

대장균 내에서 불안정한 Minisatellite DNA 영역의 클로닝 및 DNA 염기서열 결정

임선희 · 김재우¹ · 김광섭 · 정윤희 · 윤세련 · 배호정 · 안태진 · 선우양일*

동아대학교 자연과학대학 생물학과, ¹동아대학교병원 임상병리실

진핵생물의 특정 염기배열을 원핵생물 내에서 증폭시킬 때 불안정성이 비교적 빈번히 관찰되어진다. 특히 long inverted repeats나 AT-rich sequences 그리고 Z-DNA와 같은 구조를 지닌 염기배열은 대장균 내에서 매우 불안정하다. 이러한 염기서열은 대장균 내에서 부분적으로 결실되거나 완전히 손실된다. 본 연구실에서 human *SK1* 유전자에 존재하는 몇 개의 tandem repeat (TR)에 대하여 다형성을 조사하였을 때, 어떤 TR 부분은 플라스미드로부터 빈번히 결실되어 그에 대한 염기서열 결정이 어려웠다. 그 결과 이러한 부분은 클로닝될 수 없는 염기서열로 남게 되었다. 본 연구에서는 클로닝이 어려운 두 개의 TR 영역을 저온에서 클로닝하고 nebulizer나 sonicator를 이용하여 두 개의 library를 만들어 DNA 염기서열을 결정하였다. 이러한 연구는 복잡한 고등생물의 게놈연구에서 불안정한 게놈 부분의 염기서열을 결정하는데 도움을 줄 것으로 사료된다.

Key words □ high GC region, minisatellites, unclonable DNA

1990년 인간 유전체 사업에 대한 공식적 논의가 시작된 이래, 최초의 진핵세포인 효모(*Saccharomyces cerevisiae*) 유전체가 1996년에 완성되었고 인간을 비롯한 다양한 생물 종에 대한 유전체 연구가 활발히 진행되고 있다. 이러한 유전체 연구는 DNA sequencer의 발달과 더불어 genome mapping 및 genomic library 제작 기술의 발달로 가능하게 되었다. 유전체의 염기서열 결정에는 주로 bacterial artificial chromosome (BAC)이 사용되었고, 그 외에 P1 artificial chromosome (PAC) 및 yeast artificial chromosome (YAC) library등이 사용되었다(3, 4, 5, 6, 20, 21, 25). 특히 2003년 6월 현재의 기술로는 메울 수 없는 갭(Gap) 약 400 여개를 남기고 인간 게놈 프로젝트의 완성이 발표되었다(The International Human Genome Sequencing Consortium, 2003). 이러한 갭은 인간 전체 유전체의 약 1% 이하로서 30 Mb의 영역에 해당한다(9). 현재 완성된 인간 전체 게놈의 연구로 많은 정보의 사용이 가능하지만, 전체 갭이 메워지지 않은 상태에서는 여전히 매우 중요한 정보의 손실이 예상된다. 또한 인간 질병 연구의 모델계로 사용 가능한 mouse genome project 역시 거의 완료에 가까워지고 있는 시점에서 진정한 게놈의 완성을 위한 새로운 방법의 시도가 요구되어진다.

이러한 팔복할만한 유전체 연구의 성과에도 불구하고 모든 인간 염색체는 몇 개에서 수십 개의 갭을 남기고 있어 현재 시점에서 갭이 없이 완전하게 분석된 염색체는 존재하지 않는다. 그 이유로는 먼저 DNA sequencing에서의 기술적 문제, 많은 반복서열로 인한 염기서열의 배열조합에서 소프트웨어의 문제 등이

존재하지만, 현재 게놈 프로젝트에 사용된 생물학적 방법상의 문제도 제기될 수 있다. 현재 genome project는 BAC, PAC 혹은 YAC library를 이용하여 이루어지고 있으며, 이러한 library는 각 벡터 내에 전체 게놈의 제한 효소단편을 무작위로 클로닝하여, 그 중에서 원하는 영역에 해당하는 클론을 조사하여 분석하는 것으로 많은 시간과 노동력이 소모된다. 또한 어떤 경우에는 두 단편이 완전히 이어지지 않는 sequencing gap이 발생하여, 전체 게놈을 완전히 연결하는데 많은 시간과 노력이 소요된다. 이에 비해 TAR cloning 법은 직선 상 TAR vector의 양쪽 끝에 목적의 DNA 영역 5'와 3'의 염기 배열을 짧게 삽입하여 목적의 게놈 DNA와 함께 효모세포 내에 도입시켜 형질전환체 내에서 vector의 환상분자가 재구성되게 하는 것이다(14, 15, 16). 이러한 재구성은 형질전환과 더불어 직선 상 DNA 양단의 상동한 배열에서 일어나는 상동성 재조합 과정에 의해 일어난다(18). 즉 TAR vector는 적당한 제한효소를 처리하여 양 말단에 상동성 배열이 노출되도록 직선상으로 만든 후, 배양세포나 조직 혹은 혈액으로부터 분리한 소량의 고등생물 genomic DNA와 더불어 효모 내로 형질전환 시킴으로서 원하는 유전자 혹은 특정 염색체 부분을 분리할 수 있다. 그러므로 TAR cloning을 시작하기 위해서는 분리하고자 하는 부분의 양쪽 말단 DNA 염기서열에 대한 정보가 필요하다. 따라서 TAR cloning법은 인간 유전체 연구로 인해 양말단의 염기배열 정보를 알 수 있으므로 그 sequencing의 갭의 크기에 관계없이 손쉽게 메울 수 있어 genome project의 완성에 매우 효과적으로 이용될 수 있다.

