김 상 수 박사

한국생명공학연구원 유전체연구센터 인간유전체연구팀 책임연구원

Tel. +82-42-866-7189, Fax. +82-42-860-4409

E-MAIL : sskimb@kribb.re.kr

Address : 대전광역시 유성구 어은동 52번지 <우:305-333>

◈ 연구관심분야

1. EST sequencing
2. Sequence assembly
3. DNA chip database & analysis

※ 21C 프론티어 사업 "인간유전체 기능 연구"의 바이오인포매틱스 책임자로서, 책임연구원 2명, 선임연구원 1명, 석사급 연구원 8명의 바이오인포매틱스팀을 운영하고 있음.

◈ 학 력

1. B.S. in Chemistry, Seoul National University., Korea, Feb. 1981.
2. M.S. in Physical Chemistry, Seoul National University., Korea, Feb. 1983.
3. Ph.D in Physical Chemistry, Iowa State University., U. S. A., Dec. 1986.

◈ 주요경력

2000. 3. ~ 현 재 한국생명공학연구원 유전체연구센터 책임연구원
1999. 1. ~ 2000. 2. (주) LG화학 Bioinformatics 팀장, 책임연구원
1995. 3. ~ 1998. 12. (주) LG화학 Biopharmaceutical Design 팀장, 책임연구원
1988. 12. ~ 1995. 2. (주) LG화학 Biopharmaceutical Design 팀장, 선임연구원
1986. 12. ~ 1988. 12. Dept. of Bio. Sci., Purdue University, West Lafayette (Prof. M.G. Rossmann), Postdoctoral Res. Ass.

◈ 학회활동

International Society for Computational Biology (ISCB), 한국생화학회, 대한화학회, 한국생물정보학회 회원

◈ 연구 실적 요약

- 학술잡지 논문발표 : 23 편
- 학술컨퍼런스 논문발표 : 9 편
- 국제학회 기조연설, 초청강연, 초청세미나 : 6 회
- 특허 : 14 건 (국제특허 7 건 포함)
- 연구과제 프로젝트 : 1 건

Bioinformatics for the Annotation of Helicobacter pylori isolated from a Korean patient and its comparison with isolates from Caucasian patients

Korea Res. Inst. of Biosci. & Biotech.
Genome Research Center
Sungsoo Kang & Sangsoo Kim

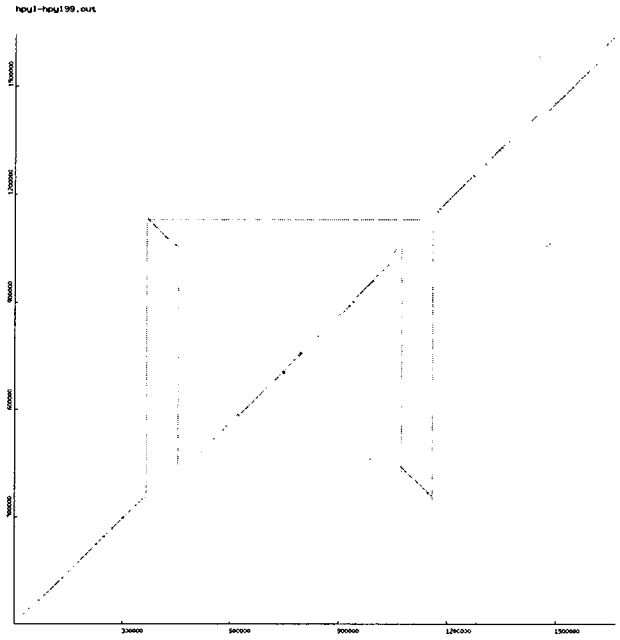
Funded by Ministry of Science & Technology
21C Frontier Program, "Human Genome Function Analysis"

Helicobacter pylori

- Causative agent of gastroduodenal disorders including gastritis, peptic ulcers & gastric cancer
- Two strains already sequenced
  - Strain 26695 (TIGR 1997)
  - Strain J99 (Astra/GTC 1999)

- 6 ~ 7% are strain-specific genes, 1/2 of them localized in the plasticity zone
- Even bigger genome diversity in Korean strains, not restricted to the plasticity zone
- Additional sequencing may help understand mechanism of antigenic variation and adaptive evolution

