임상분리 결핵 균주간의
유전적 다형성

연세대학교 대학원
의 과 학 과
김 지 원

임상분리 결핵 균주간의
유전적 다형성

연세대학교 대학원
의 과 학 과
김 지 원

임상분리 결핵 균주간의 유전적 다형성

지도교수 조 상 래

이 논문을 석사 학위논문으로 제출함

2005년 12월 일

연세대학교 대학원

의 과 학 과

김 지 원

김지원의 석사 학위논문을 인준함

심사위원 _____ 인

심사위원 _____ 인

심사위원 _____ 인

연세대학교 대학원

2005년 12월 일

감사의 글

어느덧 2년이라는 시간이 흘러 석사학위 과정이 끝나게 되었습니다. 아직 부족한 점도 많고, 더 배워야 할 것도 많은데 벌써 졸업이라고 생각하니 감회가 새롭습니다.

우선 항상 학자로서의 모습을 보여주시며, 조언을 아끼지 않으셨던 조상래 교수님께 깊은 감사를 드립니다. 그리고 처음부터 마지막까지 저를 지도해주시면서, 스스로 계획을 세우고 실험을 진행해 나갈 수 있는 능력을 가지게 해주신 송택선 교수님께 감사를 드립니다. 바쁘신 와중에도 논문심사를 맡아주셔서 조언을 해주신 김준명 교수님, 다양한 시각에서 생각할 수 있도록 지도해주신 미생물학 교실의 교수님들, 그리고 항상 따뜻하게 대해주셨던 이혜영 교수님, 조성애 교수님께도 감사의 말씀을 전합니다.

누구보다도 더 가까운 곳에서 저에게 힘이 되어주었던 우리 SNC 식구들, 항상 조언을 아끼지 않으셨던 한미영 선생님, 지영이, 지연 선생님, 재식 선생님, 원경 선생님, 보람 선생님, 영미 선생님, 은계 선생님, 은희 선생님, 주영 선생님, 정연 선생님, 민경이, 미진 선생님, 그리고 멀리 있으면서도 염려해주셨던 방선생님, 연선생님, 민선씨에게도 감사드립니다.

항상 힘이 되어주었던 소중한 친구 지효, 미정이, 로사, 보나, 한양 7공주 멤버들, 졸업하고서도 같이 고민해주고 조언해준 선배이자 친구인 승은이, 후배이지만 친구같았던 아름이, 친자매처럼 잘 따라준 선희에게 감사의 말을 전하고 싶습니다.

마지막으로 항상 저를 믿어주시고 지지해주셨던 우리 가족, 아빠, 엄마, 큰오빠, 새언니, 작은오빠, 귀여운 조카들 그리고 매일 혼자 있지만 건강하게 잘 자라준 레이에게 감사의 마음과 함께 평소에 잘 하지 못하는 말, 정말 사랑한다는 말씀을 드리고 싶습니다.

2005년 12월

김지원 올림

차 례

그림 및 표 차례 iii
국문요약 1

I. 서 론 3
1. 결핵의 현황 3
2. 결핵균 유전체 3
3. 임상분리 균주의 발병력 차이 4
4. 결핵 균주들의 유전적 다형성 6
5. 연구의 목적 7
II. 재료 및 방법 8
1. 실험재료 8
가. 사용 균주 8
나. 효소 및 시약 8
2. 실험방법 8
가. 결핵균 배양 8
나. 결핵균에서 genomic DNA 분리 8
다. 연구대상 부위의 선정 및 PCR primer의 고안 9
라. PCR 10
마. 염기서열 확인 10
바. 염기서열 분석 10
III. 결 과 14
1. K 균주와 H37Rv 균주의 염기서열 비교 14
2. K 균주와 Beijing 균주의 결손 부위 비교 14
3. K 균주와 CDC1551 균주의 결손 부위 비교 30

4. 임상분리 균주를 대상으로 PCR 수행 31
가. PCR primer의 고안 31
나. PCR template의 선정 32
다. 결손 혹은 삽입 부위에서 임상분리 균주들의 PCR 수행 33
라. 결손 부위에서 PCR 결과 분석 33
마. 삽입 부위에서 PCR 결과 분석 36
5. 결손 혹은 삽입 부위의 염기서열 분석 36
IV. 고 찰 43
V. 결 론 50