최근 염색체 19번의 염기서열 완성을 위해 새로운 몇 개의 library로 클로닝을 시도했으나 메워지지 않은 네 개의 갭이 남게 되어, TAR cloning법으로 효모 내에서 갭에 해당하는 영역을 모

*To whom correspondence should be addressed.
Tel: 82-51-200-5637, Fax: 82-51-200-7269
E-mail: yisunwoo@daunet.donga.ac.kr

두 분리하였다(17). 이러한 결과는 BAC library의 문제점을 TAR cloning 기법으로 해결한 것으로 대장균에서보다 효모에서 안정적인 DNA 단편을 분리한 것이다. 이는 AT-rich 부분이나 긴 inverted-repeat가 대장균 내에서 보다 효모에서 더 안정적이라는 보고와 일치한다(7, 8, 10, 11, 13). 이렇게 효모에서 분리된 DNA를 다시 대장균으로 YAC/BAC의 형태로 삽입하여야 하고, 이 BAC DNA를 이용하여 DNA 염기서열을 결정하였다. 그러나 앞서 설명한 것처럼 BAC의 경우 대장균 내에서 매우 불안정하여 DNA를 안정화될 수 있는 증식조건이 필요하였다. 또한 여러 조건에서도 불안정한 DNA의 경우에는 분리된 YAC을 이용하여 DNA 염기서열을 결정하였다. 이러한 결과, Leem 등(17)에 의해 library를 사용한 방법으로 분리가 어려웠던 겹 부분을 TAR cloning 법에 의해 분리한 후, 안정성에 대한 전반적 문제점을 해결하여 겹 부분의 염기서열을 분석하여 염색체 19 번을 완성하였다(17). 결정된 염기서열의 특성을 조사한 결과 네 개의 겹 모두에서 새로운 유전자가 발견되었으며, 이 부분의 구조적 특성들이 밝혀졌다. 이 부분의 염기서열의 특성은 모두 반복서열이 매우 높은 빈도로 존재하는 것으로 특히 겹 1 부분에서는 인간의 언어능력에 관련된 것으로 밝혀진 SCK1 유전자의 존재가 확인되었다(12). 이 유전자의 인트론 영역에서 12 개의 반복서열 (tandem repeat: minisatellite)이 발견되어, 현재까지 알려진 유전자 중 가장 높은 빈도로 minisatellite가 존재함이 밝혀졌다(15). 이러한 minisatellite나 Alu와 같은 반복서열들이 겹 부분에 많이 존재함으로써 대장균 내에서 증폭되는 동안 DNA 안정성에 문제가 있을 것이 시사되었다. 특히 long inverted repeats나 AT-rich sequences, 그리고 Z-DNA와 같은 구조를 가진 염기서열의 경우는 대장균 내에서 증폭될 때 매우 불안정하다고 보고 되었다(8, 10, 22, 23, 24). 이러한 반복 서열들은 대장균 내에서 증폭되어 질 때 부분적 혹은 전체적으로 결실이 일어난다.

본 연구에서는 인간 19 번 염색체의 말단에 존재하는 SCK1 유전자 내에 존재하는 겹 부분에서 밝혀진 2 개의 minisatellites인 VNTR5 (variable number of tandem repeat)와 TR6 (tandem repeat)을 각각 분리하여 대장균 내에서 DNA 안정성을 조사하였다. 또한 tandem repeats를 PCR (polymerase chain reaction) 방법으로 증폭하여 클로닝하여 염기서열을 결정하였으나, 대장

균 내의 안정성이 문제가 되는 염기서열에 대해서 PCR 산물을 이용하여 직접 그 염기서열을 결정하였다. 특히 반복서열로 인해 염기서열 결정이 매우 어려운 부분에 대해서는 nebulizer나 sonicator를 이용하여 PCR 산물을 단편화 한 다음 작은 library를 만들어 각각의 염기서열을 결정하여 염기서열을 조합하여 전체 염기서열을 결정하였다. 이러한 연구는 복잡한 고등 생물에 존재하는 많은 반복 염기서열을 가진 유전자를 원핵생물을 이용하여 클로닝을 시도할 경우 생기는 문제점을 개선한 것으로 반복 염기서열 결정이나 클로닝이 매우 어려운 DNA 단편의 염기서열 결정에 중요한 방법으로 이용될 것으로 사료된다.

재료 및 방법

사용균주 및 배지

본 실험에 사용된 vector DNA의 증폭에는 대장균 균주인 DH5 α (*supE44*, Δ *lacU169* (Φ 80*lacZ*M15), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*)를 이용하였다. Tandem repeat PCR 산물의 클로닝에는 TOPO shotgun subcloning kit (Invitrogen; Carlsbad, CA, USA)에 포함된 Top10 (*F*-, *mcrA* *D*(*mrr-hsdRMS-mcrBC*), *f80lacZDM15* *DlacX74*, *deoR*, *recA1*, *araD139* *D*(*ara-leu*)7697, *galU*, *GalK*, *rpsL*(*Str*^R), *endA1*, *nupG*) chemically competent *E. coli*를 사용하였다. *E. coli*의 기본 배지 및 배양 조건은 기본적으로 Maniatis 등(19)의 방법을 사용하였다.