BLASTN Result of 26695 *vs* J99



Summary of Experiments

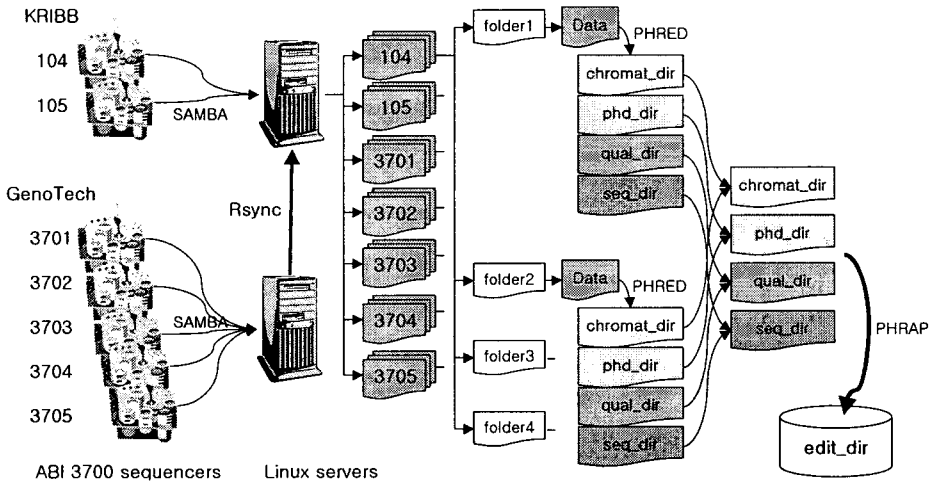
- *H. pylori* 51, Korean strain
- BAC clone physical map
- Shotgun library prep.
  - Avg 1.6~2kbp insert in pTZ19U vector

GSNU

- Robotic sequencing rxn
- ABI Prism 3700
- Forward-reverse mates
  - 20,798 reverse reads
  - 9,120 forward reads

KRIBB Genotech

Data Flow Diagram



Read name from sequencers are changed to conform to the Washington University convention used in PHRAP

104_6_A01_001_6_g.ab1

Sequencer number	Plate number	Well location	Capillary number	g = reverse b = forward

6_A01.g1

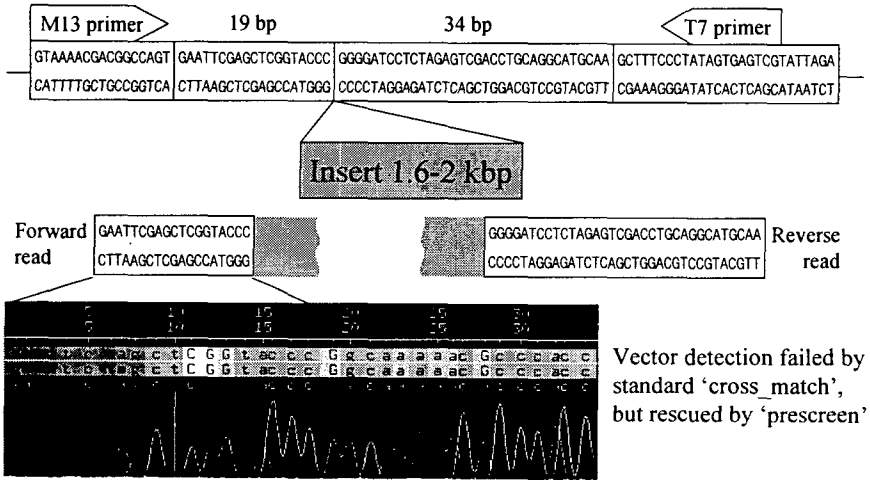
Template name used for matching forward-reverse mates

Quality Assessment

- Phred score
  - About 7% of total reads do not have significant stretch of high quality regions
  - Those plates exceeding this value have been resequenced
- Vector masking
  - Unremoved vector sequences may cause chimeric mis-assembly
  - First 80~100 residues masked for failures

Low trace quality in the beginning of reads hinders precise detection of the 19 bp vector sequence in forward reads by the standard procedure using 'cross_match'

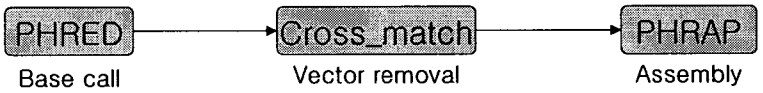
pTZ19U vector



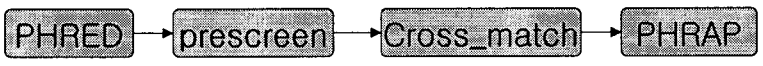
Vector detection failed by standard 'cross_match', but rescued by 'prescreen'

