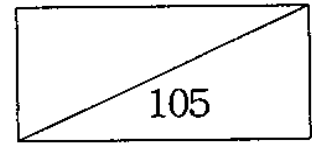


BSNB130M-829-1



T. caldophilus 게놈 연구

Genome Analysis of *T. caldophilus*

1996. 8. 30

한국과학기술연구원
부설생명공학연구소

목 차

제 1 장 서론 -----	7
제 1 절 연구 개발의 목적 및 중요성 -----	7
제 2 절 주요 연구개발 내용 및 범위 -----	7
제 2 장 연구추진방법 -----	9
제 1 절 실험재료 -----	9
제 2 절 실험방법 -----	9
1. <i>T. caldophilus</i> GK24의 배양 -----	9
2. <i>T. caldophilus</i> GK24 Library의 구성 -----	9
가. S4-plasmid library의 구성 -----	9
(1) Intact chromosome 분리 방법 -----	9
(2) Agarose blocks 상에서 DNA의 제한효소처리 -----	9
(3) PFGE를 위한 λ DNA concatemer marker의 준비 -----	10
(4) PFGE -----	10
(5) S4-plasmid library의 구성 -----	10
나. Lambda phage library의 구성 -----	10
(1) <i>T. caldophilus</i> chromosomal DNA 의 분리 -----	10
(2) Chromosomal DNA의 부분적 제한효소처리 -----	11
(3) Size fractionation by sucrose gradient -----	11
(4) 파아지 DNA의 준비 -----	11
(5) Ligation 과 Packaging 반응 -----	11
3. Library를 이용한 유용효소들의 클로닝 -----	12
가. Probe -----	12
나. Screening -----	12
4. 클로닝된 효소들의 염기서열결정 -----	12
제 3 장 결과 및 고찰 -----	13
1. 게놈지도작성 -----	13
2. 유전자군의 클로닝과 염기서열결정 -----	13
3. 특정 생합성 관련 효소유전자군의 선정과 집중연구 -----	13
4. 유전자의 전산분석과 새로운 효소의 탐색 -----	13
제 4 장 결론 및 전의사항 -----	40
제 1 절 결론 -----	40
제 2 절 전의사항 -----	40
제 5 장 참고문헌 -----	42

제 출 문

과학기술처장관 귀하

본 보고서를 “*T. caldophilus* 게놈 연구” 과제의 최종보고서로 제출합니다.

1996년 8월 30일

연구책임자 : 이대실 (분자당생물학실 책임연구원)

연구 원 : 박병철 (분자당생물학실 선임연구원)

고정현 (분자당생물학실 선임연구원)

고석훈 (분자당생물학실 연구원)

김중수 (분자당생물학실 연구원)

나시루 (분자당생물학실 연구원)

박정아 (분자당생물학실 연구원)

정보원 (분자당생물학실 연수연구원)

요 약 문

I. 제목: *T. caldophilus* 게놈 연구

II. 연구개발의 목적 및 중요성

내열성 미생물 게놈의 유전자염기서열을 결정하고 다른 미생물의 유전정보를 기초로한 분석을 통하여 새로운 효소에 대한 유전자 자원을 다량 확보할 수 있다. 이는 생명공학에 결정적인 기여를 할 것으로 본다.

1) 내열성 미생물의 효소는 70 °C 이상의 극한조건에서 안정하게 작용하는 성질을 갖고 있다. 또한 이 내열성 효소의 생화학반응은 산업환경에 쉽게 적용될 수 있고, 특수용도의 연구수단으로 유용하게 이용되고 있다. 예) 내열성 DNA중합효소 (생명과학연구의 필수 실험소재, 유전병, 유전자 검색을 통한 이용, 세계시장, 약 1000억원)

2) 내열성 효소는 당단백질이 아니므로 유전공학적인 생산이 용이할 뿐 아니라, 높은 열안정성으로 그 분리정제가 단 한번의 열처리로 가능하므로써 내열성 효소의 대량생산이 간편하다.

3) 내열성 효소생산체계는 효소공학산업의 소재를 제공하여, 다양한 생체의 생합성체계를 인공적으로 구축할 수 있어서 미래의 생물산업의 기반을 마련하여 줄뿐 아니라, 석유산업대체기술 개발의 가장 효율적인 방안이다.

4) 내열성 미생물의 게놈연구는 그간 생명공학연구소의 독자적인 추진으로 일차 지도가 작성되어 있고, 현재 국내 보유하고있는 자체 유전공학기술로도 단시일 내에 완료할 수 있다.

실행가능성: 본연구팀에서는 현재 위에서 언급한 효소외에도 40여가지의 효소와 S4 plasmid library와 λ phage library를 가지고 genome 연구를 계속 수행중에 있고 Tca, Taq, Top Polymerase와 Xylose Isomerase, 그리고 UDP-glucose pyrophosphorylase 등의 연구를 성공리에 수행한 경험이 있다.

III 연구개발의 내용 및 범위

1) 게놈지도작성: *T. caldophilus* GK24의 제한효소지도 작성, 유전적 지도 작성

2) 특정 생합성 관련 효소유전자군의 선정과 집중연구 :

- 탄수화물 관련 효소(Xylose isomerase, Pullulanase, Trehalose synthase, 그리고 UDP-glucose transferase) 의 유전자군 게놈 확인
- DNA결합효소 (DNA polymerase, Ligase)의 유전자군 게놈 확인

3) 유전자군의 클로닝과 염기서열결정

- 특정 효소 유전자군과 Ssp4 게놈 절편(314 Kb)을 중심으로 유전자 군을 탐색하고 이들을 각각 대장균에 크로닝하고 염기서열을 결정

4) 유전자의 전산분석과 새로운 효소의 탐색

- *T. caldophilus* GK24의 유전자는 G+C 함량이 70% 가까이 되어 단순한 염기배열 비교를 통해서 유전자의 특성 추정이 어려우므로 이러한 유전자분석을 위한 분석체계를 구축하고 이렇게 분석된 유전자군을 이용하여 새로운 효소를 계속해서 확보함

IV 연구개발 및 활용에 대한 진의

내열성 탄수화물 효소류의 경우에는 개개의 효소나 효소군이 실제 산업에서 한가지 공정이나 제품생산을 바꿀수 있기 때문에 세계적으로 수없이 많은 연구가 진행되고 있다. 따라서 연구과정에서 얻어진 Glucose Isomerase, Pullulanase, ADP-glucose pyrophosphorylase, GDP-mannose pyrophosphorylase 등의 효소나 효소군은 내열성이 최대의 장점이 되는 당관련 산업에 바로 적용시켜 사용할 수 있다. 또한 연구과정에서 확립한 게놈분석 및 응용기술은 분석된 유전자의 확인뿐 만 아니라 유관 게놈연구의 모델로도 매우 유용하리라고 생각된다.

- 1) 유용탄수화물의 제조기술 확보
- 2) 산업적으로 유용한 탄수화물 관련효소의 확보
- 3) 복잡한 탄수화물의 구조분석방법 개발
- 4) 유전공학기술을 이용한 산업 효소의 대량생산 체계구축
- 5) 내열성 효소의 신용도 개발연구 착수가능
- 6) Genomic library 로부터 유용단백질 및 효소 확보 가능
- 7) 게놈연구의 기술개발 기능 확보
- 8) 인력 확충 및 기기 보완이 필수적임
- 9) 국내 연구진의 상호협력 체제 구축 요망

Summary

Extremely thermophilic bacterium, *Thermus caldophilus* GK24 is a strain belonging to genus *Thermus*. Because enzymes from thermophilic organisms are designed to "take the heat", many researchers believe they may one day be useful as industrial catalysts for the production of fuels, pharmaceuticals, foods industry, and specialty chemicals, among other applications. Analysis of the genome obtained with several restriction endonucleases revealed that the *T. caldophilus* GK24 strain has a circular chromosome of about 2.18 Mb. The localization of AGPase and eight genes from genomic library constructed with *Sau3AI* digestion have been determined. Those DNA sequences are going to utilized to localize there position and determine physical map of *Thermus caldophilus* GK24. The S4 DNA fragment was purified through PFGE and it was digested with *Sau3AI* and cloned with pUC19. More than 100 clones were sequenced and analysed by BLAST and FASTA DNA analysis program with Gene Bank or EMBL databank. And also we cloned several genes include catechol dioxygenase and thioredoxin by using OS-PCR method. OS-PCR method was established for the cloning of unknown targeted genes that were related to carbohydrates enzyme. The genes encoding xylose isomerases (*xyIA*) were cloned from *Thermus caldophilus* GK24 and *Thermus flavus* AT62, and the DNA sequences were determined. They were very similar in sequence and structure each other. The genes *xyIA* encode the enzyme xylose isomerase (XI) consisting of 387 amino acids each. There was a partial xylulose kinase gene overlapping 4 bp at the end of xylose isomerase gene. The whole genome of *Thermus caldophilus* GK24 was also packaged in Lambda EMBL3 and transfection in *E.coli*. This phage library was screened with several probes like *xyIA*, *trs*, *glgC*, and *puICD* genes etc. They were identified and subcloned in *E.coli*. These clones were sequenced each other and mapped with sequences and restriction mapping. More than 500 clones and 60 Kb DNA were sequenced and more than 50 new genes were identified. They had highly homologous sequence as DNA or protein with reported genes. During these analysis for identifying the clones, we developed data bank system and analysis methods for accurate and easy handling of unknown gene sequences. All these methods, programs and also new genes are very useful to further approach of in these field and related industry.

제 1 장 서 론

제 1 절 연구 개발의 목적 및 중요성

내열성 미생물 게놈의 유전자염기서열을 결정하고 다른 미생물의 유전정보를 기초로한 분석을 통하여 새로운 효소에 대한 유전자 자원을 다량 확보할 수 있다. 이는 생명공학에 결정적인 기여를 할 것으로 본다.

- 1) 내열성 미생물의 효소는 70 °C 이상의 극한조건에서 안정하게 작용하는 성질을 갖고 있다. 또한 이 내열성 효소의 생화학반응은 산업환경에 쉽게 적용될 수 있고, 특수용도의 연구수단으로 유용하게 이용되고 있다. 예) 내열성 DNA중합효소 (생명과학연구의 필수 실험소재, 유전병, 유전자 검색을 통한 이용, 세계시장, 약1000억원)
- 2) 내열성 효소는 당단백질이 아니므로 유전공학적인 생산이 용이할 뿐 아니라, 높은 열안정성으로 그 분리정제가 단 한번의 열처리로 가능하므로서 내열성 효소의 대량생산이 간편하다.
- 3) 내열성 효소생산체계는 효소공학산업의 소재를 제공하여, 다양한 생체의 생합성체계를 인공적으로 구축할 수 있어서 미래의 생물산업의 기반을 마련하여 줄뿐 아니라, 석유산업대체기술 개발의 가장 효율적인 방안이다.
- 4) 내열성 미생물의 게놈연구는 그간 생명공학연구소의 독자적인 추진으로 일차 지도가 작성되어 있고, 현재 국내 보유하고있는 자체 유전공학기술로도 단시일 내에 완료할 수 있다.

실행가능성: 본연구팀에서는 현재 위에서 언급한 효소외에도 40여가지의 효소와 S4 plasmid library와 λ phage library를 가지고 genome 연구를 계속 수행중에 있고 Tca, Taq, Top Polymerase와 Xylose Isomerase, 그리고 UDP-glucose pyrophosphorylase 등의 연구를 성공리에 수행한 경험이 있다.