SCK1 유전자 내의 tandem repeat 영역의 클로닝

SCK1 유전자의 intron 내에 존재하는 tandem repeat의 확인은 <http://c3.biomath.mssm.edu/trf.html>의 Tandem Repeats Finder program을 이용하였다(1). 그 결과 확인된 반복서열을 PCR 증폭으로 확인하였다. 이렇게 확인된 산물을 TA cloning kit를 이용하여 ligation하고 pCR2-TOPO vector (Invitrogen) 내로 삽입하였다. 이때의 형질전환에는 Top10 균주를 사용하였다. VNTR5와 TR6의 tandem repeat 배열의 클로닝을 수행할 때에 대장균의 배양조건으로 온도를 달리하여 37°C와 30°C에서 비교하였다(17).

Table 1. Primers used in this work

Primers	Sequences	Number of GenBank
SCK1-VNTR5-F	5'-gaa gtc atg tct gaa atg cca cct c	AC138433*
VNTR5-R	5'-cac aat ggg ggc ccg tgt at	
SCK1-TR6-F	5'-ata cac ggc gcc cca ttg tg	AC138433*
TR 6-R	5'-ggg atg ggg aca ggg atg c	
Gap6-F1	5'-ccc tgc atc act ccg caa ca	AY207046*
-F2	5'-tgg ctc cac ctc cct tgc tg	
-R1	5'-cgc ctg ttg ctc ccc aca ct	

*Build 33, April 2003

Tandem repeat 배열의 PCR amplification

본 연구에 사용된 모든 PCR primer들은 Table 1에 나타내었다. Polymorphism을 조사한 PCR 반응에 사용한 genomic DNA는 상호 관련성이 없는 건강한 103 명에서 각각 2 ml의 혈액을 채취하여 QIAGEN genomic-tip 20/G을 사용하여 분리하였다. 또한 VNTR5와 TR6의 PCR 단편 증식에 사용한 게놈 DNA는 human chromosome 19 (LLNL)을 포함하고 있는 human/hamster mono-chromosomal somatic cell hybrid UV5HL9-5B를 배양하여 게놈 DNA를 분리하였다. 이러한 100 ng의 게놈 DNA와 10 μM의 primer DNA를 Takara LA Taq polymerase (Takara Co., Japan)를 이용하여 다음과 같은 조건에서 PCR을 수행하였다(50 mM KCl, 10 mM Tris-HCl, pH 9.0, 3.0 mM MgCl₂, 0.2 mM dTTP, dCTP, dGTP and dATP/50 μl). PCR 반응은 9700 Thermocycler (Perkin-Elmer)를 사용하여 94°C, 2 분간 1 cycle, 그리고 94°C, 30초 → 68°C, 4 분간 처리된 온도변화를 두 단계만 수행하는 반응을 30 cycles 반응시킨 후, 72°C에서 10분간 1 cycle의 연장 반응을 수행하였다.

또한 겹 6에 존재하는 틈새를 맞추기 위하여 먼저 PCR 증폭을 수행하였다. 이 부분은 높은 GC content로 인하여 증폭에 어려움이 있어, 여러 종류의 PCR kit을 사용하여 비교하였다. 그 결과 Thermal Ace Kit (Invitrogen)을 사용하여 PCR 증폭이 이루어졌다. 여기서 얻어진 산물은 다른 PCR 산물의 약 1/3 이하의 소량만이 증폭되어, 여러 번 PCR 증폭을 통하여 얻어진 산물을 Gel extraction kit (Qiagen)을 이용하여 정제하여 염기서열을 결정하였다. PCR 산물은 1.0%의 SeaKem GTG agarose gel을 사용하여 전기영동을 통하여 확인하였다.

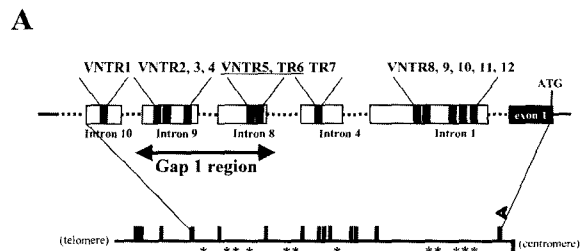
Mini-satellite DNA sequencing

위의 PCR 반응으로 확인된 2.2 kb의 VNTR5 DNA 단편을 micro-sonicator (Fisher)를 50 watt의 출력으로 0, 20, 30, 40, 50, 60 초 간격으로 sonic하여 DNA를 단편화하였다. 4.2 kb의 TR6 단편은 TOPO Shotgun subcloning Kit (Invitrogen; Carlsbad, CA, USA)의 nebulizer를 사용하여 0, 30, 50, 80, 100, 120 초를 각각 처리하여 단편화하였다. 이렇게 단편화한 DNA를 TOPO Shotgun subcloning Kit를 이용하여 ligation하고 pCR4Blunt-TOPO vector (Invitrogen) 내로 삽입하였다. 이때의 형질전환에는 Top10 균주를 사용하였고 삽입된 단편의 안정화를 위하여 30°C에서 하룻밤 배양하였다. 30 개의 형질전환체를 각각 배양하여 각각 DNA를 분리하고 vector 내에 존재하는 EcoRI 효소자리를 이용하여 삽입단편의 크기를 결정하였다. 이때 300-1,000 bp의 삽입 단편을 지닌 형질전환체를 library로 하여 DNA sequencing에 사용하였다. 겹 6의 남은 틈새 부분은 PCR로 증폭하여 PCR purification kit (Qiagen)으로 정제한 후 DNA sequencing에 사용하였다. DNA sequencing에 사용한 VNTR5와 TR5의 primers로는 pCR4Blunt-TOPO vector 내에 존재하는 T3와 T7 primer를 사용하였고, 겹6에 대해서는 PCR 증폭에 사용한 primer를 사용하였다. DNA sequencing 반응에는 ABI Big dye sequencing kit ver 1.0과 ver 3.0을 비교하여 사용하였다.

