Annexure -I

Sardar Patel University

Vallabh Vidyanagar - 388120

Final Report of the work done on the Interdisciplinary Research Project
(Report to be submitted by 15/04/2013)

1. University Reference No. G/No. DST Purse/12-13/1674

2. Period of report: from 06/06/2012 to 30/03/2013

3. Title of research project "Heli-PyD", a Comprehensive Database for Helicobacter Pylori

4. (a) Name of the Principal Investigator DR. Hetalkumar Panchal

(b) Deptt. where work has progressed

G.H.Patel P.G. Dept. of Computer Science & Technology, M.C.A, SPU

5. Effective date of starting of the project 06/06/2012

6. Grant approved and expenditure incurred during the period of the report:

a. Total amount approved Rs. 1 Lakh only/-

b. Total expenditure Rs. 1,32,324/—

C. Report of the work done: (Please attach five pages in a separate sheets) (In enclosure)

i. Brief objective of the project To Designe a Comprehensive Database for Helicobacter Pylori

ii. Work done so far and results achieved (in the following format)

a) Definition of problem handled

b) Methodology adapted

c) Detailed work (including prominent figures/data)

d) Results and Conclusion

e) Details of publications (including conference presentation)

iii. Has the progress been according to original plan of work and towards achieving the objective. If not, state reasons.

"Heli-PyD", A Comprehensive Database for *Helicobacter Pylori*

✓ Brief Objective:

To construct integrated database of *H. Pylori* consisting all type of sequential, structural and functional information at a single platform.

✓ Definition of problem handled:

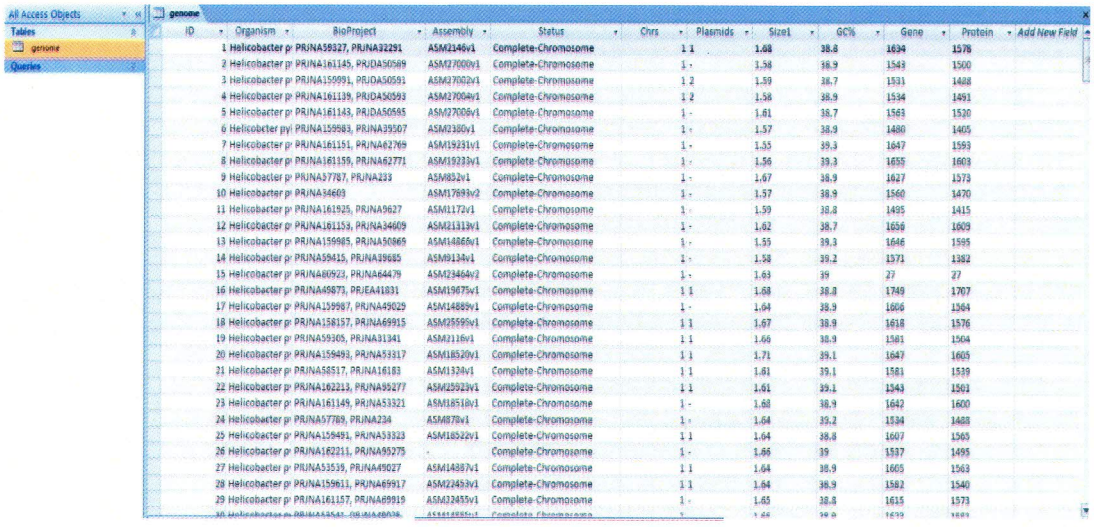
Heli-PyD a comprehensive Database for *Helicobacter pylori*

✓ Methodology adapted:

The database was designed using combination of various computational languages and database development environment. The framework of processed data was formulated using Microsoft access followed by its data normalization and database creation. An easy to use and access front end of database was developed by combination of HTML and DOTNET computational languages via package "Visual studio".