Folder name	# of reads	poor quality	good quality, but no vector by phred	total # of no vector by phred	no vector by FASTA
104/Helico_6_3700-104-KRIBB_Run_3700-104-KRIBB_2000-12-05_18	96	7	1	6	6
104/Helico_11_3700-104-KRIBB_Run_3700-104-KRIBB_2000-10-24_9	96	14	4	16	15
104/Helico_12_3700-104-KRIBB_Run_3700-104-KRIBB_2000-10-24_10	96	12	5	17	16
104/Helico_13_3700-104-KRIBB_Run_3700-104-KRIBB_2000-10-23_5	96	7	10	17	14
104/Helico_14_3700-104-KRIBB_Run_3700-104-KRIBB_2000-10-23_6	96	13	12	22	19
21_63					
104/Helico_82_3700-104-KRIBB_Run_3700-104-KRIBB_2000-12-21_64	96	6	88	94	9
104/Helico_84_3700-104-KRIBB_Run_3700-104-KRIBB_2000-12-21_65	96	12	80	92	23
104/Helico_196_3700-104-KRIBB_Run_3700-104-KRIBB_2000-12-22_66	96	1	85	86	5
104/Helico_200_3700-104-KRIBB_Run_3700-104-KRIBB_2000-12-26_71	96	5	89	94	27
104/Helico_198_3700-104-KRIBB_Run_3700-104-KRIBB_2000-12-26_70	96	2	83	85	8
Total	13344	863	3021	3715	1337
Percent total	100	6	23	28	10

Improved Vector Masking



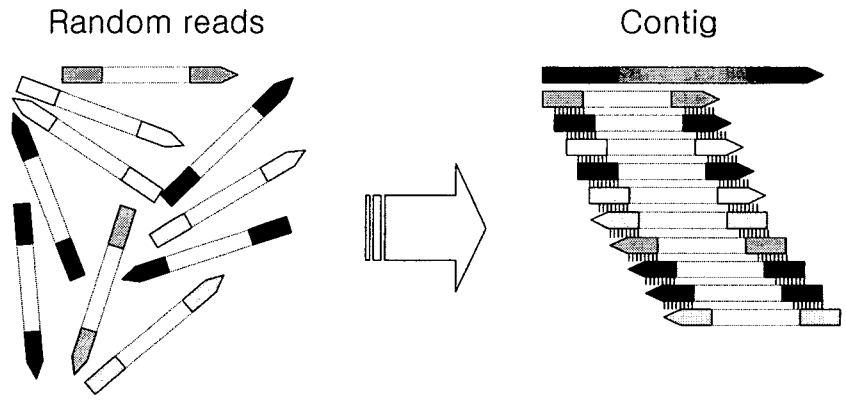
- The standard procedure for detecting vector sequences is not sensitive enough for low quality regions
- New procedure overcomes this problem by using FASTA program to detect vector sequences prior to 'cross_match' against UniVec
- 'prescreen'saves me mory and time in 'cross_match'



Summary of Sequencing

Sequencer ID	# of plates	Total reads	Poor quality	Good quality, no vector	Total # of failures Standard procedure	Total # of failures After prescreen
104	139	13344	863	3021	3715	1337
105	16	1536	93	1358	1451	267
3701	70	6720	313	3027	3294	1018
3702	16	1536	136	191	304	239
3703	11	1054	75	136	198	13
3704	10	960	85	86	152	126
3705	50	4800	281	1757	2005	823
Total	312	29950	1846	9576	11119	3823
%		100	6	32	37	13