결과 및 고찰

Human SCK1/SLI 유전자의 intron 8에 존재하는 minisatellites의 다형성 분석

본 연구실에서 인간 염색체 19 번의 p-arm 뱀 끝으로부터 겹 1, 겹 2, 겹 3 그리고 겹 6라 명명하여 네 개의 겹을 메우는 연구를 수행하였다 (17). 이 과정에서 겹1 영역에서 human SCK1/SLI 유전자가 발견되었으며 이 유전자의 intron 부분에서 12 개의 tandem repeat 부분이 발견되었다(Fig. 1A). 이 중 intron 8에 위치하는 VNTR5와 TR6는 겹에 정확히 위치하여 이 반복서열이 클로닝과 염기서열 결정에 장애가 되었을 가능성을 보여준다. 그러므로 본 연구에서는 약 12 kb인 전체 겹 중에서 tandem repeat 부분만을 따로 증폭하여 대장균 내에서의 안정성을 조사하였다. 본 연구와 유사한 시기에 NCBI에 등록된 AC138433의 fosmid에서는 이 겹 부분이 약 10 kb로 등록되어 본 연구의 결과와 차이를 나타내었다. 이러한 차이는 tandem repeat 부분의 다형성이나 tandem repeat 부분의 염기서열 일부가 결손된 fosmid의 DNA sequencing 결과일 가능성이 있다. 그러므로 겹



B

VNTR	Repeat unit	Length of repeats	Polymorphism	Heterozygosity
VNTR5	AGAGGAGGCC GGGCAGATAG ATGGCGCTTC ATCGTGAGTG AGATCGTGAG TGAGAGAT	58	yes	0.94
TR6	GTGAGTGAGATC	12	no	0

$$\text{Heterozygosity (h)} = 1 - \sum_{i=1}^k x_i^2$$

Fig. 1. Minisatellites in SCK1/SLI. (A) A schematic diagram of the sequence spanning the SCK1/SLI gene is shown. Exons are represented by boxes either above (top strand) or below (bottom strand) the line. Thirteen putative exons, coding for the Sck protein, were identified by BLAST analysis. The approximate positions of minisatellites, detected by the Tandem Repeats Finder Program (1), are indicated by asterisks. On top, a blown up portion of the insert shows the relative positions within introns of the twelve minisatellites. (B) The sequences of two minisatellite repeat units are shown. Minisatellite 5 (VNTR5) is polymorphic whereas minisatellite 6 (TR6) is mono-morphic in the population sample studied. Heterozygosity is the statistic used to evaluate the informativeness of a genetic marker, and is calculated by this formula, where x_i is the allele frequency of the ith allele at a particular locus. The degree of polymorphism is generally increased by large numbers of alleles occurring in equal frequencies.

안에 존재하는 tandem repeat 부위인 VNTR5와 TR6의 다형성을 103 명의 혈액으로부터 채취한 게놈 DNA를 이용하여 확인하였다. 이러한 혈액 DNA를 이용한 polymorphism 조사에서 먼저 VNTR5는 27개의 allele가 확인되었으며, 그 중에서 46.7 번의 반복을 나타내는 2.65 kb의 allele가 가장 높은 빈도로 나타났다 (Fig. 2). 이 VNTR5에는 많은 allele가 존재하여 heterozygosity가 0.94로 매우 높게 나타났다(Fig. 1B). Heterozygosity는 0에서 1의 범위로 표시할 때 0은 모든 개체에서 동일한 패턴을 나타내는 것이고 1은 모든 개체가 서로 다른 것으로 볼 수 있다(2). 그러므로 0.94의 값을 나타내는 VNTR5의 경우는 매우 높은 것으로 개인의 DNA typing의 인자로도 사용할 수 있다. 이에 비해 TR6는 monopolymorphic allele로서 조사한 103 명의 혈액샘플에서 동일한 패턴을 보여주는 단형성 반복서열로 밝혀졌다(Fig. 3 A, B).

대장균 내에서 불안정한 반복서열의 확인

앞서 서술한 것 같이 유사한 시기에 NCBI에 등록된 AC138433의 fosmid clone과 본 연구에서 클론된 YAC clone (AY345879)의 이 갭 1 부분의 부분의 분석 결과는 2 kb 정도 크기의 차이를 나타내었다. 이 차이의 원인이 polymorphism 때문인가를 확인하기 위해 다형성을 조사하였다. 그 결과 Fig. 2에서 보여주는 것과 같이 VNTR5의 tandem repeat 부분은 다형성

에 의한 차이를 나타내었다. 그러나 이 부분의 다형성의 차이에서도 fosmid 클론(3.0 kb)에 비해 YAC 클론이 오히려 더 작은 allele (2.2 kb)를 나타내었다. 그러므로 이 차이는 VNTR5 부분 위 polymorphism으로 인한 것이 아니라 TR6의 결과에서 보여주는 영역이 fosmid에서 결손되었다는 것을 알 수 있다(Fig. 3A). 즉 fosmid에서 보여주는 tandem repeat의 크기는 19,268-19,430에 인 위치해 해당하는 17.7 회의 반복을 지니는 173 bp에 불과하나 PCR 증폭을 통하여 조사한 결과 실제 길이는 293 회의 반복을 지닌 약 4.2 kb에 해당하였다(Fig. 3A). 이 결과는 103 명의 혈액샘플에서 모두 동일하게 확인되었으므로 polymorphism으로 보기는 매우 어렵다. 그러므로 본 연구에서 이 부분이 대장균 내에서 증폭이 어려운 반복서열로 간주하고 이 영역을 따로 PCR로 증폭하여 대장균 내에서 VNTR5와 TR6 부분의 클론닝을 시도하였다.