제 2 절 주요 연구개발 내용 및 범위

- 1) 게놈지도작성: *T. caldophilus* GK24의 제한효소지도 작성, 유전적 지도 작성

2) 특정 생합성 관련 효소유전자군의 선정과 집중연구 :

- 탄수화물 관련 효소(Xylose isomerase, Pullulanase, Trehalose synthase, 그리고 UDP-glucose transferase) 의 유전자군 계통 확인
- DNA결합효소 (DNA polymerase, Ligase)의 유전자군 계통 확인

3) 유전자군의 클로닝과 염기서열결정

- 특정 효소 유전자군과 Ssp4 계통 절편(314 Kb)을 중심으로 유전자 군을 탐색하고 이들을 각각 대장균에 크로닝하고 염기서열을 결정

4) 유전자의 전산분석과 새로운 효소의 탐색

- *T. caldophilus* GK24의 유전자는 G+C 함량이 70% 가까이 되어 단순한 염기배열 비교를 통해서 유전자의 특성 추정이 어려우므로 이러한 유전자분석을 위한 분석체계를 구축하고 이렇게 분석된 유전자군을 이용하여 새로운 효소를 계속해서 확보함

제 2 장 연구추진방법

제 1 절 실험재료

Thermus caldophilus GK24를 계놈연구를 위한 library의 제조에 이용하였다.

Escherichia coli XL1-blue MRA ($\Delta(mcrA)183$, $\Delta(mcrCB-hsdSMR-mrr)173$, *endA1*, *supE44*, *thi-1*, *gyrA96*, *relA1*, *lac*) 는 Lambda library의 host strain으로 이용하였다.

제 2 절 실험방법

1. *T. caldophilus* GK24의 배양

Thermus caldophilus GK24는 0.8 % polypeptone, 0.2 % yeast extract와 basal salt에 1 % Glucose를 탄소원으로 하여 pH를 7.2로 고정시키고 75 °C 에서 배양하였다. 이 배양액에 최종농도가 180 $\mu\text{g/ml}$ 가 되도록 chloramphenicol을 넣어 4시간 더 배양하였다.

2. *T. caldophilus* GK24 Library의 구성

가. S4-plasmid library의 구성

(1) Intact chromosome 분리 방법

배양액을 얼음에서 차게한 뒤 harvest하여 현탁용액(10 mM Tris-HCl, pH 7.6, 1 M NaCl)을 넣어 45 °C로 맞춘 뒤 동일 부피의 1 % low melting temperature agarose와 섞은 후 주형에 넣어 4 °C에서 20분간 방치하여 agarose blocks를 만들어서 동일 부피의 TC lysis solution (6 mM Tris-HCl, pH 7.6, 1 M NaCl, 100 mM EDTA, 0.5 % Sarkosyl, 1 mg/ml lysozyme, 20 $\mu\text{g/ml}$ RNase A)에 넣어 37 °C에서 30분 동안 방치하였다. TC용액을 따라낸 뒤 동일 부피의 ESP buffer (0.5 M EDTA, 1 % laural sarcosine, 1 mg/ml proteinase K)를 넣어 50 °C에서 하루동안 방치한 뒤 0.05 M EDTA(pH 8.0)로 3번 씻어 주었다.

(2) Agarose blocks 상에서 DNA의 제한효소처리

DNA-block을 수세한 후 열배 부피의 제한 효소 완충액에 넣어 4 °C에

서 1시간 방치한 뒤 동일한 완충액으로 갈아주고 30-50 units의 제한효소를 넣어 적정온도에서 12-16시간동안 반응시켰다. 이것을 50배의 TE로 씻어준 뒤 사용하였다.

(3) PFGE를 위한 λ DNA concatemer marker의 준비

10 μ g의 bacteriophage λ DNA를 TE에 녹여 37 $^{\circ}$ C로 유지시킨 뒤, 동일 온도로 유지시킨 100 μ g의 LMT agarose 용액(20 mM MgCl₂, 100 mM Tris-HCl, pH 7.6)과 섞은 후 1 % PEG 8000, 1mM ATP, 1 mM DTT, bacteriophage T4 DNA ligase, 0.75 % agarose 용액과 재빨리 섞어서 15 분 동안 얼음에서 방치하였다. 이 block을 PEG를 포함하는 ligation 용액에 넣어 상온에서 24시간동안 반응시킨 후 20 mM EDTA (pH 8.0) 용액으로 치환시킨 뒤 사용하였다.

(4) PFGE

Pulse field gel electrophoresis(PFGE)는 Bio-Rad system을 이용하였다. 전기영동 조건은 다음과 같이 하였다: 0.5 X TBE 완충용액으로 만든 1 % Sea-Kem agarose에 DNA block을 넣어 10 V/cm의 field strength 하에서 24시간 동안 전기영동하였다. 온도는 14 $^{\circ}$ C로 유지하여 수행하였다. 6-600 Kb 크기의 DNA 절편의 분리를 위해서 15 sec 에서 75 sec 로의 switching intervals의 조건을 주었다. 전기영동 후 gel을 EtBr용액에서 30분 동안 염색한후 사진을 찍었다.

(5) S4-plasmid library의 구성

T. caldophilus GK24 genomic DNA의 Ssp I 제한효소 후 PFGE에서 나온 네번째 단편(Ssp 4)을 분리한 뒤, 이 Ssp4-DNA 를 *Sau*3A I 제한효소로 digestion 하였다. 이 digest 중에서 0.5 Kb 이하의 것을 제거한 후 pUC18/19 의 *Bam*HI site 에 cloning 하였다.

나. Lambda phage library의 구성

(1) *T. caldophilus* chromosomal DNA 의 분리

앞에서 언급한 배지에 이 균주를 배양(10 ml, 24 시간) 하였다. 원심분리하여 세포를 회수한 후 4 ml의 SET (20 % sucrose, 50 mM Tris-Cl, pH 7.6, 50 mM EDTA) 완충용액에 현탁하였다. 세포를 파쇄하고 RNA를 제거하기 위하

여 40 mg의 lysozyme과 2 mg의 RNase A를 가한 후 37 °C에서 30분간 반응시켰다. 단백질을 변성시키기 위하여 최종농도가 1 %가 되는 SDS와 2 mg의 proteinase K를 가하고 동일한 방법으로 37 °C에서 30분간 반응시켰다. 이어 동량의 phenol/chloroform을 가하여 밤새(24시간) 천천히 흔들어 준 다음 원심분리하고 상층액에 대하여 최종농도 0.3 M의 sodium acetate(pH 5.2)와 2배 부피의 에탄올을 가하여 DNA를 석출시켰다. 이 DNA를 70 % 에탄올로 씻은 다음 말린 후 TE(10 mM Tris-Cl, pH 8.0, 1 mM EDTA) 완충용액에 녹여 사용하였다.

(2) Chromosomal DNA의 부분적 제한효소처리

300 μ g 정도의 chromosomal DNA를 제한효소 완충용액과 잘 섞은 후 10개의 tube에 똑같이 나누는 뒤 첫 번째 tube에 10 units의 제한효소(Sau3A I나 BamH I)를 넣고 혼합하여 절반의 부피를 두 번째 tube의 것과 혼합하는 식으로 효소의 연속적 희석한 뒤 37 °C에서 1시간 반응시킨 뒤 0.5 % agarose gel에서 전기영동하여 최적의 DNA의 부분적 절단 상태의 것을 선택하였다.

(3) Size fractionation by sucrose gradient

부분적 제한효소 처리된 chromosomal DNA에서 9-23 Kb의 크기를 가진 DNA 절편들을 얻기 위해 sucrose gradients에 의한 방법을 이용하였다. 10 %와 40 %의 sucrose 용액(1 M NaCl, 20 mM Tris-Cl, pH 7.5, 5 mM EDTA)을 이용하여 continous gradient를 SW28 tube에 만든 다음 윗부분에 부분적 제한효소처리된 DNA용액을 loading하여 25,000 rpm에서 16-24시간동안 원심분리하였다. 20 °C에서 행해졌다. Tube의 밑바닥에 capillary tube를 넣어 750 μ l씩 분획하여 0.9 % agarose gel에서 9-23 Kb의 범위에 있는 분획을 모아 TE 용액에서 투석하여 이용하였다.

(4) 파아지 DNA의 준비

Lambda EMBL3 vector DNA를 제한효소 BamH I으로 절단하여 left arm (20 Kb) 와 right arm (9 Kb) DNA를 준비하였다.

(5) Ligation 과 Packaging 반응

9-23 Kb의 크기를 가진 insert DNA와 Lambda EMBL3 vector의 left arm (20 Kb) 와 right arm (9 Kb) DNA의 최적 ligation yield를 보이는 molar

ratio를 구하여 16 °C에서 4-6 시간 동안 행해졌다.

사용한 packaging extract는 Gigapack II Gold (Stratagene) 로서 Stratagene사의 manual 에 따라 행해졌다. Packaging 후에 *Escherichia coli* XL1-blue MRA에 infection시켜 titration 하였다.

3. Library를 이용한 유용효소들의 클로닝

가. Probe

- (1) 기존의 보고된 유전자들의 염기서열로부터 PCR primer를 구성하였다.
- (2) 단백질의 아미노산서열로부터 degenerated primer 를 구성하였다.

나. Screening

Dig-labeled probes를 이용하여 2차에 걸쳐 plaque 혹은

4. 클로닝된 효소들의 염기서열결정

DNA 염기서열을 결정하기 위하여 Sanger의 dideoxynucleotide chain termination sequencing method를 변형한 방법을 이용하였다. Super-coil plasmid DNA 1-2 μ g을 NaOH로 변형시키고 ammonium acetate (pH 4.5)를 넣고 annealing하였다. 여기에 Klenow fragment와 [α - 32 P] dATP 를 넣고 섞은 후 G, A, T, C 네가지 반응을 위해 네 개의 microtube에 똑같이 나누었다. 여기에 각각 적당한 농도의 dATP/ddATP, dCTP/ddCTP, dGTP/ddGTP, dTTP/ddTTP 혼합물을 각각 넣고 반응시킨 후 dNTP를 적당량 넣어 chase 반응시킨 다음, 정지 용액을 넣어 반응을 중지시켰다. 이 반응액을 3분간 끓인 후 6 % polyacrylamide-8 M urea 겔에 loading하여 1,500-1,700 volts로 여러 시간 동안 전기 영동을 하였다. 겔을 떼어 내어 비닐로 싸서 X-ray 필름에 8시간 동안 노출 후 현상하였다.

제 3 장 결과 및 고찰

T. caldophilus GK24의 전게놈 연구가 여러 각도에서 행해졌고 또한 계속 수행중이다(그림 1). 지금까지 게놈연구를 통해 밝혀진 유전자 수는 unknown genes를 포함하여 약 350 여종으로서 <표 1>에서 그 종류들이 보여졌다.

1. 게놈지도작성

T. caldophilus GK24 (그림2)의 제한효소지도 작성이 제한효소 EcoR I, Ssp I, Spe I, Hpa I 등을 이용하여 행해졌고 일부 클로닝된 유전자들의 유전적 지도가 만들어졌다(그림 3).

2. 유전자군의 클로닝과 염기서열결정

특정 효소 유전자군과 Ssp4 게놈 절편(314 Kb)(그림 4)을 중심으로 유전자 군을 탐색하고 이들을 각각 대장균에 클로닝하고 염기서열을 결정하였다. 또한 λ phage library에 의해서도 유전자군을 클로닝하였다.

3. 특정 생합성 관련 효소유전자군의 선정과 집중연구

가. 탄수화물 관련 효소들인 Xylose isomerases (그림 5), Pullulanase, Trehalose synthase (그림 6), 그리고 UDP-glucose transferase의 유전자군 게놈이 확인되었다.

나. DNA결합효소들인 DNA polymerase와 Ligase의 유전자군 게놈이 확인되었다.

다. 16S/5S rRNA-tRNA^{Leu, Val, Glu etc}의 유전자군 게놈이 확인되었다 (그림 7과 8).

라. 이들 유전자군들의 linkage map들이 그림 9에서 보여졌다.

4. 유전자의 전산분석과 새로운 효소의 탐색

T. caldophilus GK24의 유전자는 G+C 함량이 70% 가까이 되어 단순한 염기배열 비교를 통해서도 유전자의 특성 추정이 어려우므로 이러한 유전자전산분석을 위한 분석체계를 구축하고 이렇게 분석된 유전자군을 이용하여 새로운 효소를 계속해서 확보하고자 하였다 (그림 10).