Contig Assembly

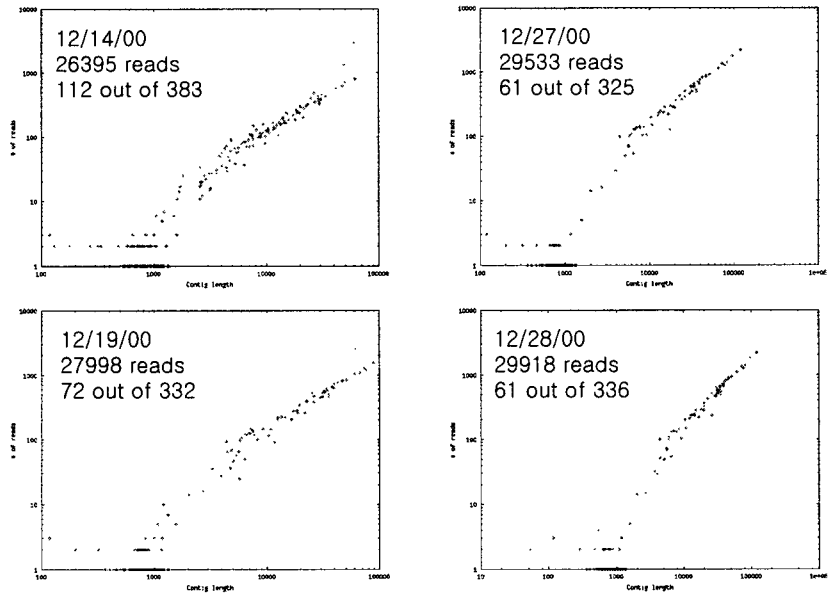


Phrap Assembly

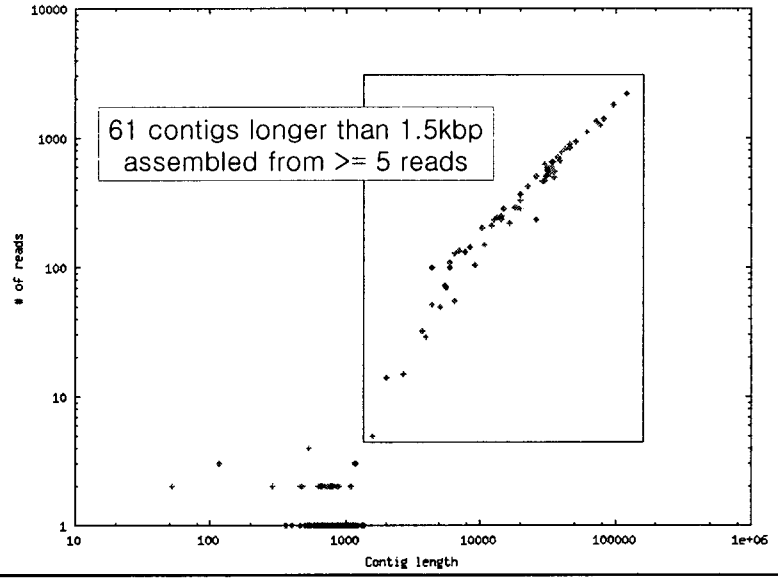
- A total of 29,918 reads excluding 32 failed runs were used in phrap run
- 30 min run on a Pentium-III 800MHz Linux box
- Max. memory consumption of 450MB