참고문헌 52

영문요약 59

그림 차례

그림 1. 실험에서 사용된 균주의 RFLP 결과 12
그림 2. 결손 부위에서 임상분리 균주간의 다형성 34
그림 3. 삽입 부위에서 임상분리 균주간의 다형성 35
그림 4. Del_21의 PCR 결과 39
그림 5. Del_21의 염기서열 분석과 Rv1243c의 아미노산서열 분석 40
그림 6. Ins_10, Ins_11의 PCR 결과 41
그림 7. Ins_10, Ins_11의 염기서열 분석과 Rv0746의 아미노산서열 분석 42

표 차례

표 1. PCR에 사용된 primer들 13
표 2. K 균주에서 일어난 결손과 삽입 15
표 3. K 균주에서 확인된 결손 부위 16
표 4. K 균주에서 확인된 삽입 부위 23
표 5. 결손에 의해 영향을 받은 ORF 28
표 6. 삽입에 의해 영향을 받은 ORF 29

국문요약

임상분리 결핵 균주간의 유전적 다형성

결핵은 결핵균 (*Mycobacterium tuberculosis*)에 의하여 발생하는 만성 세균감염증으로 세계보건기구 (World Health Organization; WHO)에서 'global emergency'로 선포하였을 만큼 세계적인 보건 문제이다. *M. tuberculosis* complex에 속하는 균주들은 전체 뉴클레오타이드 서열이 99.9%의 유사성을 보이고 있음에도 불구하고 숙주 특이성과 발병력에서 많은 차이를 보이고 있다. 최근 환자에서 분리된 결핵균 임상분리 균주들에서 균주에 따라 병원성과 관련된 표현형들이 차이가 있음이 밝혀지고 있는데 이와 같은 발병력의 차이를 설명할 수 있는 분자유전학적인 연구는 아직 충분하지 않다.

본 연구는 결핵균 임상분리 균주들에서 보이는 발병력의 차이가 각 균주들이 가지고 있는 유전 정보의 차이에 기인할 것이라는 가정에 따라 몇 가지 임상분리 균주들의 유전적 다형성을 분석하였다. 이를 위하여 최근에 밝혀진 임상분리 결핵균 K 균주의 유전체 염기서열을 결핵균 표준균주 H37Rv의 염기서열과 비교분석하여 이 균주들 사이에 차이 (결손 혹은 삽입)를 보이는 염기서열들을 연구 대상으로 선정하였고 국내 및 국외에서 분리된 임상분리 균주들을 대상으로 해당 DNA 절편을 PCR로 증폭하고 염기서열을 결정하여 각 부위의 유전적 다형성을 규명하였다.

K 균주와 H37Rv 균주의 전체 염기서열을 비교하여 K 균주에서 153 개의 결손과 124 개의 삽입이 일어났음을 확인하였다. 결손과 삽입에 의해서 각 96개와 51개의 open reading frame (ORF)이 영향을 받았으며 이들 대부분은 PE, PPE family, hypothetical protein, transposase을 coding하고 있었다. 각 부위에서 균주간의 다형성을 확인하기 위해 54 부위의 결손과 58 부위의 삽입을 대상

으로 40개의 임상분리 균주에서 PCR을 수행하였다. 임상분리 균주 A, a, b, d, 13, 18은 전반적으로 H37Rv와 유사한 삽입 및 결손을 보이는 반면 나머지 임상분리 균주들은 각 부위에 따라서 다양한 크기의 결손과 삽입을 가지고 있음을 확인하였다. Del_3, Del_16, Del_18 부위는 K 균주에서만 결손되었고, Del_29, Del_36은 모든 K family에 속하는 균주들에서 결손된 것으로 확인되었다. 따라서 이들 부위에 위치하고 있는 유전자들이 K 균주에 특이적인 발병력에 기여할 것으로 예상된다. 결손과 삽입 중에서 21 부위를 선정하여 염기서열을 결정하여 분석한 결과 PCR 산물의 크기가 같을 경우 염기서열이 동일하였고 PCR 산물의 크기가 다를 경우에 결손이나 삽입에 의하여 염기서열의 변화가 있었으며, 이로 인하여 아미노산 서열에도 변화가 있음을 확인하였다.