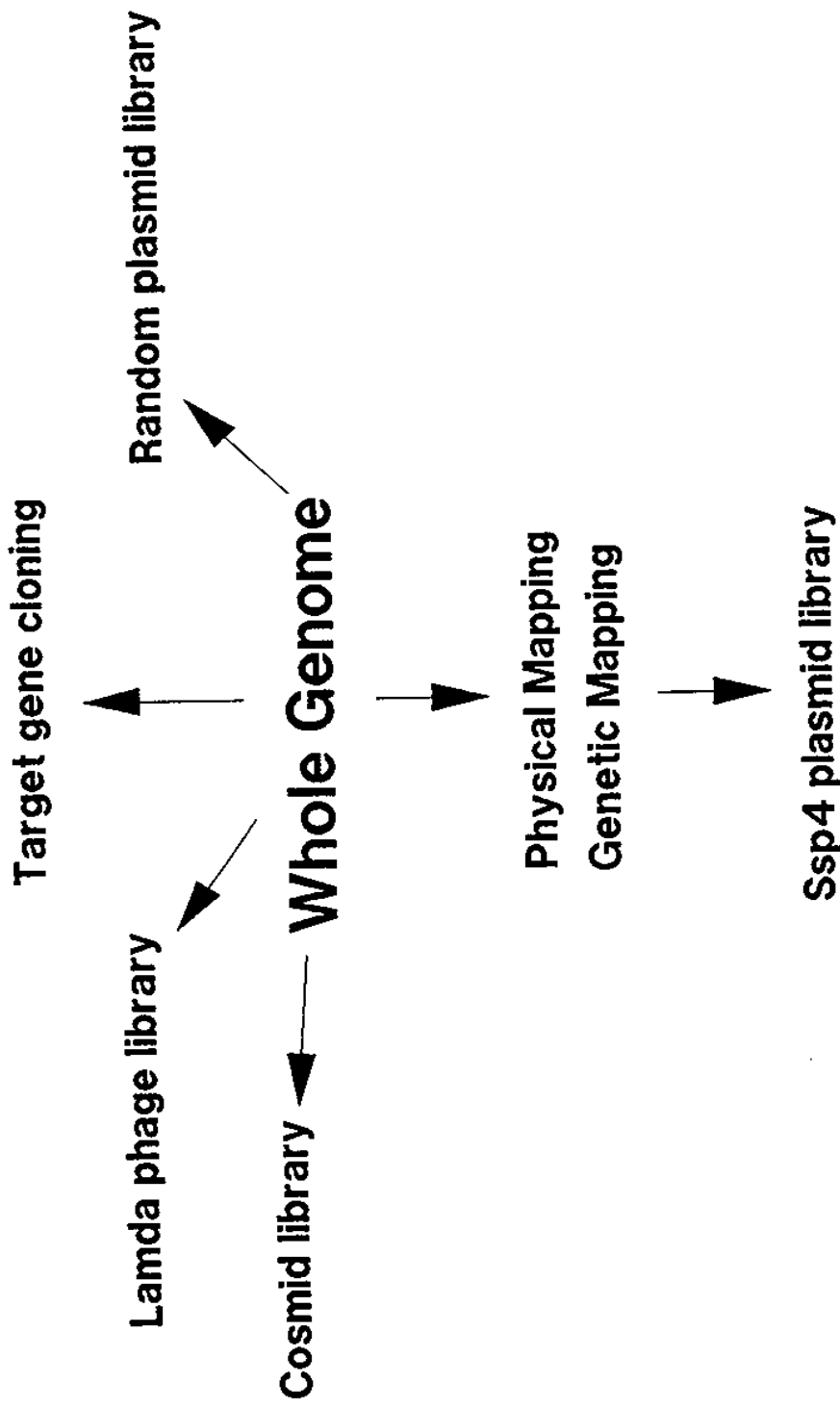


그림 1. 전게놈 연구의 전반적인 개요

Table 1. Genes from *T. caldophilus* GK24

Name	Gene	symbol	Gene length (Kbp)	Homology with those of <i>E. coli</i> (%)
AGPase		<i>glgC</i>	2.7	10
Catechol dioxygenase		<i>bphC</i>	1.0	80(<i>Pseudomonas</i>)
DNA ligase		<i>lig</i>	2.0	50
Galactokinase		<i>galK</i>	0.5	
GDP-Mannose pyrophosphorylase		<i>gpt</i>	1.0	65
Leu-tRNA ligase		<i>leuS</i>		
Pullulanase		<i>pul</i>	2.5	
RNA polymerase		<i>rpoB</i>	0.5	
23S rRNA		<i>rrn</i>	1.2	
16S rRNA/tRNA		<i>rrn/trn</i>	2.5	
phosphoribosyl-AMP cyclohydrolase		<i>hisI</i>		
phosphoribosyl-ATP pyrophosphorylase		<i>hisE</i>		
Tca polymerase		<i>polA</i>	2.5	95
Thioredoxin		<i>trxA</i>	0.5	70
Xylose isomerase		<i>xyIA</i>	1.2	80
Xylulose Kinase		<i>xyIB</i>	1.5	76
xyl operon repressor		<i>xyIR</i>	1.3	82
unknown genes				
			(90 ea)*	40
			(120 ea)**	50
			(100 ea)***	40
Total			160/2200	

* Obtained from random plasmid library of total genomic DNA

** Obtained from random plasmid library of Ssp4 genomic DNA

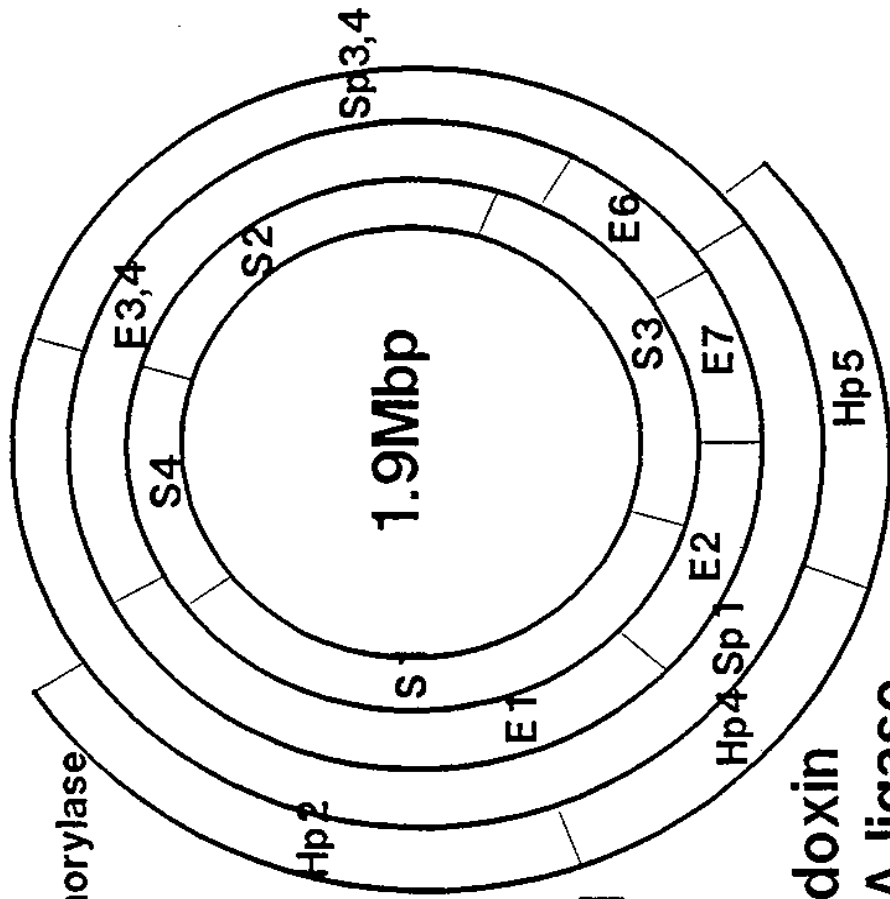
*** Obtained Lambda phage library of total genomic DNA



그림 2. *T. caldophilus* GK24의 전자 현미경 사진

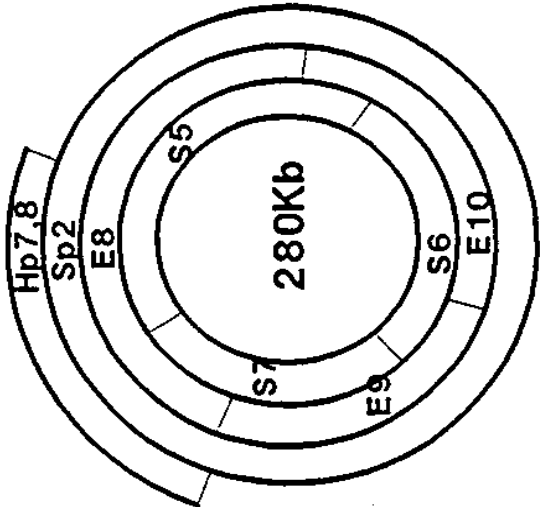
Xylose isomerase

ADP-glucose
pyrophosphorylase



Tca pol

Thioredoxin
DNA ligase



GDP-mannose pyrophosphorylase
Catechol dioxygenase

tRNA-16S rRNA

23S rRNA

Ribosomal protein L5

그림 3. *T. caldophilus* GK24의 Physical map과 genetic map. 한 개의 chromosomal DNA와 한 개의 plasmid로 구성되어 있다.

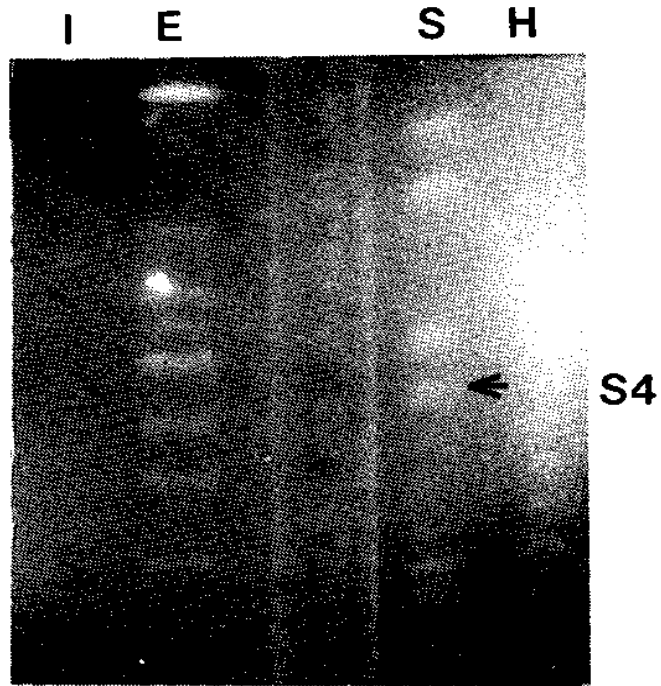


그림 4. *T. caldophilus* GK24 genomic DNA의 제한효소 처리후의 pulse field gel상에서의 pattern. Lane 설명; I, intact form; E, EcoR I; S, Ssp I; H, Hpa I. Pulse time은 15-75 sec이고 14 °C, 200 volts에서 21시간 running 하였다. 화살표는 Ssp I 처리한 경우에서 네번째의 크기를 가진 절편으로 S4 plasmid library의 제조에 이용되었다.