prescreen

- 15 min run with 200MB of memory

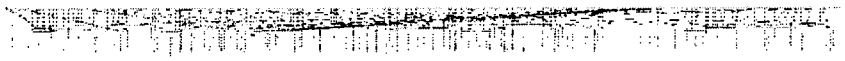


Distribution of Contig size

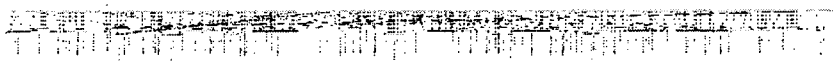


Comparison with known strains

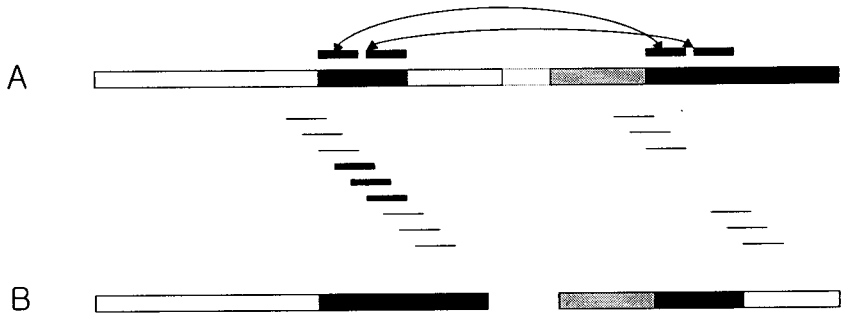
26695 *vs* 51



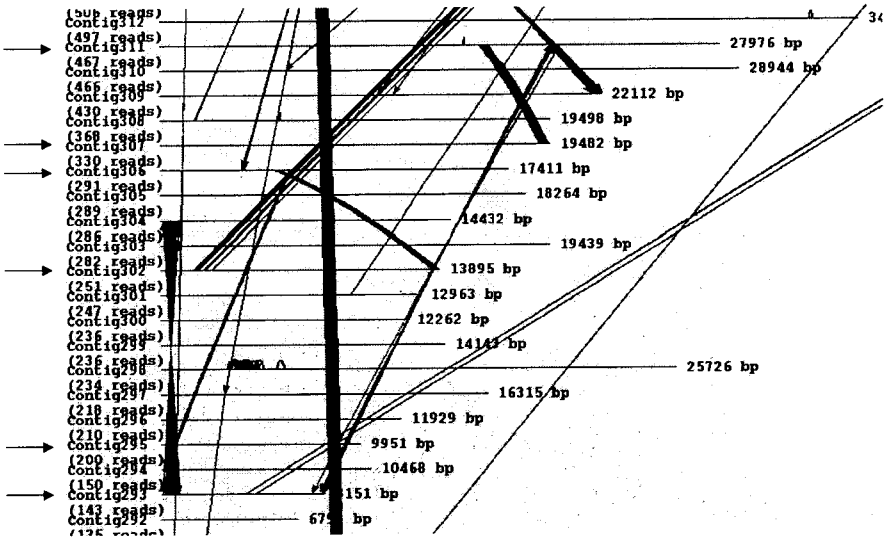
J99 *vs* 51



Repeat Detection



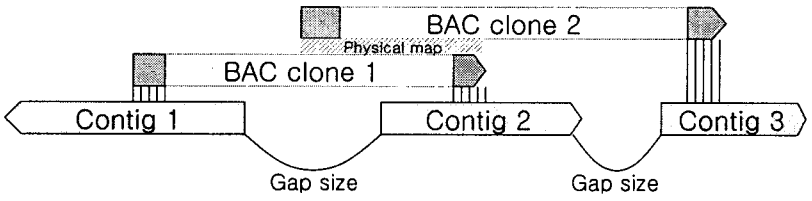
PHRAPVIEW detecting repeats

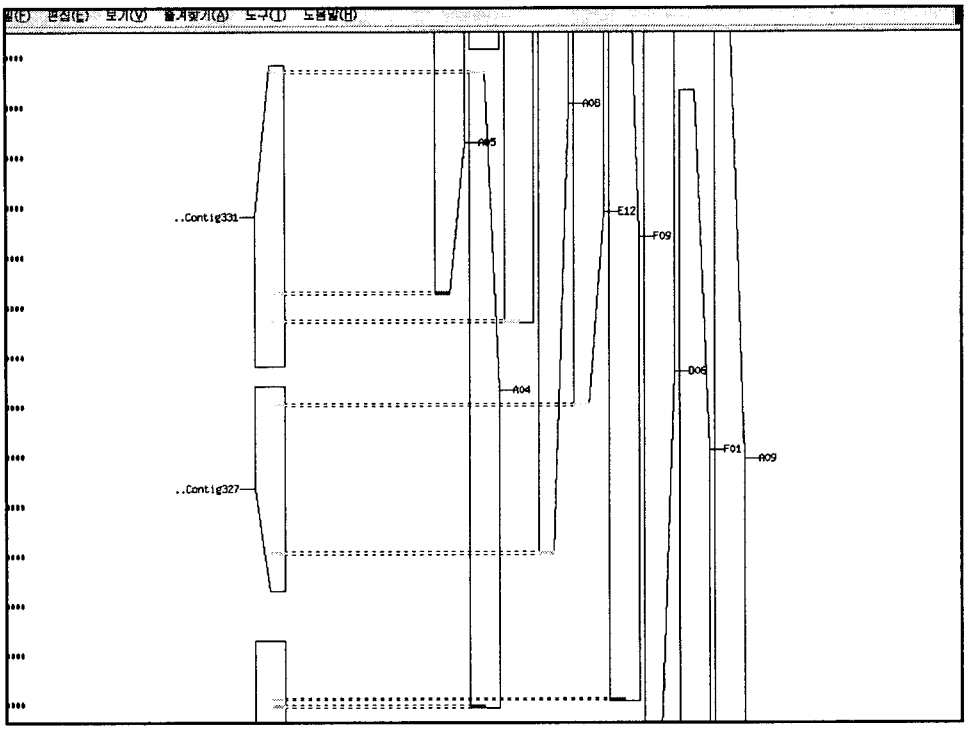


Finishing (H.S.Park *et al*)

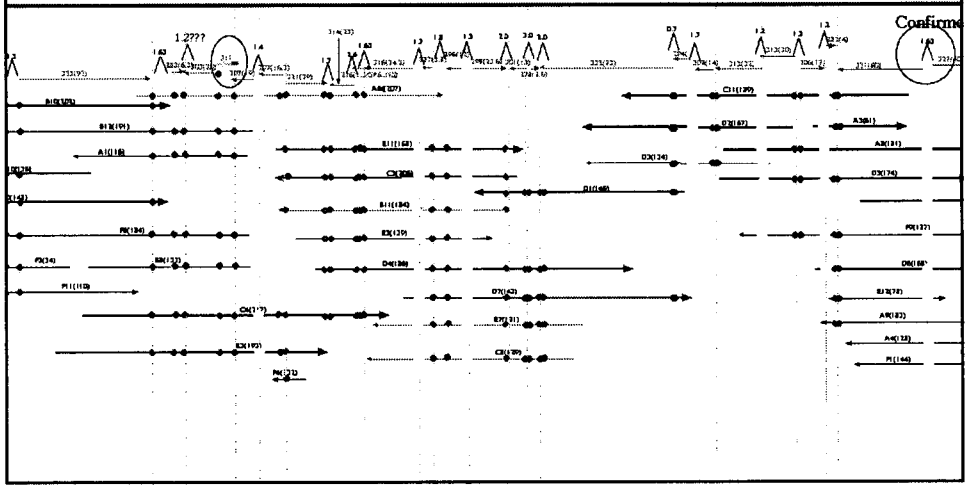
- Objectives
  - Determine the order and orientation of 61 contigs
  - Measure the size of gaps and filling
  - Sequence confirmation of low-depth or unidirectional region
- Methods
  - BAC clone physical map (prepared by GSNU)
  - BAC end-sequencing
  - BAC clones & Contig mapping
  - PCR and primer walking