지금까지의 비교유전체 연구는 DNA microarray를 기반으로 하여 수행되었기 때문에 이미 알려져 있는 염기서열에서 결손된 유전자들을 탐색하는 것에는 유용하지만 근본적으로 새롭게 삽입된 유전자들에 대한 정보를 제공할 수는 없다. PCR 및 sequencing을 이용하는 본 연구는 결손이 일어난 염기서열에 대한 정보 뿐 아니라 삽입된 염기서열의 탐색도 가능하기 때문에 기존의 연구결과보다 광범위한 결과를 축적할 수 있었다. 이러한 결과는 각 임상분리 균주에서 발견되는 발병력의 차이에 대한 유전적인 원인을 규명할 수 있는 기초 정보들을 제공함으로써 결핵균의 발병인자를 탐색하고 병인기전을 규명하는데 기반이 될 수 있을 것이다.

핵심되는 말 : 결핵균, 임상분리 균주, 유전적 다형성

임상분리 결핵 균주간의 유전적 다형성

<지도교수 **조 상 래**>

연세대학교 대학원 의과학과

김 지 원

I. 서론

1. 결핵의 현황

결핵은 결핵균 (*Mycobacterium tuberculosis*)에 의해서 발생하는 만성 세균감염증으로, 1993년에 세계보건기구 (World Health Organization; WHO)에서 'global emergency'로 선포하였을 만큼 세계적인 보건 문제이다. 최근의 WHO 추계에 의하면, 세계 인구의 1/3이 결핵에 감염되어 있고, 매년 800만 명의 새로운 결핵 환자가 발생하며, 매년 200만 명이 결핵으로 사망한다고 한다[1]. 우리나라에서도 전 인구의 약 1/3 이상이 결핵에 감염되어 있는 것으로 추정되며, 매년 약 12만 명의 새로운 결핵 환자가 발생하고 있다[2].

2. 결핵균 유전체

결핵균은 크기가 0.2~0.5 × 1~4 ㎛ 정도 되는 호기성의 그램 양성균으로 아포를 형성하지 않고, 편모가 없으며, 세포벽의 많은 지질 때문에 항산성을 띤다. 분열 증식한 균은 집락을 이루는데 균이 서로 엉겨 붙어 새끼줄처럼 보이는 균삭을 이루고, 이 균삭의 형태로 결핵을 유발하게 된다. 결핵균의 유전체는 분자량 약 2.5 × 10^9의 크기로 *Mycobacterium* 속의 다른 균주들의 유전체 (3.1~

4.5×10^9)보다 작다. 이는 많은 유전자의 결손을 거쳐 필수적인 유전자만을 남긴 채, 숙주에 기생하도록 진화했기 때문인 것으로 보인다.

결핵균에 대한 최초의 유전체 분석은 표준 균주인 H37Rv를 대상으로 영국의 Sanger 센터에서 이루어졌다[3] (GenBank accession no., NC_000962). 결핵균의 유전체는 4,411,519 염기쌍으로 이루어졌고, G+C 함량이 65.6%로 다른 세균에 비해 높은 편이다. 전체 유전체에서 지질 대사에 관련된 유전자들이 매우 높은 비중을 차지하는 것으로 알려졌고, 특징적으로 전체 유전자의 10%를 차지하는 독특한 반복되는 아미노산 서열을 갖는 단백질의 유전자군이 발견되었다. 이들을 PE, PPE family라 부르는데, 이들은 G+C 함량이 매우 높은 유전자들이며, 여기에서 생성된 단백질들은 결핵균의 면역적인 특이성과 다양한 항원성을 제공하는 것으로 알려졌다.

결핵균의 유전체에는 균주 특이적인 동일한 DNA 절편이 여러 개 존재하는 것이 밝혀져 있는데, 그 중 일부는 mobile genetic element인 insertion sequence (IS)로 알려졌다[4,5]. 이러한 IS element들은 결핵 균주에 따라 유전체의 서로 다른 위치에 불규칙하게 삽입되어 있기 때문에, 적절한 제한효소로 결핵균의 유전체를 절단하였을 때 나타나는 IS element를 포함하는 DNA 절편의 길이의 다양성을 이용하여 각 결핵 균주를 구분할 수 있는 유용한 정보를 제공한다[6]. 이 가운데 *M. tuberculosis*에만 존재한다고 알려진 IS*6110*은 결핵의 진단에 이용된다[4,5].