OVERLAPPING GENE SEQUENCE OF XYL OPERON

1170

TCGAGACCAAGCGGCCCGGGGTTATGCCCTGGAACGCCTGGACCAGCTGGCGGTGGAGTACCTCCTGGGGGTCCGGGGGTGAGGGCGG
 L E T K R R R G Y A L E R L D Q L A V E Y L L G V R G U
 XYLOSE ISOMERASE C-TERMINAL M R A
 XYLULOSE KINASE

1260

CCATCGGCTTGGACCTGGGAACGAGCGGGCTCAAGGCCCTGGTGCTGGACGAGGAGGGTAGAAAGCGCGCTGAGGCCCGGGCCGGTTACC
 A I G L D L G T S G L K A L V L D E E G R K R A E A R A G Y

putative rho-independent termination 1350

CCCTTCACACCCCGAGGCCGGGCTGGACGGAGCAGGACCCCGAGACTGGGCCCGGGCCCTGAAGGAGGTGTTCCGGGCCCTGGCGCGGA
 P L H T P R P G W T E Q D P Q D W A R A L K E V F R A L A P

그림 5. *Xyl* operon의 중복되는 유전자의 염기서열

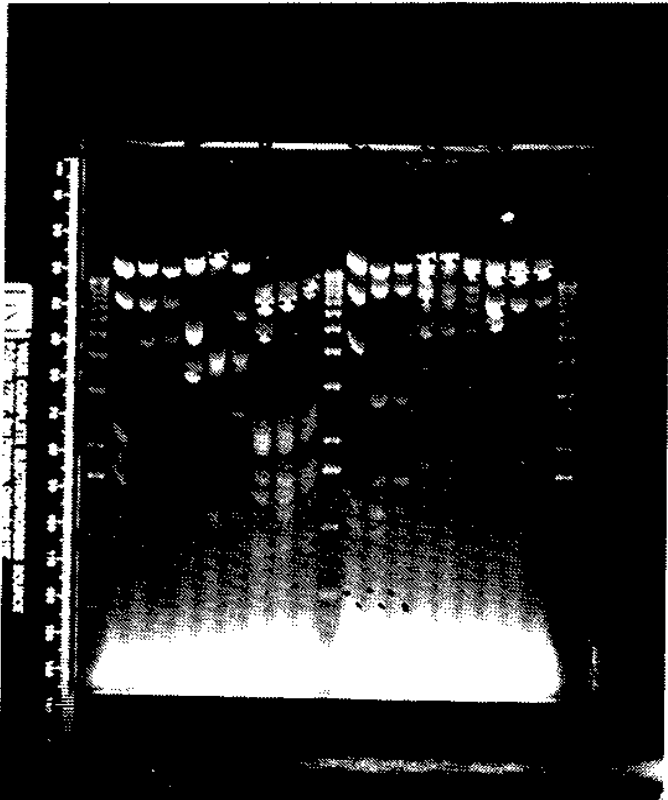
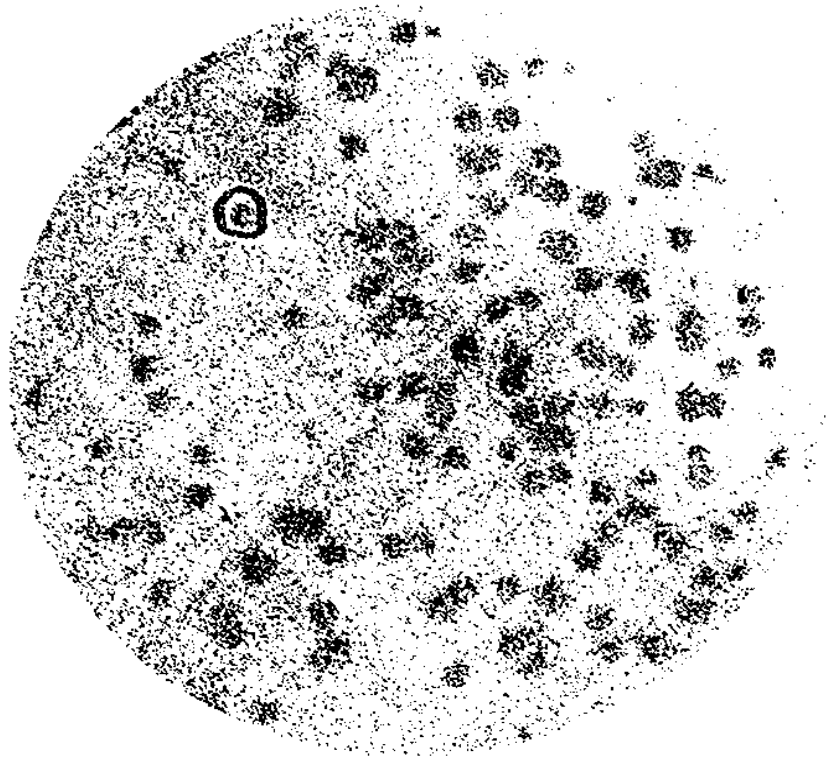
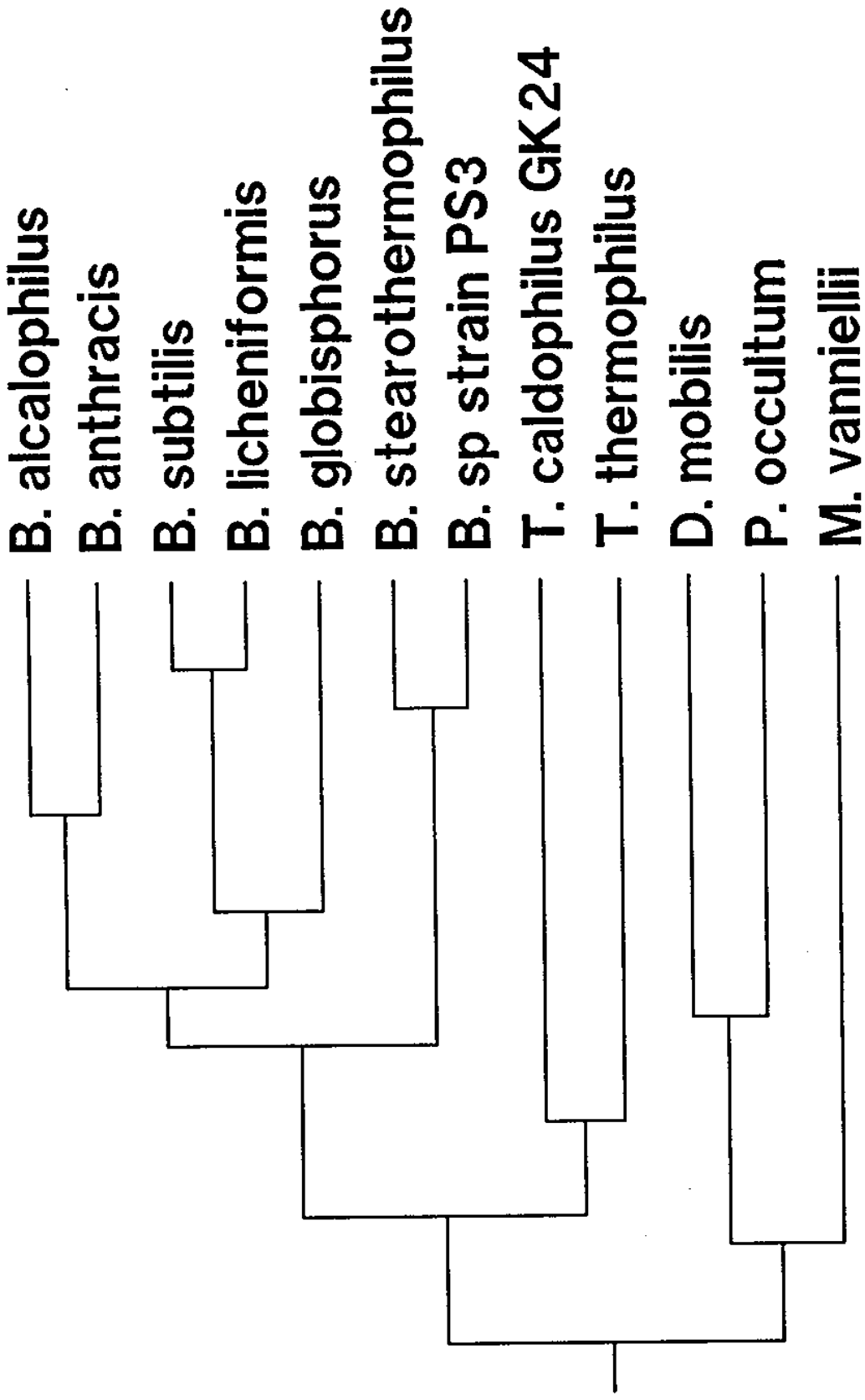
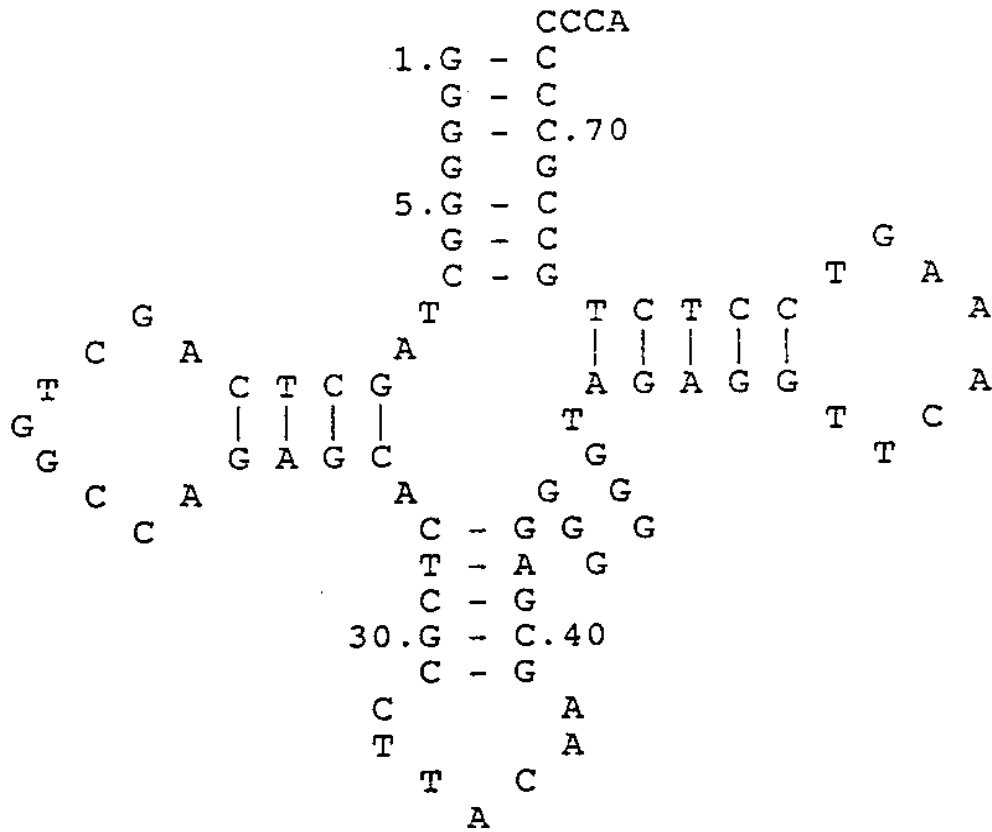


그림 6. Trehalose 유전자의 클로닝. 위: λ library에서의 screening에서 hybridization 결과 보여진 positive plaques.

아래: Southern hybridization 결과에서 보여진 positive clones의 pattern.

UPGMA with 23S rRNA Sequences





**Predicted Val-tRNA
Structure/
Cloned from *T.
caldophilus* GK24**

(Sequence analysis was carried with regular methods of PCGENE program)

그림 8. 클로닝된 Val-tRNA 유전자의 예상 2차구조. 수식된 염기들은 표시되지 않았다. Stem 부분에서의 Watson-Crick 염기쌍은 single dash로 표시되었고 dot는 canonical 염기쌍을 나타낸다.