✓ Detailed work (including prominent figures/data):

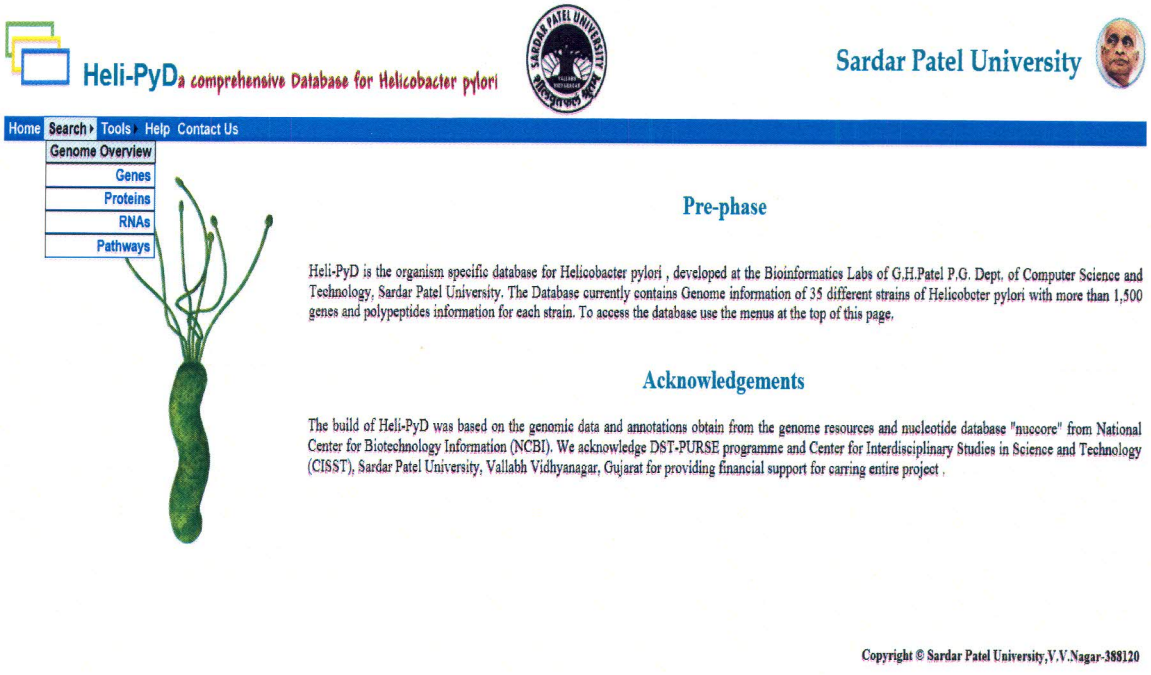
1. Primarily the required biological data of Helicobacter pylori was collected from different biological online repositories and a compiled datasheet was produced after data filtration using Microsoft access as follow:



ID	Organism	BioProject	Assembly	Status	Chrs	Plasmids	Size1	GC%	Gene	Protein	Add New Field
1	Helicobacter p	PRJNA59327, PRJNA32291	ASM2146v1	Complete-Chromosome	1	1	1.68	38.8	1634	1578	
2	Helicobacter p	PRJNA161145, PRJDA50589	ASM27000v1	Complete-Chromosome	1	-	1.58	38.9	1543	1500	
3	Helicobacter p	PRJNA159991, PRJDA50591	ASM27002v1	Complete-Chromosome	1	2	1.59	38.7	1531	1488	
4	Helicobacter p	PRJNA161139, PRJDA50593	ASM27004v1	Complete-Chromosome	1	2	1.58	38.9	1534	1491	
5	Helicobacter p	PRJNA161143, PRJDA50595	ASM27006v1	Complete-Chromosome	1	-	1.61	38.7	1563	1520	
6	Helicobcter pyl	PRJNA159983, PRJNA39507	ASM2380v1	Complete-Chromosome	1	-	1.57	38.9	1480	1405	
7	Helicobacter p	PRJNA161151, PRJNA62769	ASM19231v1	Complete-Chromosome	1	-	1.55	39.3	1647	1593	
8	Helicobacter p	PRJNA161159, PRJNA62771	ASM19233v1	Complete-Chromosome	1	-	1.56	39.3	1655	1603	
9	Helicobacter p	PRJNA37787, PRJNA233	ASM852v1	Complete-Chromosome	1	-	1.67	38.9	1627	1573	
10	Helicobacter p	PRJNA34603	ASM17893v2	Complete-Chromosome	1	-	1.57	38.9	1560	1470	
11	Helicobacter p	PRJNA161925, PRJNA9627	ASM1172v1	Complete-Chromosome	1	-	1.59	38.8	1495	1415	
12	Helicobacter p	PRJNA161153, PRJNA34609	ASM21313v1	Complete-Chromosome	1	-	1.62	38.7	1656	1609	
13	Helicobacter p	PRJNA159985, PRJNA50969	ASM14865v1	Complete-Chromosome	1	-	1.55	39.3	1646	1595	
14	Helicobacter p	PRJNA59415, PRJNA39685	ASM9134v1	Complete-Chromosome	1	-	1.58	39.2	1371	1382	
15	Helicobacter p	PRJNA60923, PRJNA64479	ASM23464v2	Complete-Chromosome	1	-	1.63	39	27	27	
16	Helicobacter p	PRJNA49873, PRJEA41831	ASM19675v1	Complete-Chromosome	1	1	1.68	38.8	1749	1707	
17	Helicobacter p	PRJNA159987, PRJNA49029	ASM14889v1	Complete-Chromosome	1	-	1.64	38.9	1606	1564	
18	Helicobacter p	PRJNA158157, PRJNA69915	ASM25595v1	Complete-Chromosome	1	1	1.67	38.9	1618	1576	
19	Helicobacter p	PRJNA59305, PRJNA31341	ASM2116v1	Complete-Chromosome	1	1	1.66	38.9	1581	1504	
20	Helicobacter p	PRJNA159493, PRJNA53317	ASM18520v1	Complete-Chromosome	1	1	1.71	39.1	1647	1605	
21	Helicobacter p	PRJNA58517, PRJNA16183	ASM1324v1	Complete-Chromosome	1	1	1.61	39.1	1581	1539	
22	Helicobacter p	PRJNA162213, PRJNA95277	ASM25923v1	Complete-Chromosome	1	1	1.61	39.1	1543	1501	
23	Helicobacter p	PRJNA161149, PRJNA53321	ASM18518v1	Complete-Chromosome	1	-	1.68	38.9	1642	1600	
24	Helicobacter p	PRJNA57789, PRJNA234	ASM878v1	Complete-Chromosome	1	-	1.64	39.2	1534	1488	
25	Helicobacter p	PRJNA159491, PRJNA53323	ASM18522v1	Complete-Chromosome	1	1	1.64	38.8	1607	1565	
26	Helicobacter p	PRJNA162211, PRJNA95275	-	Complete-Chromosome	1	-	1.66	39	1537	1495	
27	Helicobacter p	PRJNA53539, PRJNA49027	ASM14887v1	Complete-Chromosome	1	1	1.64	38.9	1605	1563	
28	Helicobacter p	PRJNA159611, PRJNA69917	ASM22453v1	Complete-Chromosome	1	1	1.64	38.9	1582	1540	
29	Helicobacter p	PRJNA161157, PRJNA69919	ASM32455v1	Complete-Chromosome	1	-	1.65	38.8	1615	1573	

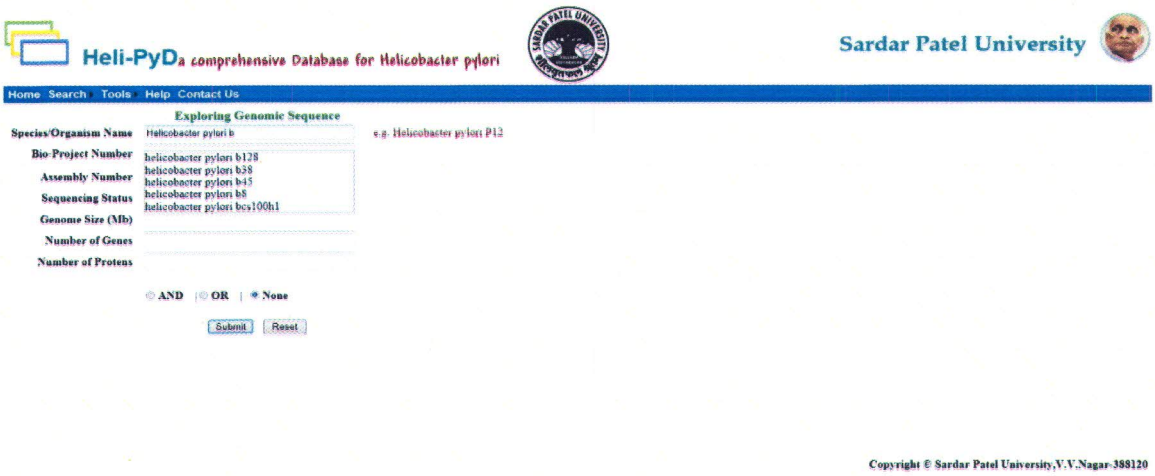
2. The secondary genomic information which were not available, were generated using different bioinformatics tools. (e.g. : Genomic- CGPlot)

3. The datasheet were then normalized and updated to fit the created data structure of database.
4. Considering the front end, precisely the front outlook of database was created using HTML. Home page of Database and the menus were created using HTML coding.