먼저 VNTR5 부분과 TR6 부분의 PCR amplication을 수행하기 위하여 본 연구에서 분리된 YAC 클론과 human chromosome 19 (LLNL)을 포함하고 있는 human/hamster monochromosomal somatic cell hybrid UV5HL9-5B의 게놈 DNA를 사용하여 증폭하였다. 그 결과 YAC 클론과 UV5HL9-5B의 게놈 DNA에서 동일한 크기의 PCR 산물이 확인되어(자료 미제시), YAC 클론에서는 fosmid에서 보여준 것과 같은 결손이 없었다. TA cloning kit을 이용하여 이러한 PCR 산물의 클론닝을 시도하였다. 이렇게 얻어진 형질전환체 20 개를 무작위로 선택한 후 37°C에서 배양하여 DNA를 분리하여 VNTR5와 TR6의 클론닝을 확인하였다. 반복서열을 지니지 않는 효모의 LEU2 유전자의 PCR 산물은 약 95%의 비율로 클론닝이 이루어지나, VNTR5는 약 10% 이하의 클론닝 빈도를 나타냈고 그 나머지는 결손된 형태로 작은 산물만이 삽입되었다. 더욱이 TR6는 20 개의 형질전환체 모두에서 완전한 길이의 DNA가 발견되지 않았다. 이러한 결과는 대장균

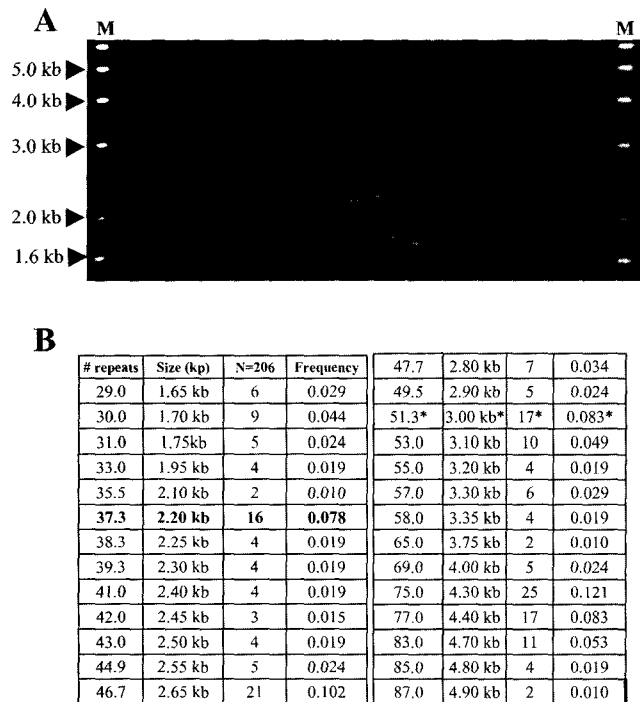


Fig. 2. Polymorphism of VNTR5. (A) The lengths of VNTRs were analyzed in DNA from 103 individuals by PCR with a diagnostic pair of primers (Table 1). (B) Allele frequency, size of PCR products, and repeat number are shown. In bold is minisatellite allele identified in a YAC clone. Asterisks show the minsatellite allele in a fosmid clone. M, 1 kb size marker.

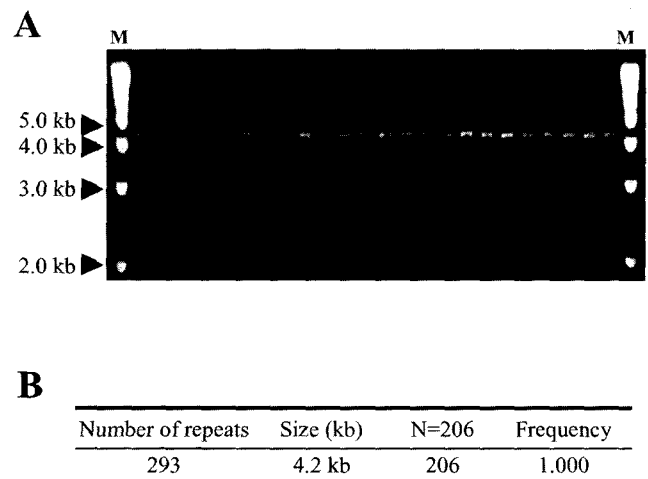


Fig. 3. Monomorphism of TR6. (A) The lengths of TR were analyzed in DNA from 103 individuals by PCR with a diagnostic pair of primers (Table 1). (B) Allele frequency, size of PCR products, and repeat number are shown. M, 1 kb size marker.

에서 이러한 반복서열이 매우 불안정함을 의미한다. 그러므로 불안정한 BAC DNA를 분리하고자할 때 사용하는 배양조건인 30 °C에서 다시 20 개의 형질전환체를 조사하였다. 그 결과, VNTR5는 25%의 클로닝 빈도를 나타내었고 TR6 부분도 또한 약 5% 빈도로 클로닝이 이루어졌다. 이렇게 클로닝된 TR6 단편은 같은 조건에서 계속 계대배양을 하여도 안정적으로 증식되었다. 이러한 결과는 BAC과 같이 큰 분자에서는 더욱 필수적인 것으로 확인되었으므로(13, 17), BAC library와 같은 library의 구축에도 30°C와 같이 다소 낮은 온도가 DNA 안정성에 도움을 주는 것으로 사료된다.