Linkage map

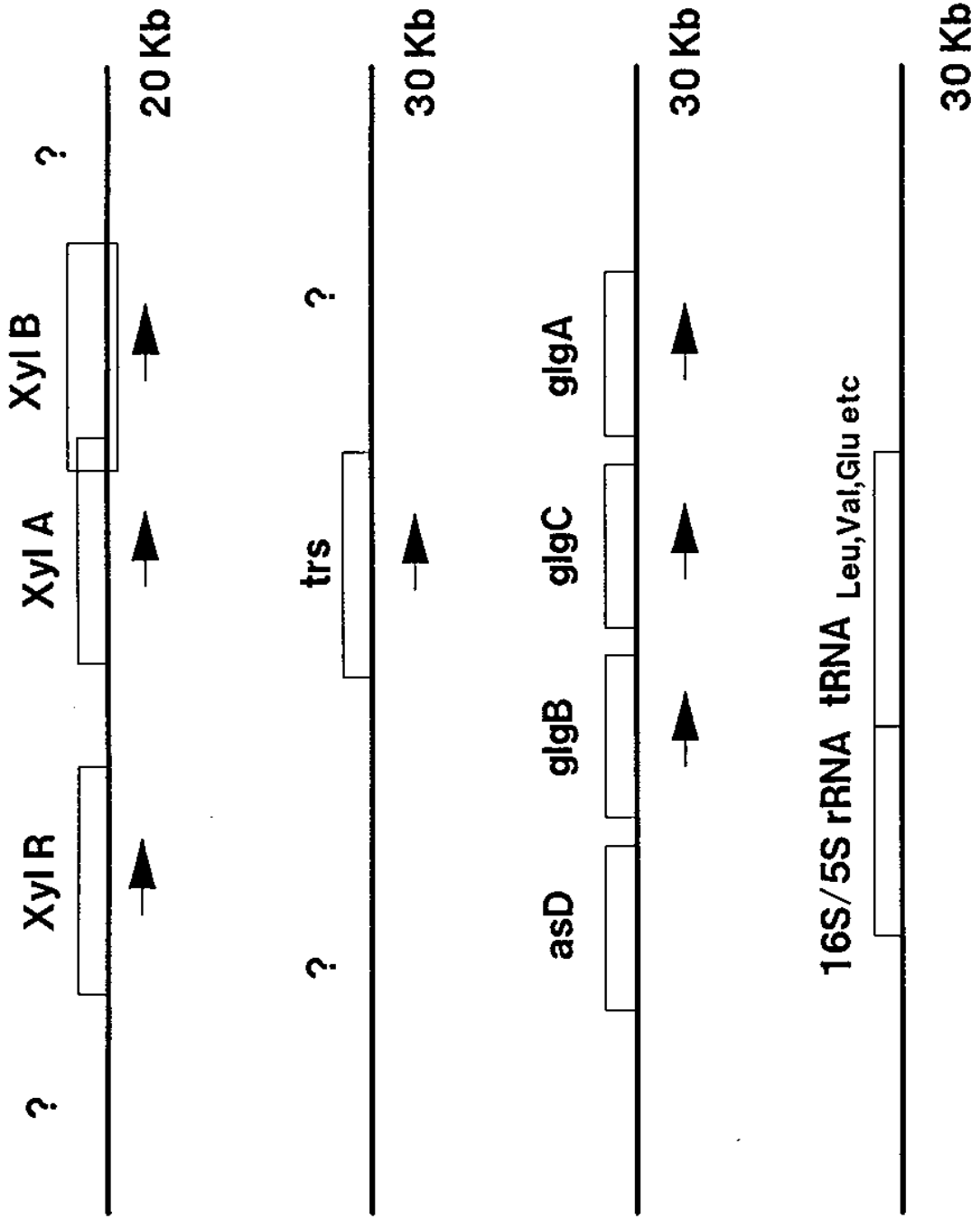


그림 9. Genome search결과에서 구성된 linkage map

BLASTN 1.2.12 [4-Sep-92] [Build 12:32:01 May 19 1994]
Smallest

Sequences producing High-scoring Segment Pairs:			High Score	Poisson Probability P(N)	N
I01968 I01968	Sequence 1 from patent US 4795699		284	1.6e-14	1
Z21946 SCTRXABA	S.clavuligerus trxB gene encoding thior...		235	2.3e-10	1
X51463 SOTHM4	Spinach mRNA for thioredoxin m (clone p...		221	3.5e-09	1

>I01968|I01968 Sequence 1 from patent US 4795699
Length = 1900

Plus Strand HSPs:

Score = 284 (78.5 bits), Expect = 1.6e-14, P = 1.6e-14
Identities = 80/109 (73%), Positives = 80/109 (73%), Strand = Plus

```
Query: 101 GGTCTGGTGGACTTCTGGGCGGAGTGGTGCGCCCCCTGCCGGATGATCGCCCCCATCCT 160
      | ||| || || ||||| ||||| || ||| ||||| ||| ||
Sbjct: 421 GATCCTCGTCGATTTCTGGGCAGAGTGGTGCGGTCCGTGCAAGATGATCGCCCCGATTCT 480

Query: 161 GGAGGAGATCGCCAAAGAGTACGAGGGGAAGCTTCTGGTGGCCAAGCTG 209
      ||| || ||||| | || | ||| || | || || | ||
Sbjct: 481 GGATGAAATCGCTGACGAATATCAGGGCAAACCTGACCGTTGCAAACCTG 529
```

>Z21946|SCTRXABA S.clavuligerus trxB gene encoding thioredoxin reductase and trxA gene encoding thioredoxin.ept
Length = 1806

Plus Strand HSPs:

Score = 235 (64.9 bits), Expect = 2.3e-10, P = 2.3e-10
Identities = 59/74 (79%), Positives = 59/74 (79%), Strand = Plus

```
Query: 102 GTCCTGGTGGACTTCTGGGCGGAGTGGTGCGCCCCCTGCCGGATGATCGCCCCCATCCTG 161
      || ||||| ||||| ||||| || ||||| ||||| |||
Sbjct: 1415 GTTCTGGTGGACTTCTGGGCCGAGTGGTGCGGTCCGTGCCCGCCAGATCGCCCCGTCGCTG 1474

Query: 162 GAGGAGATCGCCAA 175
      ||| ||| || |
Sbjct: 1475 GAGGCGATCACCGA 1488
```

>X51463|SOTHM4 Spinach mRNA for thioredoxin m (clone p6SocTRM-4)
Length = 893

Plus Strand HSPs:

Score = 221 (61.1 bits), Expect = 3.5e-09, P = 3.5e-09
Identities = 73/109 (66%), Positives = 73/109 (66%), Strand = Plus

```
Query: 101 GGTCTGGTGGACTTCTGGGCGGAGTGGTGCGCCCCCTGCCGGATGATCGCCCCCATCCT 160
      ||| ||||| ||||| ||||| | ||||| | | ||||| | |
Sbjct: 328 GGTGATGGTGGACTTCTGGGCCCATGGTGTGGACCCTGCAAGCTCATCGCCCCGGTGAT 387

Query: 161 GGAGGAGATCGCCAAAGAGTACGAGGGGAAGCTTCTGGTGGCCAAGCTG 209
      || ||| | || ||||| || || ||| | || |||||
Sbjct: 388 CGACGAGCTGGCAAAGAATATTCCGGAAAGATAGCAGTATACAAGCTG 436
```

그림 10. 유전자의 전산분석(계속)

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Poisson Probability P(N)	N
J03759 BOVKAD Bovine branched chain alpha-keto acid d...	295	1.8e-15	1
M76554 NEMPDHIIAS Ascaris suum pyruvate dehydrogenase typ...	186	3.1e-06	1

>J03759|BOVKAD Bovine branched chain alpha-keto acid dehydrogenase mRNA,
complete cds. cds
Length = 1817

Plus Strand HSPs:

Score = 295 (81.5 bits), Expect = 1.8e-15, P = 1.8e-15
Identities = 99/149 (66%), Positives = 99/149 (66%), Strand = Plus

```

Query:   52 GGTGGCGGTCTGCACCTTCGGGGACGGGGCCACGAGCGAGGGGGACTGGTACGCCGGGAT 111
        |||||  |||  ||| || || |||||  | || |||||  |||||  |
Sbjct:  820 GGTGGTCATCTGTTACTTTGGAGAGGGGGCGGCCAGTGAGGGGGACGCCACGCCGGCTT 879

Query:   112 CAACTTCGCCCGCCGTGCAGGGGGCGCCCGCGGTCTTCGTCTGTGAGAACAACCTTTACGC 171
        |||||  |||||  |||  | | | |||  |||||  |||||  |||||  |||||
Sbjct:  880 CAACTTCGCCCGCCACCTCGAGTGCCCCATCATCTTCTTCTGTGGAACAACGGCTACGC 939

Query:   172 CATCAGCGTGGACGACGGGGAGATGTACC 200
        ||||  | | | | | ||||  |||||
Sbjct:  940 CATCTCCACGCCACCTCGGAGCAGTACC 968
    
```

>M76554|NEMPDHIIAS Ascaris suum pyruvate dehydrogenase type II alpha subunit
mRNA, 3' end. cds
Length = 1673

Plus Strand HSPs:

Score = 186 (51.4 bits), Expect = 3.1e-06, P = 3.1e-06
Identities = 70/111 (63%), Positives = 70/111 (63%), Strand = Plus

```

Query:   68 TTCGGGGACGGGGCCACGAGCGAGGGGGACTGGTACGCCGGGATCAACTTCGCCCGCGTG 127
        |||||  |||||  |||||  | | ||||  | || | |||  |||  |||  ||
Sbjct:  555 TTCGGTGACGGTGCCACCAACCAGGGACAGCTTTATGAGTCGATGAACATGGCTAAATTG 614

Query:   128 CAGGGGGCGCCCGCGGTCTTCGTCTGTGAGAACAACCTTTACGCCATCAGC 178
        || | |||  | | | | | |||  |||||  |||||  ||||  |||  ||
Sbjct:  615 TGGGAGTTGCCCGTGCTGTACGTGTGTGAGAACAACGGATACGGCATGGGC 665
    
```

Database: pir
 67,423 sequences; 19,747,297 total residues.
 Searching.....done

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Poisson Probability P(N)	N
S QRECLH leucine transport protein livH - Escherichia coli	83	1.7e-05	1
S D37074 leucine transport protein livH - Escherichia coli	83	1.7e-05	1
S JH0668 high-affinity branched-chain amino acid transpor...	83	1.7e-05	1
S B36125 branched-chain amino acid transport protein braD...	83	1.7e-05	1

>S|QRECLH leucine transport protein livH - Escherichia coli
 Length = 308

Score = 83 (39.6 bits), Expect = 1.7e-05, P = 1.7e-05
 Identities = 13/20 (65%), Positives = 18/20 (90%)

Query: 24 LGYTMVYGVLELINFHSEI 43
 +GYTMVYG++ +INFH E+
 Sbjct: 27 IGYTMVYGIIGMINFAHGEV 46

>S|D37074 leucine transport protein livH - Escherichia coli
 Length = 308

Score = 83 (39.6 bits), Expect = 1.7e-05, P = 1.7e-05
 Identities = 13/20 (65%), Positives = 18/20 (90%)

Query: 24 LGYTMVYGVLELINFHSEI 43
 +GYTMVYG++ +INFH E+
 Sbjct: 27 IGYTMVYGIIGMINFAHGEV 46

>S|JH0668 high-affinity branched-chain amino acid transport LIV-I system LivA
 protein - Salmonella typhimurium
 Length = 308

Score = 83 (39.6 bits), Expect = 1.7e-05, P = 1.7e-05
 Identities = 13/20 (65%), Positives = 18/20 (90%)

Query: 24 LGYTMVYGVLELINFHSEI 43
 +GYTMVYG++ +INFH E+
 Sbjct: 27 IGYTMVYGIIGMINFAHGEV 46

>S|B36125 branched-chain amino acid transport protein braD - Pseudomonas
 aeruginosa (strain PAO)
 Length = 307

Score = 83 (39.6 bits), Expect = 1.7e-05, P = 1.7e-05
 Identities = 13/20 (65%), Positives = 18/20 (90%)