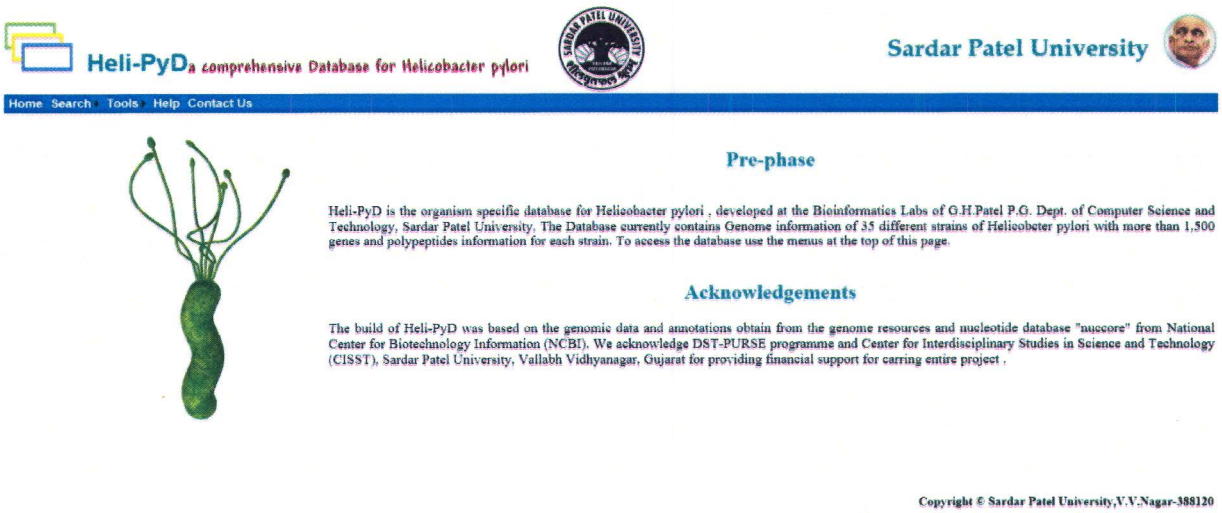
5. The logical connection of front end with the core database was created using DOTNET language. (Attachment: front look and logical Code of genome browser page) The logical coding is done in such a way that it facilitates user to auto fill almost all the possible textual field while submitting the search details

6. Each broad search spectrum mainly has multiple inputs filed. To facilitate user, “and”, “or” and “none” radio buttons are provided; so user can search by single and multiple criteria or can search by combination of these search field according to their choice and need.