Mapping Contigs with BAC Clones

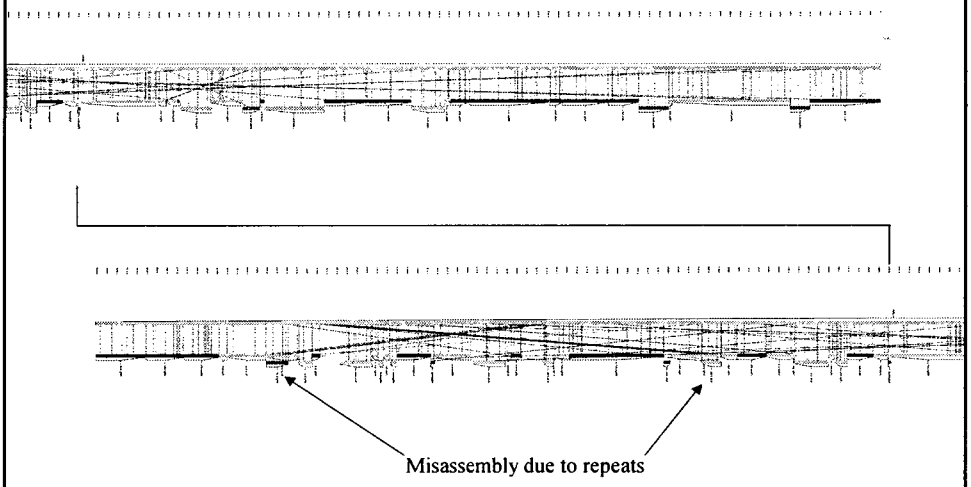




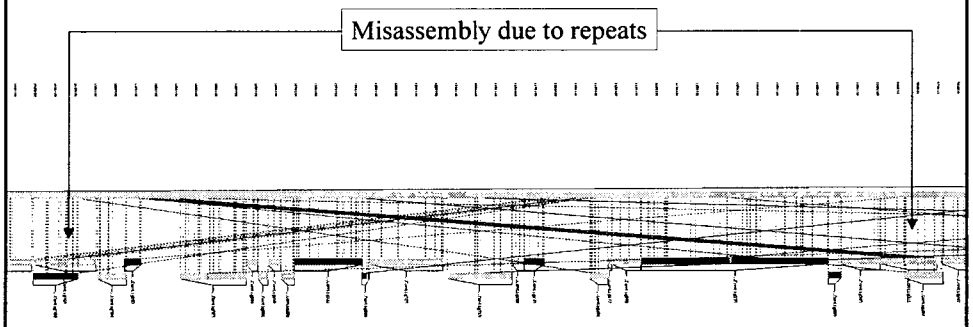
High resolution physical map depicting contig-to-BAC clone mapping confirmed by PCR (H.S.Park *et al*)

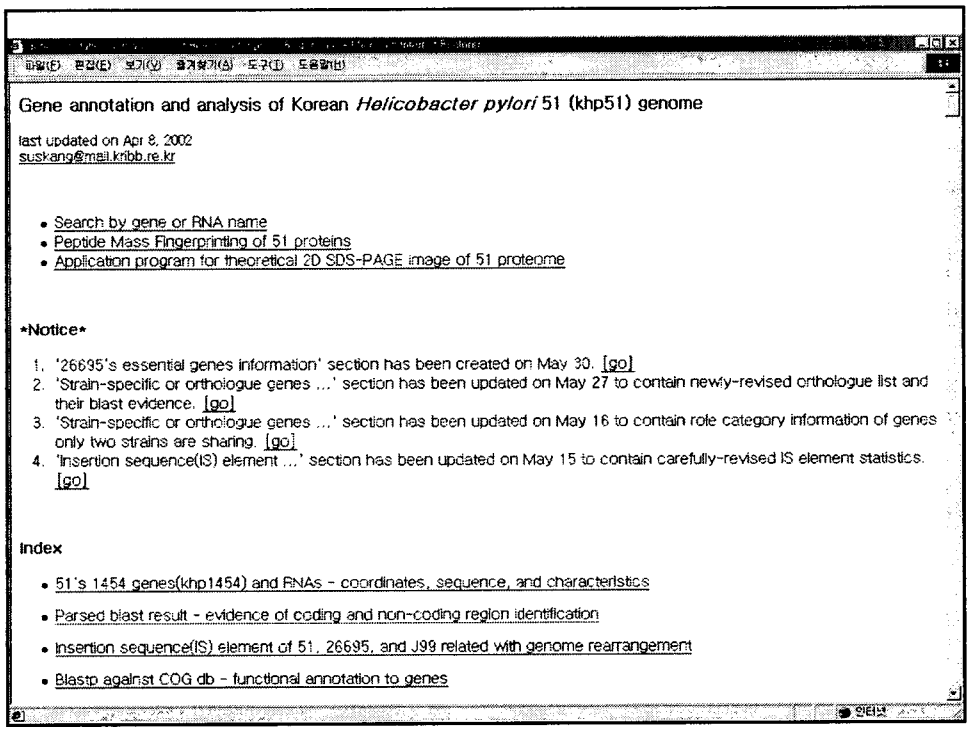


Retrospective comparison with the finished data confirmed the correctness of assembly except for two repeat-containing contigs