Query: 24 LGYTMVYGVLELINFHSEI 43
 +GYTMVYG++ +INFH E+
 Sbjct: 26 IGYTMVYGIIGMINFAHGEV 45

Smallest

	High Score	Poisson Probability P(N)	N
Sequences producing High-scoring Segment Pairs:			
S A32993 transcription factor Eryf1 - chicken	52	1.4e-05	4
<p>>S A32993 transcription factor Eryf1 - chicken Length = 304</p> <p>Score = 52 (24.5 bits), Expect = 2.2, P = 0.89 Identities = 9/16 (56%), Positives = 10/16 (62%)</p> <p>Query: 6 PPSGGPARPGAGLLP 21 PP G P P +G LLP Sbjct: 82 PPRGSPPHSSGPLL 97</p> <p>Score = 42 (19.8 bits), Expect = 70., Poisson P(2) = 0.038 Identities = 9/17 (52%), Positives = 10/17 (58%)</p> <p>Query: 9 GGPARGAGLLPGHTP 25 GG R AGLL + P Sbjct: 28 GGERTEAGLLASYPP 44</p> <p>Score = 37 (17.5 bits), Expect = 3.9e+02, Poisson P(3) = 0.0024 Identities = 7/16 (43%), Positives = 9/16 (56%)</p> <p>Query: 1 PWFPPPPSGGPARGA 16 P P PPP+ P + A Sbjct: 249 PPPPPPPAAAPPQSDA 264</p> <p>Score = 36 (17.0 bits), Expect = 5.5e+02, Poisson P(4) = 1.4e-05 Identities = 6/14 (42%), Positives = 7/14 (50%)</p> <p>Query: 5 PPPSGGPARGAGG 18 PP G P+ GG Sbjct: 224 PPGGNPSATAGGG 237</p>			

Smallest

		High	Poisson	
		Score	Probability	N
			P(N)	
Sequences producing High-scoring Segment Pairs:				
S RNECC	DNA-directed RNA polymerase (EC 2.7.7.6) beta' c...	118	9.9e-11	1
S S32681	RNA polymerase beta' chain - Buchnera aphidicola...	117	1.4e-10	1
S S31146	DNA-directed RNA polymerase (EC 2.7.7.6) beta' c...	108	2.4e-09	1
S S20583	DNA-directed RNA polymerase (EC 2.7.7.6) gamma c...	107	3.4e-09	1
S B42361	DNA-directed RNA polymerase (EC 2.7.7.6) gamma c...	106	4.6e-09	1
S S14550	DNA-directed RNA polymerase (EC 2.7.7.6) gamma c...	106	4.6e-09	1
S S20588	DNA-directed RNA polymerase (EC 2.7.7.6) gamma c...	103	1.2e-08	1
S S20587	DNA-directed RNA polymerase (EC 2.7.7.6) gamma c...	102	1.7e-08	1
S S20584	DNA-directed RNA polymerase (EC 2.7.7.6) gamma c...	102	1.7e-08	1
S S20586	DNA-directed RNA polymerase (EC 2.7.7.6) gamma c...	100	3.2e-08	1
S S20585	DNA-directed RNA polymerase (EC 2.7.7.6) gamma c...	99	4.4e-08	1
S RNZMB1	DNA-directed RNA polymerase (EC 2.7.7.6) beta'-1...	98	6.0e-08	1
S RNZC1	DNA-directed RNA polymerase (EC 2.7.7.6) beta'-1...	98	6.0e-08	1
S B29959	DNA-directed RNA polymerase (EC 2.7.7.6) beta' c...	97	8.3e-08	1
S A05033	rpoC protein homolog - common tobacco chloroplast	97	8.3e-08	1
S RNLC1	DNA-directed RNA polymerase (EC 2.7.7.6) beta'-1...	94	2.2e-07	1
S S01574	DNA-directed RNA polymerase (EC 2.7.7.6) beta'-1...	94	2.2e-07	1
S JN0420	DNA-directed RNA polymerase (EC 2.7.7.6) beta' c...	94	2.2e-07	1
S A32838	DNA-directed RNA polymerase (EC 2.7.7.6) gamma c...	75	9.7e-05	1
S RNEGB1	DNA-directed RNA polymerase (EC 2.7.7.6) beta'-1...	75	9.7e-05	1
S S34555	DNA-directed RNA polymerase (EC 2.7.7.6) beta'-1...	75	9.7e-05	1

>S|RNECC DNA-directed RNA polymerase (EC 2.7.7.6) beta' chain (version 1) -
Escherichia coli
Length = 1407

Score = 118 (54.6 bits), Expect = 9.9e-11, P = 9.9e-11
Identities = 21/28 (75%), Positives = 25/28 (89%)

Query: 14 VRRYRMGHIELATPAAHIWFKDVPSKI 41
VRR RMGHIELA+P AHIWF+K +PS+I
Sbjct: 97 VRRERMGHIELASPTAHIWFLKSLPSRI 124

>S|S32681 RNA polymerase beta' chain - Buchnera aphidicola (fragment)
Length = 209

Score = 117 (54.1 bits), Expect = 1.4e-10, P = 1.4e-10
Identities = 21/31 (67%), Positives = 27/31 (87%)

Query: 11 KSIVRRYRMGHIELATPAAHIWFKDVPSKI 41
+S VRR RMGHIEL++P AHIWF+K +PS+I
Sbjct: 94 QSKVRRERMGHIELSSPTAHIWFLKSLPSRI 124

>S|S31146 DNA-directed RNA polymerase (EC 2.7.7.6) beta' chain - Mycobacterium
leprae
Length = 1316

Score = 108 (50.0 bits), Expect = 2.4e-09, P = 2.4e-09
Identities = 19/28 (67%), Positives = 22/28 (78%)

Query: 14 VRRYRMGHIELATPAAHIWFKDVPSKI 41
VRR RMGHIELA P HIW+ K VPS++
Sbjct: 87 VRRERMGHIELAAPVTHIWIYFKGVPSRL 114

>S|S20583 DNA-directed RNA polymerase (EC 2.7.7.6) gamma chain - Cyanophora

paradoxa cyanelle (fragment)
Length = 208

Score = 107 (49.5 bits), Expect = 3.4e-09, P = 3.4e-09
Identities = 19/31 (61%), Positives = 25/31 (80%)

Query: 11 KSIVRRYRMGHIELATPAAHIWFKDVPSKI 41
+S VRR+RMG+IELA P HIW++K +PS I
Sbjct: 58 ESQVRRHRMGYIELAAPVTHIWIYKLGIPSYI 88

>S|B42361 DNA-directed RNA polymerase (EC 2.7.7.6) gamma chain - Anabaena sp.
(PCC 7120)
Length = 625

Score = 106 (49.1 bits), Expect = 4.6e-09, P = 4.6e-09
Identities = 18/31 (58%), Positives = 26/31 (83%)

Query: 11 KSIVRRYRMGHIELATPAAHIWFKDVPSKI 41
+S VRR+RMG+I+LA P AH+W++K +PS I
Sbjct: 95 ESRVRRHRMGYIKLAAPVAHVWYKLGIPSYI 125

>S|S14550 DNA-directed RNA polymerase (EC 2.7.7.6) gamma chain - Anabaena sp.
(PCC 7120)
Length = 625

Score = 106 (49.1 bits), Expect = 4.6e-09, P = 4.6e-09
Identities = 18/31 (58%), Positives = 26/31 (83%)

Query: 11 KSIVRRYRMGHIELATPAAHIWFKDVPSKI 41
+S VRR+RMG+I+LA P AH+W++K +PS I
Sbjct: 95 ESRVRRHRMGYIKLAAPVAHVWYKLGIPSYI 125

>S|S20588 DNA-directed RNA polymerase (EC 2.7.7.6) gamma chain - Synechococcus
sp. (PCC 7942) (fragment)
Length = 203

Score = 103 (47.7 bits), Expect = 1.2e-08, P = 1.2e-08
Identities = 18/31 (58%), Positives = 25/31 (80%)

Query: 11 KSIVRRYRMGHIELATPAAHIWFKDVPSKI 41
+S VRR+RMG I+LA P AH+W++K +PS I
Sbjct: 58 ESRVRRHRMGFIKLAAPVAHVWYKLGIPSYI 88

>S|S20587 DNA-directed RNA polymerase (EC 2.7.7.6) gamma chain -
Prochlorococcus sp. (strain DV1) (fragment)
Length = 204

Score = 102 (47.2 bits), Expect = 1.7e-08, P = 1.7e-08
Identities = 16/31 (51%), Positives = 26/31 (83%)

Query: 11 KSIVRRYRMGHIELATPAAHIWFKDVPSKI 41
+S VRR+RMG+I+LA P +H+W++K +PS +
Sbjct: 58 ESRVRRHRMGYIKLAAPVSHVWYKLGIPSYV 88

>S|S20584 DNA-directed RNA polymerase (EC 2.7.7.6) gamma chain - Prochlorothrix
hollandica (fragment)
Length = 203

Score = 102 (47.2 bits), Expect = 1.7e-08, P = 1.7e-08

BLASTN 1.2.12 [4-Sep-92] [Build 12:19:01 Sep 23 1994]
Smallest

Sequences producing High-scoring Segment Pairs:	High Score	Poisson Probability P(N)	N
D10390 PSEFAOAB P. fragi faoA and faoB genes, complete cds.	195	5.5e-07	1

>D10390|PSEFAOAB P. fragi faoA and faoB genes, complete cds.
Length = 3480

Minus Strand HSPs:

Score = 195 (53.9 bits), Expect = 5.5e-07, P = 5.5e-07
Identities = 59/84 (70%), Positives = 59/84 (70%), Strand = Minus

```
Query: 127 CAAGGCCAAGTTCGGTCAGCCGGAAGTCAACCTCGGCATCAGCCCTGGTATCGGTGGCTC 68
      || ||||| |||| | ||||| ||||| || ||||| ||| || | || ||| |
Sbjct: 444 CAGCGCCAAGATCGGCCTGCCGGAAGTCAAACCTGGGCATCTACCCGGGCTTTGGCGGCAC 503

Query: 67 CCAGCGTCTTACCCGCGCGGTGGG 44
      | |||| | || | | ||
Sbjct: 504 CGTACGTCTGCCACGCTGATCGG 527
```

BLASTN 1.2.12 [4-Sep-92] [Build 12:19:01 Sep 23 1994]

Database: gb

269,478 sequences; 248,499,214 total residues.

Searching.....done

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Poisson Probability P(N)	N
D90223 PSEBRACG P.aeruginosa braCDEFG operon for branch...	174	2.8e-05	1

>D90223|PSEBRACG P.aeruginosa braCDEFG operon for branched-chain amino acid transport (LIV-I) components. acid hydrolase (bphD) genes,
Length = 5302

Plus Strand HSPs:

Score = 174 (48.1 bits), Expect = 2.8e-05, P = 2.8e-05
Identities = 46/60 (76%), Positives = 46/60 (76%), Strand = Plus

Query: 100 GCTGGGGTACACCATGGTCTACGGAGTCTTGGAGCTCATCAACTTCGCCCACTCCGAGAT 159
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct: 1612 GATCGGCTACACCATGGTCTACGGCATCATCGGCATGATCAACTTCGCCCACGGCGAGGT 1671

BLASTN 1.2.12 [4-Sep-92] [Build 12:19:01 Sep 23 1994]

Database: gb
269,478 sequences; 248,499,214 total residues.
Searching.....done

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Poisson Probability P(N)	N
X16278 TTFUS Thermus thermophilus fus gene for elongat...	135	0.15	1
>X16278 TTFUS Thermus thermophilus fus gene for elongation factor G (EF-G). Length = 2082			

Plus Strand HSPs:

Score = 135 (37.3 bits), Expect = 0.16, P = 0.15
Identities = 63/108 (58%), Positives = 63/108 (58%), Strand = Plus

```

Query:   76 CTGATCCTCACCTCCACGCCAAGTACCGGGAGGAGGGGGAGACCTTCGTACCCAAGTAC 135
      | | |||| | | ||||| ||||| |||| | | | |
Sbjct: 1066 CGGCTCCTCCGCATGCACGCCAACCACCGGGAGGAGGTGGAGGAGCTCAAGCGGGCGAC 1125

Query:   136 CTGCGGATTTTGGAGCCCGGGAGAGCACTAACCCCAAGGAGATCCTG 183
      ||| |   ||| | |   ||   || | | || ||| | |||
Sbjct: 1126 CTGGCGCCGTGGTGGGCCTCAAGGAGACGATCACCGGGGACACCCTG 1173

```


BLASTN 1.2.12 [4-Sep-92] [Build 12:19:01 Sep 23 1994]

Database: gb
269,478 sequences; 248,499,214 total residues.