반복서열 및 GC-rich 부분의 염기서열 결정

위에서 클로닝된 DNA의 염기서열을 결정하여 실제 TR6가 모두 반복서열을 나타내는지 확인하였다. VNTR5와 TR6는 반복서열 부분이므로 바깥쪽의 배열을 읽고나면 그 다음부터는 염기서열 결정이 매우 어렵다. 그러므로 클로닝된 단편을 *Eco*R1 효소로 분리하여 정제한 후, 이를 작은 library를 구축하기 위하여 TR6는 nebulizer (Invitrogen)를 사용하고 VNTR5는 sonicator를 이용

하여 DNA를 단편화하였다(재료 및 방법 참조, Fig. 4). 이렇게 만들어진 library로부터 정제한 DNA를 사용하여 염기서열을 결정하였다. DNA sequencing 반응에는 ABI Big dye sequencing kit ver 1.0과 ver 3.0을 비교한 결과, GC content가 높고 반복서열이 많은 경우에는 ver 3.0에서 현저히 좋은 결과를 나타내었다. 이렇게 결정된 VNTR5의 염기서열은 NCBI에 등록된 AC138433의 fosmid clone과 본 연구실에서 분리된 YAC clone (AY345879)과 동일한 염기서열을 나타내었다. 이에 비해 TR6 부분은 Fig. 5에서 보여주는 것과 같이 PCR product와 동일한 크기의 반복서열의 존재가 확인되었다. 이는 AC138433의 fosmid clone과는 약 4 kb의 반복서열 크기의 차이를 나타내는 것이다. 그러므로 YAC 클론은 대장균에서 분리한 BAC 클론에 비해 반복서열에 대한 안정성이 높은 것으로 사료된다. 이러한 결과로부터 TAR cloning을 통한 YAC의 분리는 대장균 내에서의 증폭과정을 거치지 않으므로 유전체 연구에 또 다른 새로운 방법으로 그 효용성이 제기될 수 있다.

염색체 19번에 존재하는 겹 6는 TAR cloning 법으로 분리한 후, retrofitting 과정을 거쳐 저온에서 완전한 크기의 BAC의 형

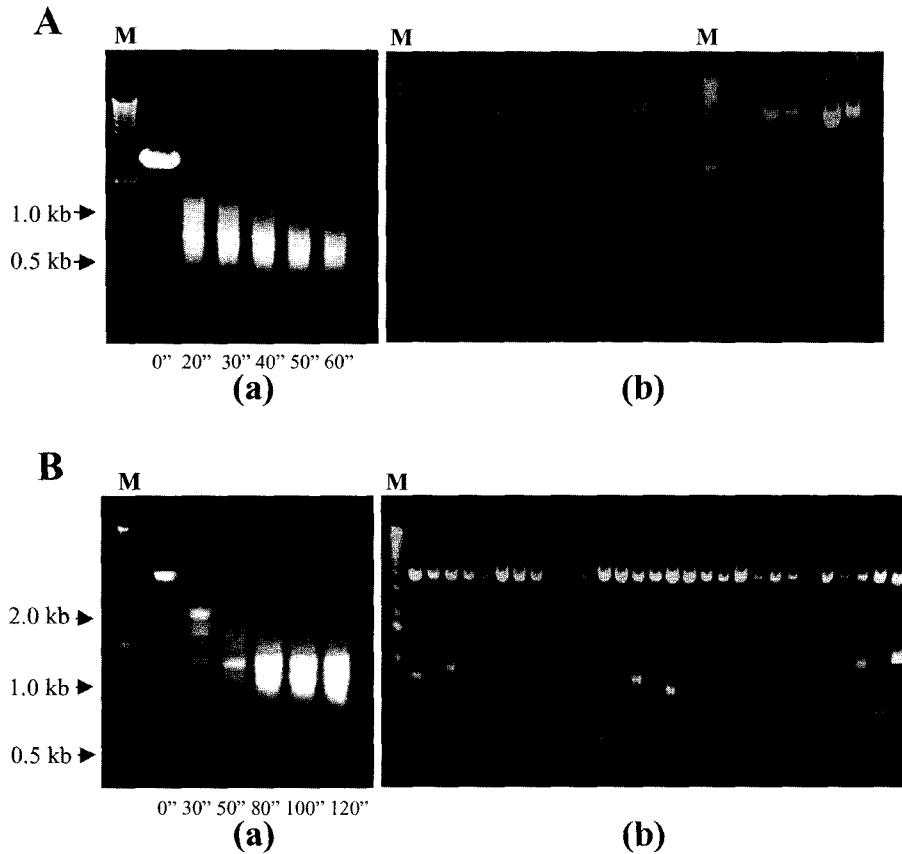


Fig. 4. Construction of small libraries for VNTR5 and TR6. (A) (a) showed the agarose gel electrophoresis of VNTR5 DNA samples sheared by a sonicator at different times. After shearing, the DNA was ligated into pCR4 Blunt-TOPO vector to form a library. Twenty-one clones from the library were digested with *Eco*R1 enzyme and analysis of insert size on an agarose gel (b). (B) (a) showed the agarose gel electrophoresis of TR6 DNA samples sheared by a nebulizer at different times. Twenty-nine clones from the library were digested with *Eco*R1 enzyme and analysis of insert size on an agarose gel (b). M, 1 kb size marker.

