TeCK Database: A comprehensive collection of telomeric and centromeric sequences with their associated proteins

Ragul Gowthaman*, Saranya Krishnamoorthy, Renuka Devi Nandakumar, Vishalini Ayyarappan

Department of Bioinformatics, School of Chemical and Biotechnology, Shanmugha Arts Science Technology and Research Academy (SASTRA) SASTRA University, Thanjavur, India; Ragul Gowthaman* - Email: g.ragul@gmail.com; * Corresponding author

Abstract:
Telomeres and centromere are two essential features of all eukaryotic chromosomes. They provide function that is necessary for the stability of chromosomes. We developed a comprehensive database named TeCK, which covers a gamut of sequence and other related information about telomeric patterns, telomere repeat sequences, centromere sequences and centromeric patterns present in chromosomes. It also contains information about telomerase ribo-nucleoprotein complexes, centromere binding protein and centromere DNA-binding protein complexes. The database also includes a collection of all kinetochore-associated proteins including inner, outer and central kinetochore proteins. The database can be searched using a user-friendly web interface.

Availability: http://www.bioinfosastra.com/services/teck/index.html

Keywords: telomere; centromere; chromosome; kinetochore proteins

Background:
Telomeres [1] are a series of repeated DNA sequences located at the end of chromosomes, with strand asymmetry in GC content, resulting in one G-rich strand and one C-rich strand. They serve to assure that a chromosome is replicated properly during cell division. During this process some of the telomeres are lost. Eventually little or no telomere remains, and the cell dies. [2] Telomerase is an enzyme that adds specific DNA sequence repeats, ("TTAGGG" in all vertebrates) to the 3' end of DNA strands, in the telomere regions at the ends of chromosomes, thus preventing it from getting shorter. Centromeres [3] are DNA sequences contained in the heterochromatin, responsible for the segregation of each chromosome into daughter cells during cell division. The kinetochore [4] assembled on the centromere links the chromosome to the mitotic spindle during mitosis. As a cell tends to become cancerous [5], it divides more often, and its telomeres become very short, causing death of cells, which can be prevented by activating the enzyme telomerase. By understanding these regions of chromosomes, cancer and aging can be treated effectively and thus is a potential target of drug design approach. Hence, a database for analyzing and interpreting chromosomal information is essential. So far there had been no attempts to annotate and curate these parts of chromosomes.

Methodology:
TECK is developed using MySQL [6] a relational database management system that serves as the backend for storing data. IIS (Internet Information Server) is used as the web server and PHP5 (Hypertext preprocessor) [7] a widely used scripting language driven by Zend engine is used as the web interface.

The nucleotide information is retrieved from NCBI, [8] a national resource repository for molecular biological information. About 160 entries are obtained for both telomere and centromere. The information about the associated proteins is retrieved from SWISSPROT, [9] a protein database. About 154 entries are obtained for telomerase, 162 entries are obtained for centromere binding proteins and 170 entries are obtained for kinetochore. Hyperlinks are provided to the corresponding databases for each entry with their IDs displayed.

Each entry is given an unique ID called TeCK ID consisting of three parts, the first part indicating the class and the second depicting the organism and the last denoting the position of that particular entry in the table. About 16 tables had been created all of which are linked using their unique TeCK ID. Each protein entry is described using twenty parameters such as protein name, gene name, source, lineage, function, other interacting proteins etc. The

nucleotide information is described using parameters such as chromosome number, keywords, lineage, the articles along with the author details and references. BLAST **[10]** is a search tool for comparing biological sequences. A stand alone BLAST Program obtained from NCBI has been installed for searching similar sequences.

Usabilty / Accessibility:
TeCK database can be accessed via internet and the screen shot of the homepage is displayed [Figure 1A]. There are three ways by which the user can query the database. The first one is the keyword search in home page that can be performed for five categories (Telomere, Centromere and its associated proteins along with Kinetochore proteins) by giving any text data related to organism name, protein name, function etc. to retrieve the required data. The