Searching.....done

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Poisson P(N)	Probability	N
X07744 TTRPEGL Thermus thermophilus genes trpL, trpE a...	338	1.6e-36	2	
X14445 HSINT2 Human int-2 proto-oncogene.	117	0.00045	2	

>X07744|TTRPEGL Thermus thermophilus genes trpL, trpE and trpG.
Length = 2340

Plus Strand HSPs:

Score = 338 (93.4 bits), Expect = 1.1e-19, P = 1.1e-19
Identities = 70/73 (95%), Positives = 70/73 (95%), Strand = Plus

```
Query: 71 GGCCCGGGGCGGGAGGCCTTCTCCCGGGGCACACCCCGGGGCTTTGTTTTGGGGGACGG 130
      | |||
Sbjct: 108 GCCCGGGGCGGGAGGCCTTCTCCCGGGGCACACCCCGGGGCTTTGTTTTGGGGGACGG 167

Query: 131 CATGGAGCGGATC 143
      |||
Sbjct: 168 CATGGAGCGGATC 180
```

Score = 295 (81.5 bits), Expect = 8.3e-16, Poisson P(2) = 1.6e-36
Identities = 75/95 (78%), Positives = 75/95 (78%), Strand = Plus

```
Query: 32 GGCCATGGCCCTTCCCTCCGCCCTCTGGTGGCCCGGCTAGGCCCGGGGCGGGAGGCCTTC 91
      |||
Sbjct: 68 GGCCATGGCCCTTCCCTCCGCCCTCTGGTGGCCCGGCTAGGCCCGGGGCGGGAGGCCTT 127

Query: 92 TCCCGGGGCACACCCCGGGGCTTTGTTTTGGGGG 126
      |||
Sbjct: 128 TCCCGGGGCACACCCCGGGGCTTTGTTTTGGGG 162
```

Sequences producing High-scoring Segment Pairs:		High Score	Smallest Poisson Probability P(N)	N
L27989 MSGRPOB	Mycobacterium tuberculosis RNA polymera...	225	1.2e-09	1
U00006 ECOUW89	E. coli chromosomal region from 89.2 to...	182	5.5e-06	1
V00339 ECRPOBC	E. coli operon rpoBC coding for the bet...	182	5.5e-06	1

>L27989|MSGRPOB Mycobacterium tuberculosis RNA polymerase beta-subunit (rpoB) gene, complete cds and RNA polymerase beta'-subunit rpoC gene, ern
Length = 5084

Plus Strand HSPs:

Score = 225 (62.2 bits), Expect = 1.2e-09, P = 1.2e-09
Identities = 61/81 (75%), Positives = 61/81 (75%), Strand = Plus

```

Query:   69 CATCGTCCGCCGCTACCGCATGGGGCACATTGAGCTCGCCACCCCGCCGCCACATCTG 128
      || || || || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct:  4895 CAAGGTGCGTCGTGAGCGGATGGGCCACATCGAGCTTGCCGCGCCCGTCACCCACATCTG 4954

Query:   129 GTTCGTCAAGGACGTCCCCTC 149
      || | | | | | | | | | | | | | |
Sbjct:  4955 GTACTTCAAGGGTGTGCCCTC 4975
    
```

>U00006|ECOUW89 E. coli chromosomal region from 89.2 to 92.8 minutes.
Length = 176,195

Plus Strand HSPs:

Score = 182 (50.3 bits), Expect = 5.5e-06, P = 5.5e-06
Identities = 70/112 (62%), Positives = 70/112 (62%), Strand = Plus

```

Query:   45 GAGCGGTGCGGGAGGTCGGAAGAGCATCGTCCGCCGCTACCGCATGGGGCACATTGAGCT 104
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct:  50851 GTGCGGCGTTGAAGTGACCCAGACTAAAGTACGCCGTGAGCGTATGGGCCACATCGAACT 50910

Query:   105 CGCCACCCCGCCGCCACATCTGGTTCGTCAAGGACGTCCCCTCCAAGATC 156
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct:  50911 GGCTTCCCGACTGCGCACATCTGGTTCCTGAAATCGCTGCCGTCCCGTATC 50962
    
```

>V00339|ECRPOBC E. coli operon rpoBC coding for the beta- and beta'-subunits of RNA polymerase (genes rpoC and rpoB), and genes rplL, rlpJ, rplA, and
Length = 12,337

Plus Strand HSPs:

Score = 182 (50.3 bits), Expect = 5.5e-06, P = 5.5e-06
Identities = 70/112 (62%), Positives = 70/112 (62%), Strand = Plus

```

Query:   45 GAGCGGTGCGGGAGGTCGGAAGAGCATCGTCCGCCGCTACCGCATGGGGCACATTGAGCT 104
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct:  7334 GTGCGGCGTTGAAGTGACCCAGACTAAAGTACGCCGTGAGCGTATGGGCCACATCGAACT 7393

Query:   105 CGCCACCCCGCCGCCACATCTGGTTCGTCAAGGACGTCCCCTCCAAGATC 156
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct:  7394 GGCTTCCCGACTGCGCACATCTGGTTCCTGAAATCGCTGCCGTCCCGTATC 7445
    
```

BLASTP 1.3.0 [16-Nov-92] [Build 20:30:50 Jun 18 1993]

Reference: Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman (1990). Basic local alignment search tool. J. Mol. Biol. 215:403-410.

WARNING: Invalid residue ``□'' encountered in query sequence will be skipped.
Query= NO NAME
(47 residues)

Database: pir
67,423 sequences; 19,747,297 total residues.
Searching.....done

Sequences producing High-scoring Segment Pairs:	High Score	Smallest Poisson Probability P(N)	N
S S31223 transcription factor brain-1 - mouse	60	1.3e-05	3
S S22429 Collagen alpha 3(IX) chain precursor - Chicken	43	1.5e-05	5
S KGZQHL histidine-rich glycoprotein precursor - Plasmodi...	54	1.8e-05	3
S JQ0405 hypothetical 119.5K protein (uvrA region) - Micr...	44	1.9e-05	5
S GNWVR3 structural polyprotein - rubella virus (strain M33)	45	6.8e-05	5
S A47446 Qin=HNF-3/fork head family transcription factor ...	54	0.00010	2

>S|S31223 transcription factor brain-1 - mouse
Length = 495

Score = 60 (27.9 bits), Expect = 0.20, P = 0.18
Identities = 12/27 (44%), Positives = 12/27 (44%)

Query: 12 HHAGPLHHRGAHHSSPPRRRTGSHGG 38
H LHRG H PPP H G
Sbjct: 156 HAGTALHHRGPPHLP PPPPPHGHGPG 182

Score = 48 (22.3 bits), Expect = 13., Poisson P(2) = 0.0020
Identities = 10/29 (34%), Positives = 11/29 (37%)

Query: 9 HLGHHAGPLHHRGAHHSSPPRRRTGSHG 37
H HH H HH+ PP G G
Sbjct: 269 HHHHHHAHPHPPHHAQPPHGGGGAG 297

Score = 44 (20.5 bits), Expect = 50., Poisson P(3) = 1.3e-05
Identities = 8/22 (36%), Positives = 10/22 (45%)

Query: 18 HHRGAHHSSPPRRRTGSHGGP 39
HH HH+ P P + G P
Sbjct: 268 HHHHHHAHPHPPHHAQPP 289

Score = 36 (16.7 bits), Expect = 7.6e+02, Poisson P(4) = 0.00021
Identities = 7/20 (35%), Positives = 9/20 (45%)

Query: 4 PAVAQHLGHHAGPLHHRGAH 23
P +A+H HH H H
Sbjct: 262 PELAEHHHHHHHAHPHPPH 281

Score = 32 (14.9 bits), Expect = 2.9e+03, Poisson P(5) = 0.00065
Identities = 7/18 (38%), Positives = 8/18 (44%)

Query: 27 PPPRRRTGSHGGPGRQL 44

xxr84, 448 nt vs t library
 using DNA matrix

initn initl
 < 4 670 670:=====

2989699 residues in 5719 sequences
 statistics exclude scores greater than 125
 mean initn score: 28.2 (12.50)
 mean initl score: 28.2 (12.47)
 5046 scores better than 1 saved, ktup: 6, fact: 6
 joining threshold: 50 scan time: 0:00:12

The best scores are: initn initl opt
 I09223 Sequence 1 from patent WO 8901520. 149 149 252
 65.5% identity in 165 nt overlap

	260	270	280	290	300	310
xxr84	TCCCTGCCTTTTGAGGAGGTGATGTAGGTGTACGAACCCAAACCGGAGCACAGGTTTACC					
				::: : ::::	:: :: ::::	::: : ::
I09223	ATGAACTACCAGCCACCCCGAGGACAGGTTTACC					
			10	20	30	

	320	330	340	350	360	370
xxr84	TT-TGCCTTTGGACTGTGGACAATGTGGACCGTGATCCCTTCGGGGACACGGTTCGGGAG					
	::	: ::	::: : ::	:: : : X::: : ::	::: : ::	::: : ::
I09223	TTCGGACTGTGGACCGTTCGGCTGGCAGGGACGGGACCCCTTCGGTGACGCCACGGCGGCGC					
	40	50	60	70	80	90

	380	390	400	410	420	430
xxr84	AGACTGGACCCGGTTTACGTGGTTCATAAGCTGGCGGAGCTTGGGGCCTACGGGGTAAAC					
	::	::: : ::	::: : ::	::: : ::	::: : ::	::: : ::
I09223	GCCCTCGACCCGGTTCGAGTCGGTTCGGCGGCTGGCCGAGCTGGGCGCCACGGCGTCACG					
	100	110	120	130	140	150

	440
xxr84	CTTCACGACGAGGAC
	: ::: : X :::
I09223	TTCCACGACGACGACCTCATCCCCTTCGGCTCCAGCGACAGCGAGCGGAGGAGCACGTC
	160 170 180 190 200 210

using DNA matrix

The best scores are:

			initn	initl	opt
TTHTAQPIA	Thermus aquaticus DNA polymerase (PolI)	gen	160	92	531

TTHTA

QP1A	Thermus aquaticus DNA polymerase (PolI)	gen	160	92	231
93.3% identity in 132 nt overlap					

		10	20	30	40
n28r		TGATTACGAATTCATGAGGGGGATGCTGCCCTC--TGAGCC			
		X::::::::::::::::::::::::::v ^:::::			
TTHTAQ	GCTACCCGGGGCGGGTGGTGGGAGGGTAACATGAGGGGGATGCTGCCCTCTTTGAGCC				
	90 100 110 120 130 140				

	50	60	70	80	90
n28r	CAAGGGCCGGGTC--TCTGGTGGACGG---CTCTCTGG-CTACCG-ACCTT-CACGGCCT				
	::::::::::::: ::::::::::: : :::: :::::: :::::: :::: ::::				
TTHTAQ	CAAGGGCCGGGTCCTCCTGGTGGACGGCCACCACCTGGCCTACCGCACCTCCACGCCCT				
	150 160 170 180 190 200				

	100	110	120	130
n28r	GAAGGGCTACACCACCAGCCGGGGGAGCCGGTGCAGGCGGTCTACGG			
	:::::::::: :::::::::::::::::::::			
TTHTAQ	GAAGGGCCTCACCACCAGCCGGGGGAGCCGGTGCAGGCGGTCTACGGCTTCGCCAAGAG			
	210 220 230 240 250 260			

BLASTN 1.2.12 [4-Sep-92] [Build 12:32:01 May 19 1994]
Smallest

Sequences producing High-scoring Segment Pairs:	High Score	Poisson Probability P(N)	N
D00332 STYBRNQ Salmonella typhimurium brnQ gene and fl...	179	1.7e-05	1

>D00332|STYBRNQ Salmonella typhimurium brnQ gene and flanks.
Length = 1603

Minus Strand HSPs:

Score = 179 (49.5 bits), Expect = 1.7e-05, P = 1.7e-05
Identities = 59/88 (67%), Positives = 59/88 (67%), Strand = Minus

```
Query: 143 TAATTCGACCAGTATCATCGCGCCGGACTTGCGCAGCTGATCGATCAGGGAAGTCTTTGA 84
      ||||| ||| |||||||||||||||| | || | | | | | || |||
Sbjct: 1314 TAATTCACCCGTATCATCGCGCCGGCCATGTTTATCAGTCTGCTTTTTGGTATCCTTGA 1373

Query: 83 CGCCATCCTAGCACCTTGTTTGGGCGAT 56
      || |||| |||| || || |||||
Sbjct: 1374 CGGCATCAAAGCATCTGCCTTCGGCGAT 1401
```

glcfl, 271 nt vs b library
using DNA matrix

The best scores are:

		initn	initl	opt
RCAPRKBA	R.sphaeroides phosphoribulokinase gene, 3'	116	84	124
STMGLNA	Streptomyces coelicolor glutamine synthetas	113	68	138
STMAMY	Streptomyces thermoviolaceus alpha-amylase	109	63	104
RCAPRKBA	R.sphaeroides phosphoribulokinase gene, 3'	116	84	124

60.0% identity in 120 nt overlap

	120	130	140	150	160	170
glcfl	TGGAGTACGTCTTCCGCAACCACACGAACCCGGAGCGGGTCTTCGTCACGGGGTGCAGCG					
				:: X::	::	::
RCAPRK	AGAAGGCCCGCGGCCCATCTGCAGAACGGGGCGAAGCGGGTGTCTCGTCTCGGCGCCCTGCG					
	2490	2500	2510	2520	2530	2540

	180	190	200	210	220	230
glcfl	CCGGGGCCTACGGGGCGGTGCTTTTGGCGGACAAGATCCTCGCCACCTACAAAAACGCC					
	:::::X::	::	::	::	::	::
RCAPRK	ACGGGGCCGACCGGACCATCGTCTATGGGG---TGAACCATGCGACGCTCACCGCGGACG					
	2550	2560	2570	2580	2590	

	240	250	260	270
glcfl	AGATCGCCGTCT-GCTGGCCCGCCGACCACATATG			
	: ::	::	::	::
RCAPRK	ACCTCGTGGTCTCGAATGCCTCCTGCACCACCAACTGCCTCTCGCCGGTGGCCAAGGTGC			
	2600	2610	2620	2630

STMGLNA Streptomyces coelicolor glutamine synthetas 113 68 138
55.1% identity in 265 nt overlap

	10	20	30
glcfl	AAACTTCTTCGTCCCCTACTGCACGGGGGAACCTTCCA		
	::	::	::
STMGLN	CGAGCAGGGCTACGCCGGCCTGTCCGACACCCCGCTACTACATCGGGCGCATCCTCAA		
	1100	1110	1120

	40	50	60	70	80	90
glcfl	ACGTGGGCCGGGCCACGG---TGGACTAA-CGGGGGCTTCCAAGGTCCCACCACAGGG					
	: : ::	: :	: :	: :	: :	: :
STMGLN	GCACGCCCGCTCGCTGCTGGCCTTACCAACCCGACGGTGAACCTCGTACCACCGCC-TGG					
	1160	1170	1180	1190	1200	1210

	100	110	120	130	140	150
glcfl	GGCCAGG----AACGCCAGGCCGCCCTGGAGTACGTCTTCGCAACCACACGAACCCGG					
	::	::	::	X::	::	::
STMGLN	TGCCGGGCTTCGAGGCGCCGGTGAACCTGGTGTACTCGCAGCGCAACCGCTCGGCCGCGA					
	1220	1230	1240	1250	1260	1270

	160	170	180	190	200	210
glcfl	AGCGGGTCTTCGTCACGGGGTGCAGCGCCGGGGCCTACGGGGCGGTGCTTTTGGCGGACA					
	::: : :	::: X :	: : :	: : :	: : :	: : :
STMGLN	TGCGGATCCCGATCACGGGCTCGAACCCGAAGGCCAA--GCGCGTCGAGTTCGCGCCCC					
	1280	1290	1300	1310	1320	

	220	230	240	250	260
glcfl	AGA-TCCTC--GCCACCTACAAAACGCCAGATCGCCGTCTGCTGGCCGCCGGCACCAT				
	:: : ::	: : :	: : :	: : :	: : :
STMGLN	GGACGCCCTCCGGCAACC--CGTACCTGGCGTTCTCGGCG-CTGCTGCTGGCGGGCCTGGA				
	1330	1340	1350	1360	1370

제 4 장 결론 및 건의사항

제 1 절 결론

1. 게놈지도작성: *T. caldophilus* GK24 (그림1)의 제한효소지도 작성과 유전적 지도 작성이 수행되었다.
2. 특정 생합성 관련 효소유전자군의 선정과 집중연구
가. 탄수화물 관련 효소들인 Xylose isomerase, Pullulanase, Trehalose synthase, 그리고 UDP-glucose transferase의 유전자군 게놈이 확인되었다.
나. DNA결합효소들인 DNA polymerase와 Ligase의 유전자군 게놈이 확인되었다.
3. 유전자군의 클로닝과 염기서열결정: 특정 효소 유전자군과 Ssp4 게놈 절편(314 Kb)을 중심으로 유전자 군을 탐색하고 이들을 각각 대장균에 클로닝하고 염기서열을 결정하였다.
4. 유전자의 전산분석과 새로운 효소의 탐색: *T. caldophilus* GK24의 유전자는 G+C 함량이 70 % 가까이 되어 단순한 염기배열 비교를 통해서만 유전자의 특성 추정이 어려우므로 이러한 유전자분석을 위한 분석체계를 구축하고 이렇게 분석된 유전자군을 이용하여 새로운 효소를 계속해서 확보하고자 하였다.

제 2 절 건의사항

내열성 탄수화물 효소류의 경우에는 개개의 효소나 효소군이 실제 산업에서 한가지 공정이나 제품생산을 바꿀수 있기 때문에 세계적으로 수없이 많은 연구가 진행되고 있다. 따라서 연구과정에서 얻어진 Glucose Isomerase, Pullulanase, ADP-glucose pyrophosphorylase, GDP-mannose pyrophosphorylase 등의 효소나 효소군은 내열성이 최대의 장점이 되는 당관련 산업에 바로 적용시켜 사용할 수 있다. 또한 연구과정에서 확립한 게놈분석 및 응용기술은 분석된 유전자의 확인뿐만 아니라 유관 게놈연구의 모델로도 매우 유용하리라고 생각된다.

1. 유용탄수화물의 제조기술 확보
2. 산업적으로 유용한 탄수화물 관련효소의 확보
3. 복잡한 탄수화물의 구조분석방법 개발

4. 유전공학기술을 이용한 산업 효소의 대량생산 체제 구축
5. 내열성 효소의 신용도 개발연구 착수가능
6. Genomic library 로부터 유용단백질 및 효소 확보 가능
7. 계놈연구의 기술개발 기능 확보
8. 인력 확충 및 기기 보완이 필수적임
9. 국내 연구진의 상호협력 체제 구축 요망

제 5 장 참고문헌

1. Taguchi, H., Yamashita, M., Matsuzawa, H., and Ohta, T., *J. Biochem.* 91, 1343-1348. 1982.
2. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W., *Methods in Enzymol.* (Gene expression technol.), 185, 60-89. Academic press. 1990.
3. Kizawa, H., Shinoyama, H., and Yasui, T., *Agric. Biol. Chem.*, 55(3), 671-678. 1991.
4. Lamli, U. K.
5. Plant, A. R., Morgan, H. W., and Daniel, R. M., *Enzyme. Microb. Biotechnol.* vol 8. 668-672. 1986.
6. Smart, J. B., *Appl. Microbiol. Biotechnol.* 34; 495-501. 1991.
7. Eguchi, H., Wakagi, T., and Oshima, T., *Biophysica Biochemica Acta*, 99; 133-137. 1989.
8. Suzuki, Y., Hatagaki, K., and Oda, K., *Applied Microb. Biotechnol.* 34; 707-714. 1991.
9. Love, Ü. D. K., McAnulty, J., Wallace, C., Caughey, P. A., Saul, D. and Berquist, P. L., *Agric. and Env. Microbiol.* 1017-1024. 1990
10. Sambrock, J., Fritsch, E. F., and Maniatis, T., *Molecular Cloning, A laboratory manual* 2nd Ed. Cold Spring Harbor Lab. Press 1989.
11. Kristjansson, J. K., *Thermophilic bacteria*, CRC Press Inc. 1992
12. Ulrich, J. T., Mcfeters, G. A., and Temple, K. L., *J. Bacteriol.* 110, 691, 1972.
13. Cowan, D. A., Daniel, R. M., Martin, A. M., and Morgan, H. W., *Biotechnol. Bioeng.*, 26, 1141, 1984.
14. Takase, M. and Horikoshi, K., *Agric. Biol. Chem.*, 53, 559, 1989.
15. Nakamura, N., Sashihara, N., Nagayama, H., and Horikoshi, K., *Starch*, 41, 112, 1989.
16. Sashihara, N., Nakamura, N., Nagayama, H., and Horikoshi, K., *Fems Microbiol. Lett.*, 49, 385, 1988.
17. Dekker, K., Yomagata, H., Sakaguchi, K., and Udaka, S., *J.*

Bacteriol.,173, 3078, 1991.

18. Dekker, K., Sugiura, A., Yomagata, H., Sakaguchi, K., and Udaka, S., *Applied Microb. Biotechnol.*, 36, 727, 1992.

19. Moracci, M., La Volpe, A., Pulitzer, J. F., Rossi, M., and Ciaramella, M.,*J. Bacteriol.* 174, 873, 1992.

20. Bachmann, S. L. and McCarthy, A. J., *Appl. and Env. Microbiol.* 57, 8, 2121, 1991.

21. Dobberstein, J. and Emeis, C. C., *Applied Microb. Biotechnol.*, 35, 210, 1991.

22. Biley, P., MacKenzie, C. R., Puls, J., and Schneider, H., *Bio. Technol.*, 4, 731, 1986.

23. Poutanen, K., Raettoe, M., Puls, J., and Viikari, L., *J. Biotechnol.* 6, 49, 1987.

24. Amore, R., and C. P. Hollenberg. *Nucleic Acids Res.* 17,7515, 1989

25. Carrel, H. L., J. P. Gulsker,, V. Burger, F. Manfre, D. Tritsch, and J. P. Biellman. *Proc. Natl. Acad. Sci. USA*, 86, 4440, 1989.

26. Brook and Freeze, 1969. *J. Bacteriol.* 98:289-297.

27. Brock, T. D. 1986. *Thermophiles* pp. 27-28. John Wiley & Sons, Inc., New York.

28. Smith, C. L., J. Econome, A. Schutt, S. Klco, and C. R. Cantor. 1987. A physical map of the *E. coli* K-12 genome. *Science* 236:1448 -1453.

29. Suwanto, A., and S. Kaplan. 1989. Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1. genome:genome size, fragment identificaion, and gene localization. *J. Bacteriol.* 171:5840-5849.

30. Baril, C. C., Richaud, G. Baranton, and I. Saint Girons. 1989. Linear chromosome of *Borrelia burgdorferi*. *Res. Microbiol.* 140:507-516

31. Ferdows, M. S., and A. G. Barbour. 1989. Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent. *Proc. Natl. Acad. Sci. USA* 86:5969-5973.

32. Schwartz, D. C., W. Saffran, J. Welsh, R. Haas, M. Goldenberg, and C. R. Cantor. 1983. New techniques for purifying large DNAs and studying their properties and packing. *Cold Spring Harbor. Symp. Quant. Biol.* 47:189-195.

33. Bancroft, I., C. P. Wolk, and E. V. Oren. 1989. Physical and genetic maps of the genome of the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* 171:5940-5948.
34. Krawiec, S., and M. Riley. 1990. Organization of the bacterial chromosome. *Microbiol. Rev.* 54:502-539.
35. Kwon, S.-T., Kim, J. S., Park, J. H. and Lee D.-S. 1992.
36. Taguchi, H., Yanashita, M., Matsuzawa, H., and Otha, T. 1982. *J. Biochem.* 91, 1343-1348.
37. Matsuzawa, H., Hamaoki, M & Otha, T. 1983. *Agri. Biol. Chem.* 47, 25-28.
38. Feinberg, A. P. and Vogelstein, B., *Anal. Biochem.* 137, 266