*CTTGATTTGGAAGTCAGGCTTTTCCACCAGGGCAGAGACCTTGTGTTTGTTC
 ACTGGTGTATCCCCAGCTCCTGGCCCACAGAAGGTGCTCAAGAAATATGTGT
 CGAATGAATGAGAGTCTGGGCTTCTCCCTCTCCCTGCATCACTCCGCAACAG
 CCGGGCCCGGCTCTGGCGGACTTTCTAGCCCCTGCCACCTTCAGCTACG
 CCGGCGGGCCGACCTGGGTCTCTGACCACCACGCCCTATCCCTGCAGCCC
 CTCCTCCTCGGATGGCTCCACCTCCCTTGTCTGGGCCCCGCCCCAGGCTCGC
 AGTCTCCCGCCCCGAGAGCACACGCACCTGGGCACGTTGATATCCCGGTGC
 GCCCGGGGCGCACGGGACGCCTTGTGAAGGCCACTTTGGCGCCGACGCC
 CACGGCCAGGGCGAAGCTGATGACCCCGCACACGACGTAGGCTGTGCTGC
 CCCCAGGGCCCTCGCCCCCGGGCCCCCGGCACCCCCAGTCCGGC
 CCCCTTACCACCCGGCCTGGCCGGGCCCTGGCCCCCGCCCCGCA
 CCCCCAGCGCCCCCGGGCCCCAGCTAGCGGGCGGGCGGCGTGGTGG
 CCCAACCGGGCGTGTCTGTAATTGGAGCAGGAGGCCTGCGCCAGGC
 CATGTGACGGTGTCTACAGCAAAAGCGGTAGTGGCAGGTGCCACA
 GCAGAAACGGTAGGAACCGGTGCTCCAGTTGAAGGTGGCTTCGTA
 CTGGCCCATGACATCGTAGTACCCGTGGCAGAGCTCGGCGGGAGG
 GGGCGCCCGGGCCGCGCCCGCCCCCGCCCGGGGGCCGGTCTCT
 GGTGCCGTTGGCGCCTGGGCTCGCCGCGCCCCCGCCGCCCGTAC
 CGCCCCGGTCAGGCGCCGCAGGTGCGCCAGCAGGGCGGGCAGCG
 GGCCCCGCGGGCTCGGCGCTCGTGGCGTTGGACGGGCGCGCCCTGG
 CCTGGCCGGCGCTAGAGGCCAGGAGTACGAGGAGCAGGAGGGCCG
 GCATGGGGCTTGCAGGGGGTTCGCACTGGGCCGCCAGGCTGCGGGG
 AGAACGAGAGCAGAGGTTGGAGCGCTGGGCGGGAGAGAAACGGTC
 AGAGGGTGTAGAAACGTAGAGATGGGTGGGCAGAGAGGCTGGGGAGGGC
 GGGCGGGGGCGGCAGAGTGTGGGGAGCAACAGGCGCCAGGAGGCAGGA
 AGGGCGGAGGGAGCTGAGCTGCGTGGGCCCAAGTTGGCCGCCACGCG
 CGGTGGGCGACGATGGGAGAGTAATGGAAACGCGCCCGGGCACGGCTGG
 GGGAGAGAAATGAGCATGGGCGGACAGAACCGGGATCAGGCGAGAGATGA
 ATGCGGGAGAAGCGATTGGGAGGGTTCGACAGACACCAGAGGGGTGTGAC*

Fig. 6. DNA sequence of the Gap6-final Gap DNA fragment amplified by PCR. 644 bp DNA sequence of gap are shown in bold. Primers are shown in italic and underlined.

감사의 글

이 논문은 2002학년도 동아대학교 공모과제 교내 학술연구비에 의하여 연구되었음.

참고문헌

1. Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573-580.
2. Birren, B., E.D. Green, P. Hieter, S. Klapholz, R.M. Myers, H. Riethman, and J. Roskams. 1999. *Genome analysis*, vol. 4, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
3. Burke, D.T., G.F. Carle, and M.V. Olson. 1987. Cloning of large segments of DNA into yeast by means of artificial chromosome vectors. *Science* 236, 806-812.
4. Carrano, A.V., P.J. de Jong, E. Branscomb, T. Slezak, and B.W. Watkins. 1989. Constructing chromosome- and region-specific cosmid maps of the human genome. *Genome* 31, 1059-1065.
5. Chartier, F.L., J.T. Keer, M.J. Sutcliffe, D.A. Henriques, P. Mileham, and S.D. Brown. 1992. Construction of a mouse yeast artificial chromosome library in a recombination-deficient strain of yeast. *Nature Genet.* 1, 132-136.

6. Clarke, L., and J. Carbon. 1980. Isolation of a yeast centromere and construction of a functional small circular chromosome. *Nature* 287, 504-509.
7. Gardner, M.J., S.J. Shallom, J.M. Carlton, S.L. Salzberg, V. Nene, A. Shoaibi, A. Ciecko, J. Lynn, M. Rizzo, B. Weaver, B. Jarrahi, M. Brenner, B. Parvizi, L. Tallon, A. Moazzez, D. Granger, C. Fujii, C. Hansen, J. Pederson, T. Feldblyum, J. Peterson, B. Suh, S. Angiuoli, M. Perte, J. Allen, J. Selengut, O. White, L.M. Cummings, H.O. Smith, M.D. Adams, J.C. Venter, D.J. Carucci, S.L. Hoffman, C.M. Fraser. 2002. Sequence of *Plasmodium falciparum* chromosomes 2, 10, 11 and 14. *Nature* 419, 531-534.
8. Glockner, G., L. Eichinger, K. Szafranski, J.A. Pachebat, A.T. Bankier, P.H. Dear, R. Lehmann, C. Baumgart, G. Parra, J.F. Abril, R. Guigo, K. Kumpf, B. Tunggal, E. Cox, M.A. Quail, M. Platzer, A. Rosenthal, A.A. Noegel, B.G. Barrell, M.A. Rajandream, J.G. Williams, R.R. Kay, A. Kuspa, R. Gibbs, R. Sugang, D. Muzny, B. Desany, K. Zeng, B. Zhu, P. de Jong, T. Dinger, G. Gerisch, P. Philippsen, M. Schleicher, S.C. Schuster, T. Winckler. 2002. Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. *Nature* 418, 79-85.
9. Grimwood, J., and J. Schmutz. 2003. Genomics: six in seventh. *Nature* 425, 775-776.
10. Hagan, C.E., and G.J. Warren. 1982. Lethality of palindromic DNA and its use in selection of recombinant plasmids. *Gene* 19,

- 147-151.
11. Hayashi, Y., E. Heard, and M. Fried. 1993. A large inverted duplicated DNA region associated with an amplified oncogene is stably maintained in a YAC. *Hum. Mol. Genet.* 2, 133-138.
 12. Kojima, T., Y. Yoshikawa, S. Takada, M. Sato, T. Nakamura, N. Takahashi, N.G. Copeland, D.J. Gilbert, N.A. Jenkins, and N. Mori. 2001. Genomic organization of the Shc-related phosphotyrosine adapters and characterization of the full-length Sck/ShcB: specific association of p68-Sck/ShcB. *Biochem. Biophys. Res. Commun.* 284, 1039-1047.
 13. Kouprina, N., S.-H., Leem, G., Solomon, A., Ly, M., Koriabine, J., Otstot, E., Pak, A., Dutra, S., Zhao, J. C., Barrett, and V. Larionov. 2003. Segments missing from the draft human genome sequence can be isolated by TAR cloning in yeast. *EMBO Reports* 4, 257-262.
 14. Larionov, V., N. Kouprina, N. Nikolaishvili, and M.A. Resnick. 1994. Recombination during transformation as a source of chimeric mammalian artificial chromosomes in yeast (YACs). *Nucleic Acids Res.* 22, 4154-4162.
 15. Larionov, V., N. Kouprina, J. Graves, X.-N. Chen, J.R. Korenberg, and M.A. Resnick. 1996. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. U.S.A.* 93, 491-496.
 16. Larionov, V. 1998. Direct isolation of specific chromosomal regions and entire genes by TAR cloning. *Genet. Eng. Princ. Methods* 21, 141-159.
 17. Leem, S.-H., N. Kouprina, J. Grimwood, J.-H. Kim, M., Mullochandov, Y.-H., Yoon, J.-Y., Chae, J., Morgan, S., Lucas, P., Richardson, C., Detter, T., Glavina, E., Rubin, J. C. Barrett, and V. Larionov. 2004. Closing the gaps on human chromosome 19 revealed genes with a high density of repetitive tandemly arrayed elements. *Genome Res.* 14, 239-246.
 18. Ma, H., S. Kunes, P.J. Schatz, and D. Botstein. 1987. Plasmid construction by homologous recombination in yeast. *Gene* 58, 201-216.
 19. Maniatis T., E.F. Fritsch, and J. Sanbrook. 1997. Molecular cloning a laboratory manual, Cold spring Harbor Laboratory, Cold Spring Harbor Press, N. Y.
 20. Murray, A.W., and J.W. Szostak. 1983. Construction of artificial chromosomes in yeast. *Nature* 305, 189-193.
 21. Osoegawa, K., A.G. Mammoser, C. Wu, E. Frengen, C. Zeng, J.J. Catanese, and P.J. de Jong. 2001. A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res.* 11, 493-496.
 22. Pan X, and D.R. Leach. 2000. The roles of mutS, sbcCD and recA in the propagation of TGG repeats in *Escherichia coli*. *Nucleic Acids Res.* 28, 3178-3184.
 23. Razin, S.V., E.S. Ioudinkova, E. Trifonov, and K. Scherrer. 2001. Non-clonability correlates with genome instability: a case of unique DNA region. *J. Mol. Biol.* 307, 481-486.
 24. Schroth, G.P., and P.S. Ho. 1995. Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucleic Acids Res.* 23, 1977-1983.
 25. Shizuya, H., B. Birren, U.J. Kim, V. Mancine, T. Slepak, Y. Tachiri, and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase -pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. U.S.A.* 86, 8794-8797.

(Received March 25, 2004/Accepted May 4, 2004)

ABSTRACT: Cloning and DNA Sequencing for Unstable Minisatellites DNA Regions in *E. coli*.

Sun-Hee Leem, Jae-Woo Kim¹, Kwang-Sup Kim, Yun Hee Jeong, Se-Lyun Yoon, Ho-Jeong Bae, Tae-Jin Ahn, and Yangil Sunwoo*.

(Department of Biology, Dong-A University, ¹Department of Clinical Pathology, Dong-A University Hospital, Busan 604-714, Korea)

Instability of some eukaryotic sequence propagated in prokaryotic hosts is a frequently observed phenomenon. It is well documented that long inverted repeats, AT-rich sequences with structures like Z-DNA are extremely unstable in *E. coli*. These sequences may either be under-represented or even lost when cloned in *E. coli*. When we analyzed the polymorphic pattern for several tandem repeat (TR) in human *SCK1* gene, we found some TR regions were frequently deleted from plasmids and had difficult problem for their sequencing. These regions may result in non-clonability of the DNA sequence. Here we have cloned two difficult TR regions under low temperature and made two library for DNA sequencing using a nebulizer or sonicator. This study will help to determine the unstable genomic elements in complex mammalian genome.