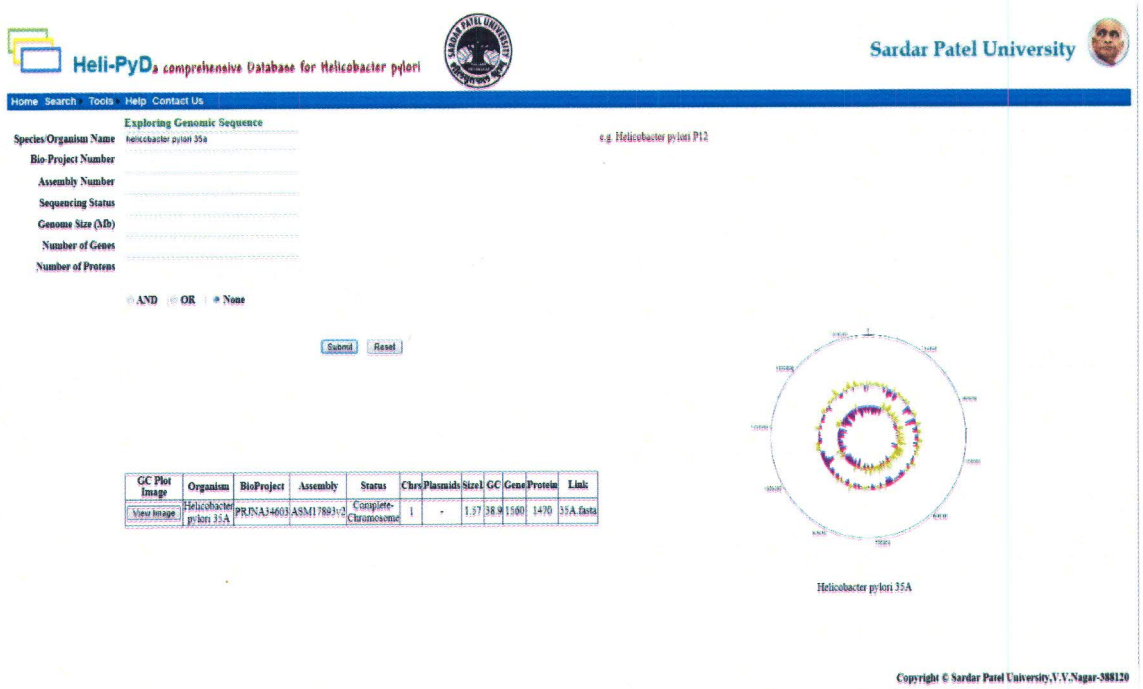
✓ <u>Results and Conclusion:</u>

- The Home page of database is as follow:



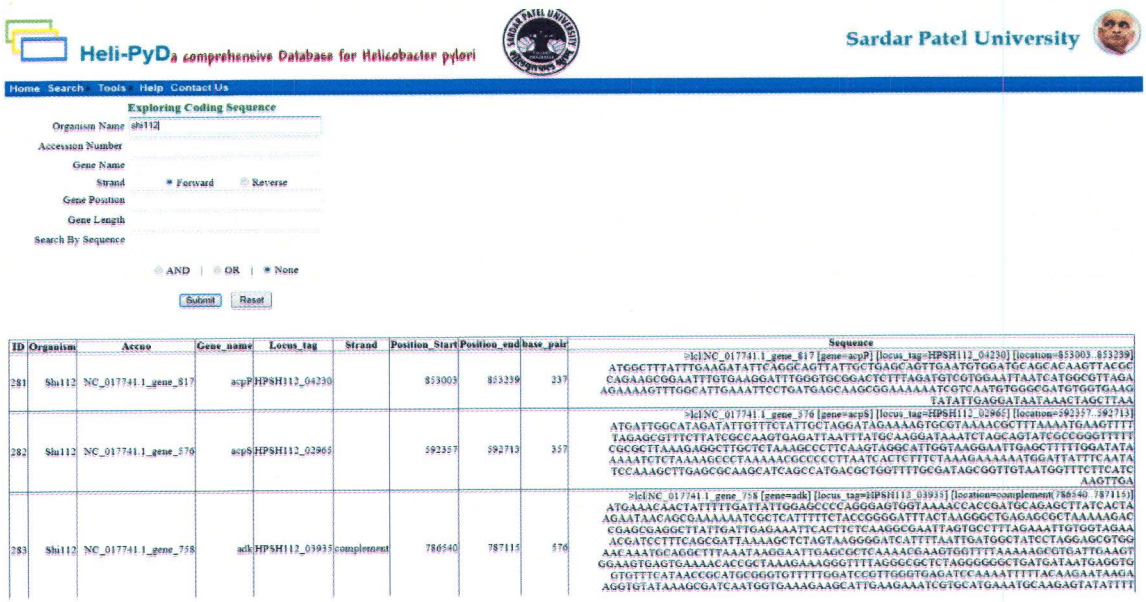
- The search menu of database has options that allow searching database for overall genomic data as well as genes related data. The overall genomic information of each strain can be view by using "Genome Overview "browser. The data can be retrieved by input of either organism name, status of sequencing, sequencing bio-project number,

