## **DDBJ** Usefulness of WABI (Web API for Biology) Environments

DNA Data Bank of Japan

Yeondae Kwon<sup>1</sup>, Yasumasa Shigemoto<sup>1,2</sup>, Yoshikazu Kuwana<sup>1,3</sup>, Hideaki Sugawara<sup>1</sup> <sup>1</sup> Center for Information Biology and DDBJ, National Institute of Genetics, <sup>2</sup> Fujitsu Limited, <sup>3</sup> Tokai software

## Summary Contact us: hsugawar@genes.nig.ac.jp DNA Data Bank of Japan (DDBJ) provides Web-based systems for biological analysis, called Web APIs for biology (WABI). **Customization** ervices at DDBJ Programming Each Web service and workflow can be used as a complete service or a building block for programmers to 2 Very flexible Maad construct more complex information processing systems. Impossible Workflows No need Workflow navigation system aims to help non-programming biologists perform analysis tasks by providing Flexible Workflow Navigation System No need next applicable services on Web browsers according to the output of a previously selected service. Disease-related Gene Workflow DDBJ Web APIs for Biology (DDBJ WABI) We are developing an information retrieval system named DDSS (Drug Discovery Web API for Biology (WABI) http://www.xml.nig.ac.jp/ and Diagnostic Support System), which presents associations among genes, diseases, and chemicals. Web APIs are the interfaces for programs to access Web services on the Internet. We propose a specificity-based ranking method that ranks a gene which causes a Web APIs are building blocks for the developments of workflows. given disease but does not cause other diseases at the top. Our ranking policy is Web APIs are implemented by both REST and SOAP. useful for drug discovery and diagnosis. DDBJ provides 129 methods (APIs) from 21 categories (services). indirect associations Provided Web APIs el transition 2<sup>nd</sup> transition n<sup>th</sup> transi d d d d d d d d Service name (# of Web APIs) Service description Keyword search and data retrieval against 20 public databases frequency of co-occurrences of the genes DDBJ(7), ARSA(4), GetEntry(44), Weighting based GIb GTOP g2 -(g Analysis functions such as homology search and multiple gene-disease association gene-gene association Blast(6), ClustalW(4), Fasta(5), VecScreen(4) Associated diseases with g2 SPS AR5A TxSearch RefSeq 60 NCBIGenomeAnnotation alianments OWLW DDBJ original database system (microbial/virus genome, insertion sequence, environmental sequence, re-evaluation of ORF in genome, Gib(11), Gtop(3), GTPS(8), GIBV(8), GIBEnv(1), GIBIS(1), SPS(2) DDBJ Workflow Navigation System on RefSeq GTOP protein structure) http://cyclamen.ddbj.nig.ac.jp/ TxSearch(5), RefSeq(1), GO(3), Ensembl(4), OMIM(2), NCBIGenomeAnnotation(4) Useful databases developed by other institutes A workflow navigation system dynamically presents users a list of services that are executable as a next Web I/F generator Meta-info step according to results of a service How to Use DDBJ Web APIs L Dire Users only need to select a service (follow a Web link) which they would like to execute from the list of An example of retrieving a DNA sequence entry for an accession number from the REST server executable services A list of services is generated from dictionaries and use LWP::UserAgent; \$ug =new LWP::UserAgent; meta-information of services. # make a request \$req = new HTTP::Request POST => 'http://xml.nig.ac.jp/rest/Invoke'; \$req->content\_type('application/x-www-form-unlencoded'); #set parameters and retrieve a DDBJ entry in a FASTA format \$req->content('service=GetEntryåmethod=getDDBJEntryåaccession=AB000100'); How to Use DDBJ Workflow Navigation System # sent the request and get response \$res = \$ua->request(\$req); For example, if you want to retrieve DDBJ entry and then execute BLAST search, (1) Workflow Navigation System (1) Click on [Keyword search] tab of the workflow navigation # show response print \$res->content; system top page. Topper Re-Trauen Betrie ve DOBJ flat file entry by a - Sanch 1001/ in human - Sanc (2) (4) DDBJ Workflows A workflow is a predefined series of applications of Web APIs. (2) Click on [DDBJ]. (3) Click on [FlatFile]. Typical analysis procedures can be carried out without any programming. (4) Input an accession number and click on [submit]. Then, a result about the DDBJ entry is displayed DDBJ currently provides 8 workflows. them all sectors multiplicities sequences Workflow Description - Minista reach to BLALT - BLAST assessed DOB2 to DNA sequence Retrieve DOBJ for the entry by accession Homology workflow Search other species which have genes similar to human genes (5) Human chromosome gene workflow Show the number of genes on each chromosome BLAST against DDBJ by DNA sequence BLAST against (06) delabere by using DNA requence as a garry it may rate of Nucleotide frequency workflow Report the pattern of nucleotide frequency distribution. (5) Click on [BLAST against DDBJ by DNA sequence] in [Next flow] tab. Blast-ClustalW workflow Run blastn and compare alignment regions of high similar sequences BLAST workflow Run multiple BLAST against DDBJ, UniProtKB/Swiss-Prot, and PDB. (6) etta. - Patrave 2000 tiet tie etta: braistatett tunter - Patrieve 2000 anges braistatetter Splicing workflow Compare the similarities between splicing structure and homology. OMIM workflow Compare the similarities of human disease genes among eukaryotes (6) Jump to the BLAST search page, where [sequence] field is automatically filled with the Na Augustein number. Bitalisten andre auf erseinen Aus (2004) his suivent Au-steration (2004) his suivent Au-senta (2004) his suivent Au-senta (2004) his suivent Au-senta (2004) his suivent Au-stances (2004) his hims tanzent (2004) suivent Au-materia (2004) suivent Au-materia (2004) suivent Au-materia (2004) suivent Au-stances (2004) suivent Au-stances (2004) suivent Au-stances (2004) suivent Au-Materia (2004) suivent A SNP workflow Extract the relation between a human gene and SNP result of the previous service. Disease-related gene workflow Extract a group of associated genes with a given disease (7) Check the options and Click on [submit]. Then you can obtain the BLAST result. How to Use DDBJ Workflows in to these algometry New 201 (1) Click on [Workflow] of WABI top page http://www.xml.nig.ac.jp/ (7) (2) Click on [SNP workflow]. (3) Input a keyword about human genes (e.g. "transporter") and click on [Search]. CookBook (4) Information on protein 3D structures and disorder regions for the genes is shown in a table like below. Provide wiki-style Web pages to share know-how in usage of WABI services, such as "How can we execute phrase search by ARSA?" and "How can we retrieve BLAST Image and Construction Statistics Non-N Image in Statistics Statistics Statistics Image in Statistics e result with an XML format?". Example programs are also available.

The work is supported by Institute for Bioinformatics Research and Development (BIRD), Japan Science and Technology Agency (JST).