second one is the advanced search option for specific requirements [Figure 1B], where the user can fill the form by specifying the input parameters such as SwissProt, Interpro, Pfam, BLOCKS, PDB, ModBase IDs or TeCk ID. Alternatively, the users can also browse the database [Figure 1C] by organisms and category. On selection of the required query, a result page containing the number of hits and the list of all entries ID with a short description is displayed. Detailed information can be obtained by clicking each entry ID [Figure 1D]. The sequences are displayed in FASTA format and the retrieval can be done using a separate 'retrieve in FASTA format' link. BLASTn can be done against TeCK database & NCBI and BLASTp against TeCK database, SwissProt & PDB. [Figure 1E & Figure 1F].
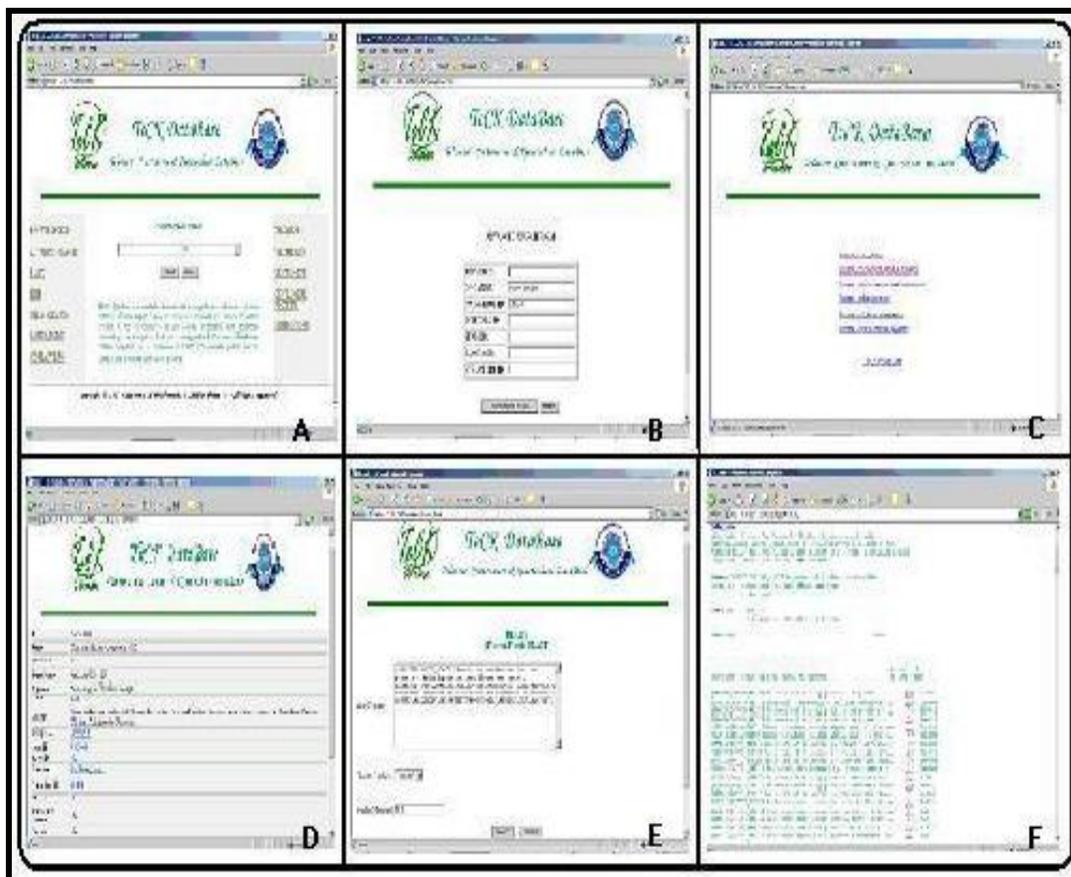

Figure 1: Screen captures of the database GUI. (A) TeCK Database Home page, (B) Advanced search page, (C) Browse page (D) Entry page (E & F) BLAST pages

Utility:
As TeCK database includes name, function, sequence, subunits, and other literature information. The data for TeCK database were obtained from various curated databases and hence TeCK database helps end users to

study, compare, analyze and interpret them and finds utility in control of cancer and ageing.

Caveats:
It should be noted that the consistency depends on the source of the original data.

References:
[01]   E. Hiyama, *et al.*, *Br J Cancer,* 96:1020 (2007) [PMID: 17353922]
[02]   http://learn.genetics.utah.edu/features/telomeres/
[03]   B. Chatterjee, *et al., Nucleic Acids Res.,* 16:1201 (1988) [PMID: 3344208]
[04]   M. S. Kwon, *et al., Mol. Biol. Cell,* 18: 2155 (2007) [PMID: 17392512]
[05]   N. Miura, *et al., Cancer Genet Cytogent.,* 93: 56 (1997) [PMID: 9062581]
[06]   http://www.mysql.com
[07]   http://www.php.net
[08]   http://www.expasy.org/sprot
[09]   http://www.ncbi.nlm.nih.gov/
[10]   S. F. Altschul *et al.,* Nucleic acid Res., 25: 2389 (1997) [PMID: 9254694]

Edited by P. Kangueane

Citation: Gowthaman *et al.*, Bioinformation 2(2): 73-75 (2007)