Expanded View of Misassembly





Gene annotation and analysis of Korean *Helicobacter pylori* 51 (khp51) genome

last updated on Apr 8, 2002
suskang@mail.kribb.re.kr

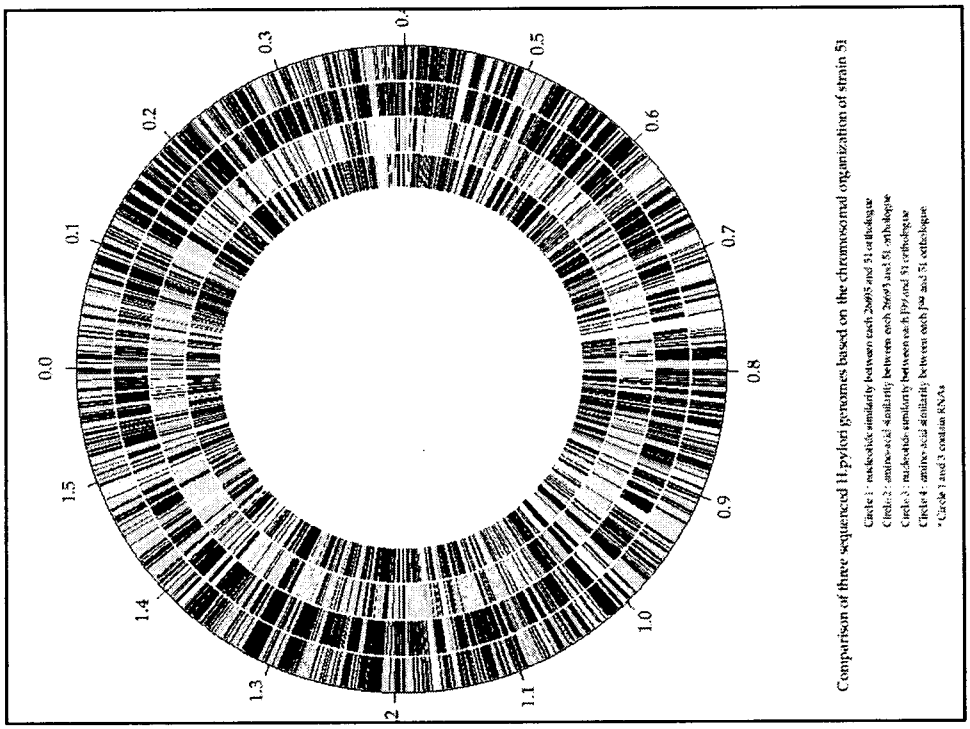
- Search by gene or RNA name
- Peptide Mass Fingerprinting of 51 proteins
- Application program for theoretical 2D SDS-PAGE image of 51 proteome