genome size and more available options



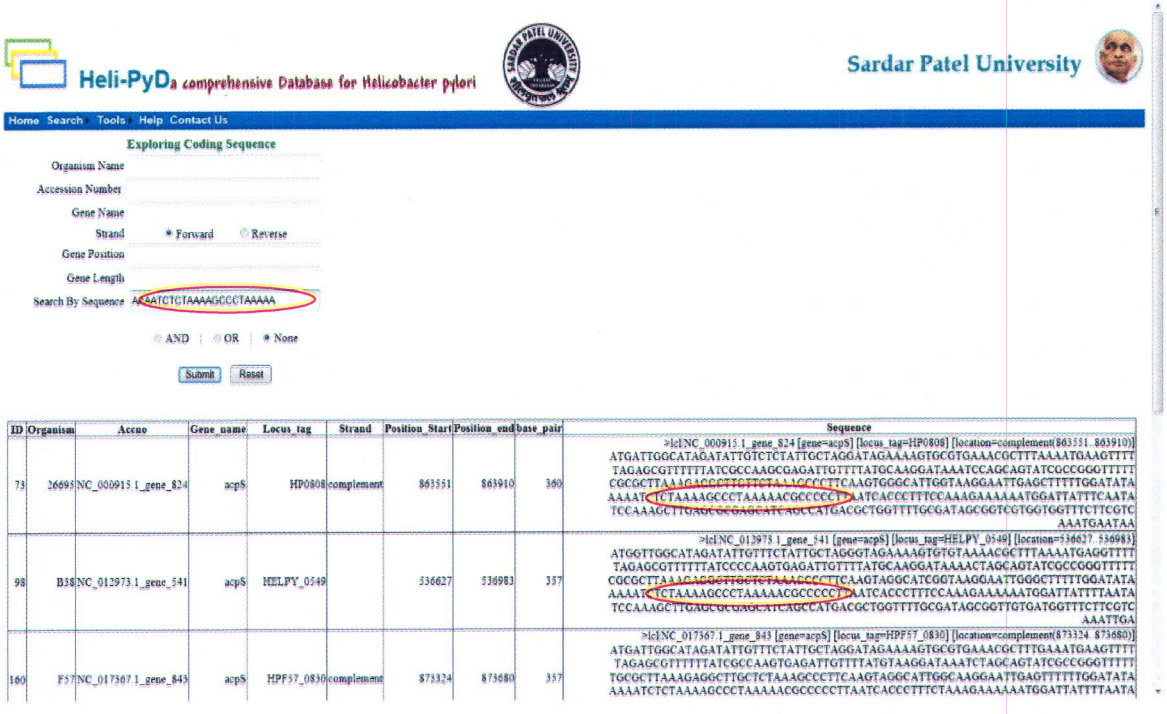
- In the above show figure the table at the bottom and GC-plot at the right of webpage represents the search for organism/species Helicobacter pylori 35A. The plot will be displayed only after clicking button "view image" in first column of result table likewise the last column link will direct user to the genomic sequence of that organism/species, from where it can be saved easily.

- Another option of search menu is "genes" which will allow browsing all the coding sequences of each specified genome.



- Database allows retrieving all the gene related information by input of any one field like Gene name, Gene length, its position, its strand preference or a stretch of nucleotide sequence. The auto fill option is available for all the above specified fields and the output is obtained in tabular format. In the above figure user has tried to enlist the genes of strain *H. pylori* shi112 which are specifically present on forward strand only. Likewise N –number of field combinations can be used according to the need of user.

- The sequence base search is biologically one of the most significant search options, which is allows the user to identify genes sharing same stretch of sequences. Thus one can get a rough idea of sequence homology within the species and between different species.



✓ <u>Details of publications (including conference presentation)</u>

Megha Vaidya and Dr. Hetalkumar Panchal, "Comparative study of data generated by structural annotation of the genome for identifying the unique parameter responsible for pathogenic property of *Helicobacter pylori*", at *3rd International Conference and Exhibition on Pharmaceutics & Novel Drug Delivery Systems"* April 8-10, 2013.

✓ <u>Has the progress been according to original plan of work and towards achieving the objective. If not, state reasons:</u>

- <u>Yes</u>

- ✓ Please indicate the difficulties, if any, experienced in implementing the project:
    - In ordering and getting the equipments and in payment of those bills.
- ✓ If project has not been completed, please indicate the reasons:
    - It is completed
- ✓ Any other information which would help in evaluation of work done on the project: