WiGID: wireless genome information database

Björn M. Ursing*

Center for Genomics and Bioinformatics, Karolinska Institute, SE-171 77 Stockholm, Sweden

ABSTRACT

Summary: WiGID, wireless genome information database, is a new application for mobile internet and can be reached through wireless application protocol (WAP). The main purpose of WiGID is to give easy access to information on completely sequenced genomes. Genome entries in WiGID can be queried by the number of open reading frames (ORFs), genus and species name and year published. Initial search results are linked to information on the full entry.

Availability: WiGID can be accessed through WAP at http://wigid.cgb.ki.se/index.wml and through the regular internet at http://wigid.cgb.ki.se.

Contact: bjorn.ursing@cgb.ki.se

DATA AND STRUCTURE

WiGID is a secondary database, with primary genome information collected manually from other databases such as Genome OnLine Database (GOLD; Bernal *et al.*, 2001; http://wit.integratedgenomics.com/GOLD/) and NCBI (http://www.ncbi.nlm.nih.gov/), as well as genome project web sites, as the information cannot be automatically downloaded. A genome entry in WiGID comprises information on a WiGID specific entry number, species and strain name, genome size, number of ORFs, accession number in EMBL nucleotide database (Stoesser *et al.*, 2002), literature reference of original publication, year published, G+C content and taxonomic category according to GOLD. In entries of eukaryotic genomes, some are single chromosomes and some are complete genomes, dependent upon genome size and the completeness of the genome information. The data is stored in one file, with one line per genome entry. All scripts are written in Python.

SEARCH

The user can query WiGID by name of organism (genus or species name), year published and number of ORFs predicted in the genome (Figure 1). Using the name as

*To whom correspondence should be addressed.

query, only the first letter(s) of the genus or species name needs to be entered. Year of publication and numbers of ORFs are selected from lists. Each query can give many results, and in order to optimize the use of the screen, the results are presented in an abbreviated form. After entering the query, the user specifies the information requested for the list of abbreviated results. This is chosen from a list comprising name, year of publication, number of ORFs, genome size, accession number of the genome sequence in EMBL nucleotide database (Stoesser *et al.*, 2002), literature reference to the genome publication and taxonomic assignment according to GOLD (Bernal *et al.*, 2001). The results are given as a list with the first letter of the genus name and the full species name followed by the specified information. In the case where the information of the complete name is requested, the whole name and strain name is presented, with no abbreviations. Each entry in the list of hits is a link to a presentation of a full entry comprising information on taxonomy of organism according to GOLD and G+C content in the genome in addition to the above-mentioned list of information.

For a single search, one receives typically 3000 bytes and sends about 800 bytes. For additional searches, the number of bytes received will only increase with the new results received as the query page is cached. The help page is approximately 1100 bytes. Although the amount of data transferred is not high, general packet radio service (GPRS), a broadband connection to WAP, is recommended to enhance access speed. WiGID has been tested using Ericsson T68 and Motorola P7389 with GPRS, and Nokia 3330 with a dialled-up WAP connection.

On the WiGID WAP-site there is also a collection of other scientific sites, e.g. BioWAP (Riikonen *et al.*, 2001), one of the first scientific WAP portals. The start page will be increased and possibly turn into a portal itself. WiGID can also be searched through the HTML pages, and the WAP interface can be viewed through WapRunner, a web-based WAP simulator from wap3 technologies GmbH. However, the simulator does not allow searches.
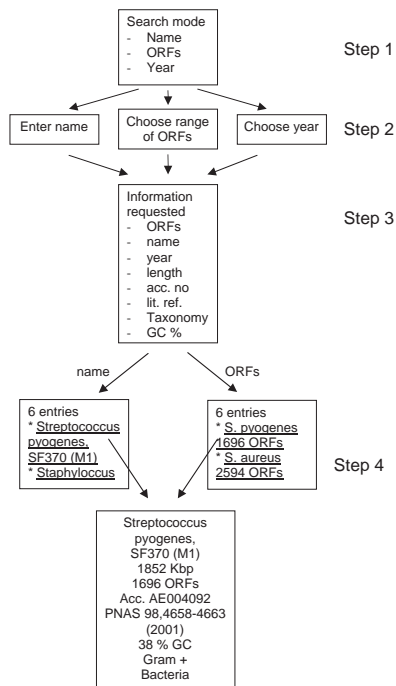
Fig. 1. A flow chart over a search in WiGID. Step 1: choose search mode. Step 2: enter query. Step 3: choose information to be shown in the summary page of abbreviated results. Step 4: choose entry for full view.

WAP

The amount of data that can be transmitted and viewed limits the usefulness of WAP. Nevertheless, the abundant access to mobile internet still makes WAP a useful tool. WiGID will never replace the big genome databases containing annotated sequences with multiple links to other databases. Instead, WiGID will serve as a tool to provide up-to-date answers when the user does not have access to stationary internet.

In the near future WAP will change with respect to bandwidth with G3, making the applications faster and allow more data to be transferred. When palm tops, which have increased screen-size and memory, are connected to the mobile internet, more extensive data can be displayed.

REFERENCES

Bernal,A., Ear,U. and Kyrpides,N. (2001) Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. *Nucleic Acids Res.*, **29**, 126–127.

Riikonen,P., Boberg,J, Salakoski,T and Vihinen,M. (2001) BioWAP, mobile internet service for bioinformatics. *Bioinformatics*, **17**, 855–856.

Stoesser,G., Baker,W., van den Broek,A., Camon,E., Garcia-Pastor,M., Kanz,C., Kulikova,T., Leinonen,R. and Lin,Q. *et al.* (2002) The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, **30**, 21–26.