of the sequences within which it occurs or the laboratory that submitted it. Perhaps the most useful search method provided is BLAST searching of SNP flanking sequences - the identification of SNPs plus their surrounding sequences that significantly match a sequence of interest.

Navigation

The site is well documented, with links from SNP records to other NCBI resources such as [GenBank](#) and [LocusLink](#). It is not possible, however, to bookmark individual SNP records. The design of the site occasionally tends towards the labyrinthine; the pages on submission make the process seem quite complex at first sight. The new interfaces for form-based data submission and searching that are currently under development will undoubtedly address such shortcomings.

Reporter's comments

Timeliness

dbSNP is updated erratically every few months with newly submitted variations. Submissions from large, ongoing projects are expected to expand the database by thousands of new SNPs at a time at irregular intervals over the next few years.

Best feature

One very useful feature of dbSNP is its integration with NCBI's [LocusLink](#), which summarizes the sequence and mapping data available for known loci, including information on alternative gene names, expression (via the NCBI [UniGene](#) database) and gene products. Given the GenBank accession number of an mRNA, one can retrieve the corresponding [LocusLink](#) entry and view all the SNPs associated with all the sequences in that entry, regardless of whether these SNPs are present in the starting mRNA sequence. For example it is possible to find SNPs present in the untranslated regions of mRNAs when all you begin with is the coding sequence accession number. Eventually, as [LocusLink](#) integrates genomic data from the Human Genome Project, it should be possible to find intronic SNPs in the same way.

Worst feature

Caution should be taken with any SNPs retrieved from the database because of the presence of non-validated SNPs identified computationally by groups such as the [Cancer genome anatomy project genetic annotation initiative](#) (CGAP-GAI). Such SNPs are annotated as 'candidates' and have not been verified empirically. A proportion (around 8% of the CGAP-GAI candidate SNPs) of the candidate SNPs are expected to be artifactual.

Wish list

In positional cloning projects the interest is invariably in a region of the genome rather than a specific gene. Consequently, it would be useful to be able to retrieve SNPs by map location. This kind of search is currently under development.

Related websites

The [SNP consortium](#) is a non-profit foundation set up by the Wellcome Trust and various pharmaceutical companies to develop up to 300,000 SNPs distributed evenly throughout the human genome and to make the information related to these SNPs publicly available. The SNP consortium site provides searchable access to quarterly SNP data releases; the data are also submitted to dbSNP. [Human genic bi-allelic sequences](#) (HGBase) is a database of intragenic (promoter to end of transcription) sequence polymorphisms including SNPs and intron variations. It contains a smaller number of variations than dbSNP.

Table of links

[dbSNP: A database of human single nucleotide polymorphisms](#)

[GenBank](#)

[LocusLink](#)

[UniGene](#)

[Cancer genome anatomy project genetic annotation initiative](#)

[SNP consortium](#)

[Human genic bi-allelic sequences](#)

References

1. [dbSNP: A database of human single nucleotide polymorphisms.](#)