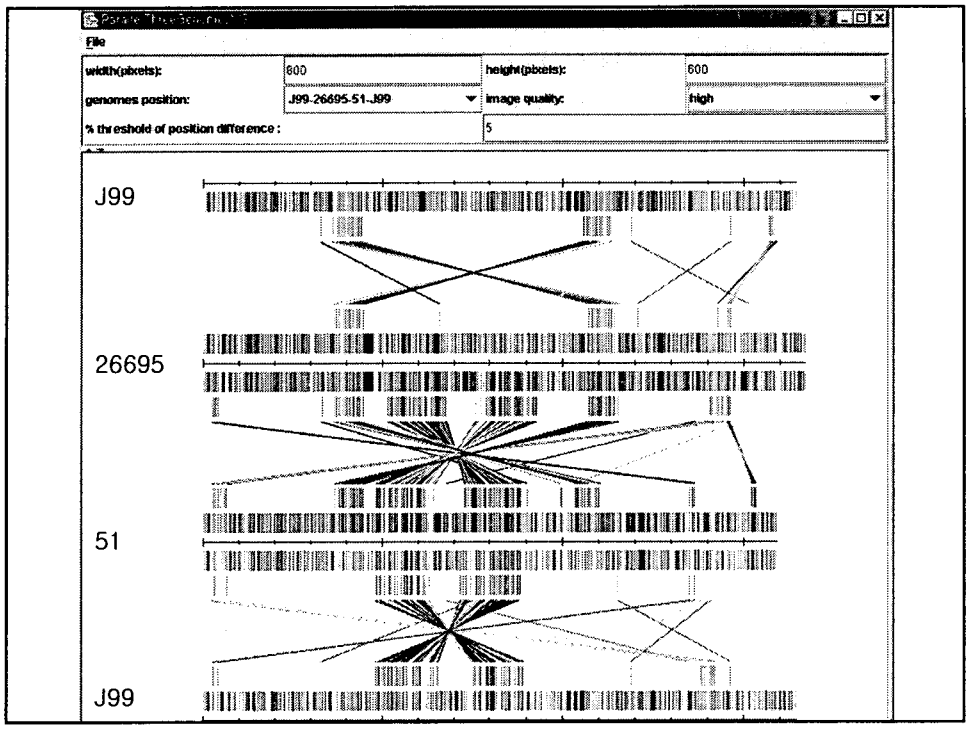
Notice

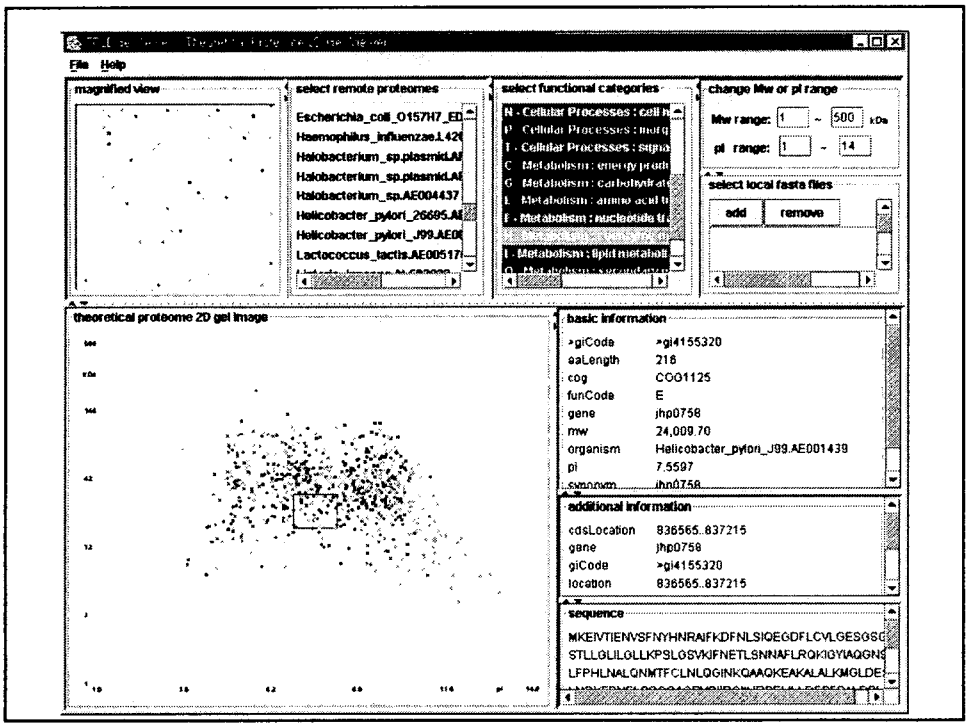
1. '26695's essential genes information' section has been created on May 30. [go]
2. 'Strain-specific or orthologue genes ...' section has been updated on May 27 to contain newly-revised orthologue list and their blast evidence. [go]
3. 'Strain-specific or orthologue genes ...' section has been updated on May 16 to contain role category information of genes only two strains are sharing. [go]
4. 'Insertion sequence(IS) element ...' section has been updated on May 15 to contain carefully-revised IS element statistics. [go]

Index

- 51's 1454 genes(khp1454) and RNAs - coordinates, sequence, and characteristics
- Parsed blast result - evidence of coding and non-coding region identification
- Insertion sequence(IS) element of 51, 26695, and J99 related with genome rearrangement
- Blastp against COG db - functional annotation to genes




File
width(pixels):
600
height(pixels):
600
genomes position:
J99-26695-51-J99
image quality:
high
% threshold of position difference :
5
J99
26695
51
J99


File
Help
magnified view
select remote proteomes
select functional categories
change Mw or pI range
Mw range:
1
500
kDa
pI range:
1
14
select local fasta files
add
remove
theoretical proteome 2D gel image
basic information
>giCode >gi4155320
aaLength 216
cog COG1125
funCode E
gene jhp0758
mw 24,009.70
organism Helicobacter_pylori_J99.AE001439
pI 7.5597
additional information
cdsLocation 836565..837215
gene jhp0758
giCode >gi4155320
location 836565..837215
sequence

Acknowledgement

- Library construction
  - Gyeongsang Nat'l U.
- Sequencing
  - KRIBB, Genotech
- Informatics
  - KRIBB
- Finishing
  - KRIBB, GSNU, Genotech

Kwang Ho Rhee
Myung Je Cho
Woo Kon Lee
Yong Sung Kim
Hong Seog Park
Yoonsoo Hahn
Jae Jong Kim
Haeyoung Jeong