

Charm (and DengueInfo)

<http://dengueinfo.org/>

Holland R.C.G., Ong S.H., Verhoef F., Mitchell W.P., Schreiber M.J.

Richard Holland, BOSC 2005

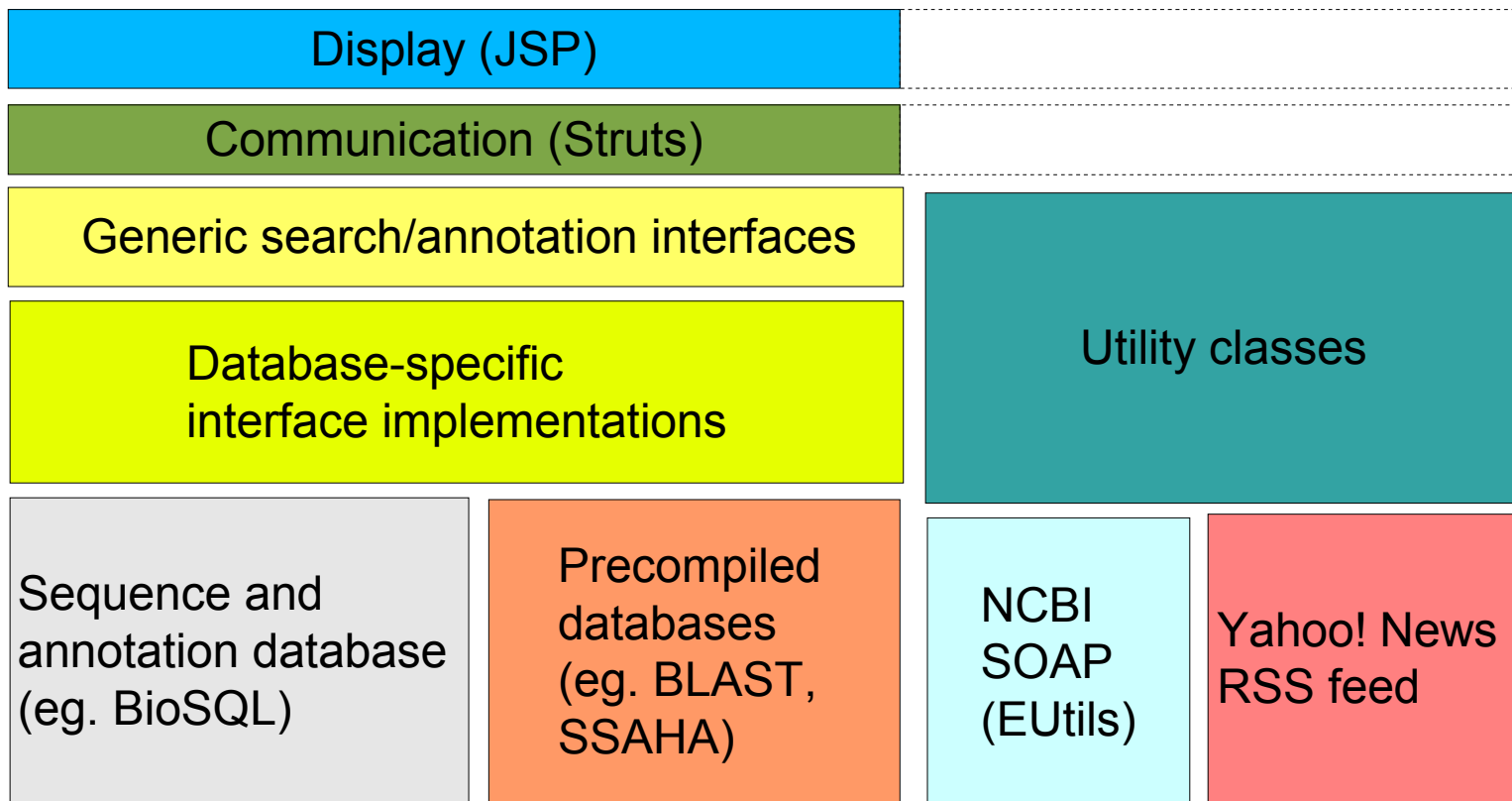
Background

- Dengue is a serious infectious tropical disease transmitted by the mosquito *Aedes aegypti* during feeding.
- **No drugs exist** for the specific treatment of dengue.
- NITD and GIS are collaborating on drug development.
- Very **small genome**.
- Complete genome **infrequently sequenced** to date.
- Needed a **searchable repository** for dengue genomes **annotatable with clinical information**.

Charm

- **Generic webapp** to interact with an **existing annotatable sequence database**.
- Defines an **extensible custom annotation ontology**.
- Able to store sequences and annotate them, and perform **complex searches**.
- **Easily extensible**, easy to create specialised versions such as DengueInfo.

Charm architecture



Searches – a bit like BIND

Complex Sequence Search

Find records where **Any** of the following conditions are met:

1. Accession

2. Date of Isolation

3. Country

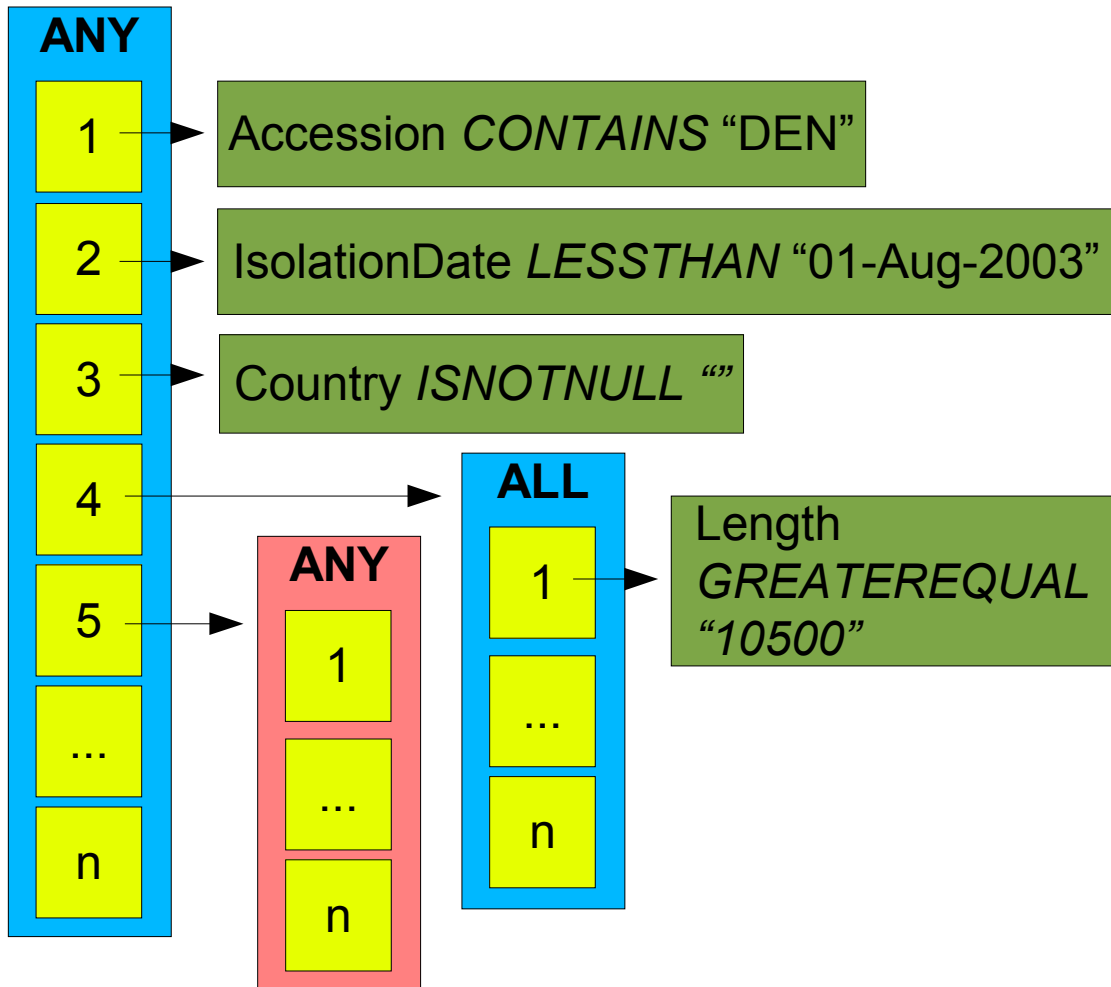
All of the following conditions are met:

1. Length

...but exclude records where **Any** of the following conditions are met:

1. Accession

Searches



- Search objects have an **ANY/ALL** flag.
- **Recursive definition.** Each term can be a...
 - field/method/value triple (**CONDITION**).
 - search object (**SUBQUERY**).
 - search object flagged to exclude matching results (**EXCLUSION**).

Searches

- Each condition is translated and **executed individually** to retrieve a set of unique IDs.
- For “**ANY**” searches, the results are the **set union** of all returned IDs.
- For “**ALL**” searches, the results are the **set intersection** of all returned IDs.
- Subqueries and Exclusions are executed as **independent searches** and their **results combined with the parent search** using union, intersection, or subtraction as appropriate.

Other searches

- **BLAST** (calls out to NCBI command line binaries, Oracle 10g reference code provided if required)
- **SSAHA** (BioJava's implementation)
- Current implementations use **preformatted databases** on disk, **rebuilt only on request** via web interface.

Search results

- Results are sets of **unique IDs** with **scores**.
- Search definition and results are **stored in session variables** to prevent needless re-entry or re-execution.
- Actual sequence details not stored, to save memory.
- Search results screen provides some basic manipulations.

Results

You searched for **{MATCH ANY [ACCESSION CONTAINS DEN]}**. There were **8** results.

Select... All None	Score	Display ID	Description
<input type="checkbox"/>	1.0	DENT1SEQ	Dengue virus type 1 complete genome.
<input type="checkbox"/>	1.0	DENSTRA	Dengue virus type 4 polyprotein precursor, gene, complete cds.
<input type="checkbox"/>	1.0	DENRCG	Dengue virus type 2 (S1 vaccine strain), complete genome.
<input type="checkbox"/>	1.0	DENCMEMSB	Dengue virus type 2 strain 16681 capsid protein (C) gene, matrix protein (M) gene, envelope protein (E) gene, nonstructural protein (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) genes, mature peptides.
<input type="checkbox"/>	1.0	DENCMEMSA	Dengue virus type 2 strain 16681-PDK53 capsid protein (C) gene, matrix protein (M) gene, envelope protein (E) gene, nonstructural protein (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5) genes, mature peptides.
<input type="checkbox"/>	1.0	DENCME	Dengue type 3 virus complete genome RNA, complete cds.
<input type="checkbox"/>	1.0	DEN2JAMCG	Dengue virus type 2 Jamaica/N.1409, complete genome.
<input type="checkbox"/>	1.0	DEN2CGA	Dengue virus type 2 complete genome.
All None			

Trim by position:	From <input type="text" value="0"/>	to <input type="text" value="0"/>	(0 for unlimited)
Translate to protein?	<input checked="" type="radio"/> No translation <input type="radio"/> Frame 1 <input type="radio"/> Frame 2 <input type="radio"/> Frame 3		
	<input type="button" value="Display Fasta"/>		
	<input type="button" value="Display Genbank-style extract"/>		
	<input type="button" value="Annotate"/>		

Annotation

- Can only annotate using terms from the custom ontology.
- Manual annotation done by selecting sequence accessions and entering **term/value pairs**.
- Automatic annotation done by adding code to the appropriate **middleware method** (called **once per batch** of sequences uploaded).

Manual annotation

Selected sequence(s):	DENT1SEQ
	<input type="button" value="Change selection"/>

Existing annotations:

- **AnnotatorNote** (DENT1SEQ)
- **Country: SG** (DENT1SEQ)
- **Modified: Wild_Type** (DENT1SEQ)
- **Owner: PUBLIC** (DENT1SEQ)
- **Serotype: DENV-1** (DENT1SEQ)
- **Severity: DHF** (DENT1SEQ)
- **Strain** (DENT1SEQ)
- **YearOfIsolation** (DENT1SEQ)

Term with value:	<input type="text" value="AnnotatorNote"/>	<input type="button" value="Annotate"/>
	<input type="text"/>	

Other features

- Password protection of annotation and admin tasks.
- Export/Import whole database via zip file.
- Add sequences manually (FASTA-like interface).
- Add sequences from GenBank files.
- Remove sequences.
- Export/Import the custom ontology as XML file (useful for adding new terms).

DengueInfo – a Charm extension

- Charm is **generic**, designed to be extended and specialised.
- Some utility classes are not used in basic implementation – written specifically for use by extended versions.
- **DengueInfo** is an example of how Charm can be extended to **suit a specialist task**.

PubMed feed

Recent PubMed Abstracts

Below are listed the 20 most recently published abstracts from PubMed which relate to dengue. Recent is defined as being within the last 90 days. These abstracts are all subject to the [NCBI Disclaimer and Copyright notice](#).



Dengue vaccines: Problems & prospects.

Chaturvedi UC, Shrivastava R, Nagar R.

201-Annapurna Apartments, No. 1, Bishop Rocky Street, and Faizabad Road, Lucknow 226007, India.

The extent of cumulative disease burden caused by dengue virus has attained an unprecedented level in recent times with sharp increase in the size of human population at risk. Dengue disease presents highly complex medical, economic and ecologic problems. The surge in publications on the development of dengue vaccines, taking advantage of new generation of biotechnology techniques indicates the profound interest and urgency in the scientific and medical communities in combating this disease. This review summarizes the importance of critical subjects like pathogenesis of dengue haemorrhagic fever and inadequacy of animal model that have adversely affected dengue vaccine development. Further, the remarkable progresses so far made in dengue vaccine research not only employing a diverse range of new strategies but also re-using old techniques to improve the existing vaccines, have been presented. The efficacy and safety of some of the new vaccine candidates have been evaluated and proven in human preclinical/clinical trials. Besides the technical advancement in vaccine development, vaccine safety and vaccine formulation have been examined.

Yahoo News feed

Recent Dengue News

[Putting more bite into fighting dengue \(@Metro KL\)](#)

Thu Jun 09 08:21:50 SGT 2005

A FTER receiving reports of more than 5,000 dengue cases since January, the Selangor Health Department has decided to increase its efforts to fight the menace.



[Dengue epidemic year alert up \(Sun Star\)](#)

Wed Jun 08 00:22:09 SGT 2005

THE City Health Office (CHO) alerted the public against a possible upsurge of dengue cases with 2005 declared as dengue epidemic year. Health officer Dr. Erlinda Cabrera reiterated her advise to the public to remove all mosquito breeding places.

[City Hall braces for floods, dengue cases \(Sun Star\)](#)

Tue Jun 07 15:03:29 SGT 2005

THE Iloilo City Government is stepping up preparations to mitigate flash flooding and the outbreak of dengue fever during the rainy season. Mayor Jerry Trefias said the declogging of the city's waterways in flood-prone areas is still ongoing.

[New dengue prevention legislation soon \(Daily News\)](#)

Tue Jun 07 03:53:30 SGT 2005

The Health Ministry will strictly implement the new Dengue prevention legislation that will be presented in Parliament shortly to completely eradicate the epidemic from the country.

Other bits

- Expanded custom ontology.
- Auto-annotation of serotypes and structural components.
- Annotators Notes.
- Synchronise with NCBI to download latest Dengue genomes.
- Additional terms available for searching.
- Additional options available for working with search results.

Clinical Information

- 1.0 DENCME Dengue type 3 virus complete genome RNA, complete cds.
- 1.0 DEN2JAMCG Dengue virus type 2 Jamaica/N.1409, complete genome.
- 1.0 DEN2CGA Dengue virus type 2 complete genome.

All | None

Trim by position: From to (0 for unlimited)

Trim by component: From to (overrides trim by position)

Translate to protein? No translation Frame 1 Frame 2 Frame 3

Accession	Serotype	Strain	Modified	Severity	Country	Date of Isolation	PubMed ID(s)	Info
DEN2CGA	DENV-2	New Guinea-C			PG	1944	2964755 2714651	
DEN2JAMCG	DENV-2	N1409			JM	1983	3388770	
DENCME	DENV-3	H87			PH	1957	2345967	
DENCMEMSA	DENV-2	16681-PDK53	Recombinant				2719554 1829885 1312269	
DENCMEMSB	DENV-2	16681	Wild_Type	DSS	TH	1964	2719554 1829885 1312269	
DENRCG	DENV-2	S1	Wild_Type	DF	PR	1969	2827375	
DENSTRA	DENV-4	814669	Wild_Type		DO		3022479 3039728	
DENT1SEQ	DENV-1	S275/90	Wild_Type	DHF	SG	1990	1585663	

Wrinkly bits

- BioJava's BioSQL support was found to be a bit flaky.
- Ontology persistence couldn't handle triples or term synonyms.
- Oracle support just didn't work at all if you used Oracle 9i or greater, due to API changes for accessing LOBs (Large Objects, anything > 4000 bytes).
- Order of annotations not preserved.
- Genbank parsers did not export References.
- All has been **fixed** and **contributed** back to BioJava.
- Working on plans to synchronise the way BioJava and the other Bio* projects use BioSQL.

Scalability

- Currently has 142 sequences, all from GenBank.
- Expect 400 by this time next year.
- Unfriendliness of UI for manual annotation will soon become apparent – data just won't fit on screen.
- Filesize and slowness of export/import database options will become more noticeable as database size increases.
- Search results will need paginating.
- Charm version 2 specifications are under development, scalability (and security) will be a priority.

Future Plans

- Being **open source**, we hope people will use Charm and contribute their ideas.
- Plans to add **free-text indexing** and searching of documents and papers.
- Make annotations editable/removable.
- Security needs work:
 - organise users into groups and implement '**copyright**' of private or protected sequences.
 - implement **tracking of changes** (additions, deletions, annotations) by username.
 - remove reliance on Tomcat-specific mechanisms (roles) to enable deployment on other application servers

Where to get it?

- To use DengueInfo, an example of Charm extension:
 - <http://dengueinfo.org/>
- Source code, Javadocs, WAR files, custom ontologies, NCBI Java client, and installation guides:
 - <http://dengueinfo.org/dist/>
 - ontology XSD is in the `web` folder of the source code of both projects.
- To use a barebones version of Charm (running off the DengueInfo database):
 - <http://dengueinfo.org/charm/>

Acknowledgements

- **Mark Schreiber** (NITD) for the concept, providing a web server and database to run it on, and code contributions.
- **Ong Swee Hoe** (GIS) for annotations and feedback.
- **Frans Verhoef** (GIS) for code contributions and feedback.
- **John Salama** (Blueprint) for insights into BIND.
- **Hilmar Lapp** (OBF) for suggesting improvements.
- **Wayne Mitchell** (GIS) for guidance and coffee.

References

- **BIND** (<http://bind.ca/>)
 - Bader G.D., Betel D., Hogue C.W. (2003) BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* 31(1):248-50.
- **BLAST** (<http://www.ncbi.nlm.nih.gov/BLAST>)
 - Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- **ODM BLAST** (<http://www.oracle.com/>)
 - Stephens S.M., Chen J.Y., Davidson M.G., Thomas S. and Trute B.M. (2005) Oracle Database 10g: a platform for BLAST search and Regular Expression pattern matching in life sciences. *Nucleic Acids Research*, Vol. 33, Database issue D675-9.
- **SSAHA** (<http://www.sanger.ac.uk/Software/analysis/SSAHA/>)
 - Ning Z., Cox A.J., Mullikin J.C. (2001) SSAHA: a fast search method for large DNA databases. *Genome Res.* 2001;11;1725-9.