3. 임상분리 균주의 발병력 차이

균주가 감염을 유발시킬 수 있는 능력을 발병력이라고 하는데, 결핵균 균주들 간에도 발병력의 차이가 있는 것으로 나타나고 있다. 결핵균이 호흡을 통하여 폐로 흡입되어도 모두 발병으로 이어지는 것이 아니라 약 10%의 사람에서 결핵으로 진행되며 나머지 90%에

서는 감염의 징후 없이 숙주의 면역작용에 의해 흡입된 결핵균이 사멸되거나 휴면 상태로 감염되어 숙주의 면역력이 약해졌을 때 발병하게 된다.

균주의 흡입에서 감염 및 발병에 이르기까지 관여하는 요인들 중 첫번째는 숙주의 유전적인 요인이다. Bellamy 등[7]은 common ancestry using microsatellite (CAM) mapping을 통하여 사람의 염색체상의 두 지역 15q와 Xq에 결핵 감수성 유전자가 존재한다는 것을 증명하였다. 이는 숙주의 유전적 요인이 결핵균 감염에 중요한 요소로 작용한다는 것을 의미한다. 두번째는 각 결핵 균주들이 숙주의 면역 반응을 유발하는 정도의 차이이다. CDC1551 균주는 성장 속도의 측면에서는 *in vivo*, *in vitro* 상에서 다른 임상 분리 균주들보다 발병력이 낮지만, 숙주의 면역반응을 보다 급격하고 왕성하게 유도하는 것으로 나타났다[8]. 세번째는 결핵균의 유전적 요인이다. 이는 균주마다 독특한 점 돌연변이 혹은 결손 및 삽입에 의한 돌연변이에 의하여 특정 단백질의 변화가 유발되고, 독성에 변화가 생기는 경우이다. 결핵균 26개 유전자에 대한 염기서열 분석 결과, 염기의 치환이 일어나 아미노산이 바뀌거나 다른 돌연변이가 일어난 경우의 95% 이상이 균주의 항생제 내성과 연관되어 있음이 밝혀졌다[9,10].

한국에서 결핵 환자들로부터 분리된 임상분리 균주들을 IS*6110*를 이용한 restriction fragment length polymorphism (RFLP) 양상을 기준으로 분류하였을 때, 독특한 RFLP 양상을 보이는 균주들이 발견되었다. 그 중에서도 가장 높은 빈도로 발견되는 집단의 대표적인 균주를 K 균주라 명명하였다[11]. K 균주는 국내의 임상분리 균주에서 가장 높은 빈도로 발견될 뿐 아니라, 고등학교에서 집단 발병에 관련이 있다는 점으로 볼 때, 발병력이 다른 집단의 균주들보다 높을 것으로 추측된다.

한편 최근에 미국에서 집단 발병을 일으킨 임상분리 균주인

CDC1551 균주[12]는 환자와 짧은 접촉을 가졌던 접촉자들을 높은 빈도로 sero-conversion시키는 것으로 확인되어, 발병력이 매우 높은 균주인 것으로 예상되었다[13]. 그러나 최근의 동물 실험 결과, CDC1551은 다른 결핵 균주들에 비해 숙주에서 높은 면역 반응을 유도시키지만, Erdman이나 다른 임상분리 균주들보다 발병력은 그리 높지 않은 것으로 밝혀졌다[8].

4. 결핵 균주들의 유전적 다형성

Mycobacterium tuberculosis complex는 유전체의 염기서열과 16s rRNA 염기서열에서 99.9%의 유사성을 보임에도 불구하고[10,14] 숙주특이성과 발병력에 있어서 많은 차이를 보이고 있다. 이러한 표현형의 차이에 대한 유전적인 배경을 규명하기 위해서는 각 균주들에 대한 비교유전체 연구가 필수적이다.

전체 염기서열이 밝혀진 결핵균 H37Rv 균주와 CDC1551 균주의 유전체의 비교 분석에 의하여 두 균주 사이에 다수의 large-sequence polymorphisms (LSPs)과 single-nucleotide polymorphisms (SNPs)이 존재한다는 것이 밝혀졌지만, 이런 유전적 다형성과 결핵 균주들의 발병력과의 관계에 대한 분자적인 기반 연구는 아직 미흡하다[12,15]. 또한 DNA microarray 기법을 이용하여 결핵 균주들 사이의 유전체를 비교하는 연구들[16,17]은 유전자 결손에 의한 LSP의 확인에는 효과적인데 비해서, 알려진 염기서열을 바탕으로 hybridization을 통해 대상 균주들 사이의 차이를 탐색하는 실험 방법의 특성상, 유전자 삽입에 의한 유전적 다형성을 규명하는 것은 불가능하다는 단점이 있다. 따라서 결손 및 삽입을 모두 포함하는 보다 정확하고 포괄적인 다형성의 연구를 위해서는 임상분리 균주들에서 차이를 보이는 부위의 염기서열을 확보함으로써, 이들을 비교 분석하여 유전적 다형성을 규명할 수 있는 실험 방법이 필수적이다.

5. 연구의 목적

본 연구는 다양한 임상분리 균주들을 대상으로 각 균주들에 존재하는 유전적 다형성을 규명하는 것을 최종적인 연구목적으로 한다. 이를 위하여 최근에 염기서열 분석이 완료된 국내 임상분리 균주인 K 균주의 유전체 염기서열을 결핵균 표준 균주 H37Rv 및 CDC1551의 유전체 염기서열과 비교하여, 각 균주 유전체에서 상이한 부위들을 검색하고, 이 결과를 바탕으로 국내에서 분리한 몇 가지 임상분리 균주에서 동일한 부위의 염기서열을 확인하고 분석함으로써 다양한 임상분리 균주에 존재하는 유전적 다형성을 밝히고자 한다.

II. 재료 및 방법

1. 실험재료

가. 사용 균주

표준 균주로는 *M. tuberculosis* H37Rv (ATCC 27294)를 사용하였고, 임상분리 균주들은 한국에서 분리된 *M. tuberculosis* K 균주를 포함하여 특징적인 IS*6110*-RFLP 양상을 보이는 39가지를 선정하여 연구대상으로 하였다 (그림 1).

나. 효소 및 시약

DNA polymerase는 SolGent (Daejeon, Korea) 제품을 사용하였고, agarose는 USB (Cleveland, OH, USA)의 Agarose-LE와 CAMBREX (Rockland, ME, USA)의 MetaPhor® Agarose를 사용하였다.

2. 실험방법

가. 결핵균 배양

균주들은 0.05% (v/v) Tween-80, ADC [0.5% (w/v) bovine serum albumin fraction V, 0.2% (w/v) glucose, 0.0003% (w/v) catalase]가 함유된 Difco™ Middlebrook 7H9 (Becton, Dickinson and Company, Sparks, MD, USA) 액체배지에서 37℃에서 3~4주일 배양하였다.

나. 결핵균에서 genomic DNA 분리

PCR 분석을 위한 결핵균 DNA는 cetyltrimethylammonium

bromide (CTAB)을 사용한 방법에 의해 준비하였다[18]. Difco™ Middlebrook 7H9 액체 배지에서 3~4주 동안 배양한 균을 원심분리하여 상층액을 제거한 다음, 5 ㎖ 배양을 기준으로 했을 때, 500 ㎕의 Tris-EDTA (pH 8.0)을 넣고 잘 풀어 주었다. 그 다음 80℃에서 20분 이상 살균하고, 25 ㎕의 lysozyme (20 ㎎/㎖)을 첨가하여 37℃에서 1시간 이상 반응시켰다. 그리고 70 ㎕의 10% SDS와 6 ㎕의 proteinase K (10 ㎎/㎖)를 첨가하여 65℃에서 10분간 반응시키고, 100 ㎕의 5 M NaCl, 80 ㎕의 CTAB-NaCl (50 ㎕의 5 M NaCl과 40 ㎕의 10% CTAB)을 첨가하여 잘 섞어준 후, 다시 65℃에서 10분간 반응시켰다. 시료와 동일한 양의 chloroform-isoamyl alcohol [24 : 1 (vol : vol)]을 넣어 잘 섞어준 후, 13,000 ×g에서 10분간 원심분리하여 상층액만 새 tube에 옮겼다. DNA 침전을 위해 시료의 6배의 isopropanol을 넣고 섞어준 다음, -20℃에서 30분간 방치하고, 13,000 ×g에서 20분간 원심분리하였다. 마지막으로 상층액을 제거하고 70% ethanol로 씻어준 후 건조시킨 DNA는 20 ㎕의 Tris-EDTA (pH 8.0)에 녹였다.

다. 연구대상 부위의 선정 및 PCR primer의 고안

K 균주와 H37Rv 또는 CDC1551 균주의 염기서열을 다양한 program들 (BLAST, FASTA33, sim4)을 이용하여 비교하여 각 균주의 유전체 사이에서 결손 및 삽입이 일어난 부위를 확인하였다. 확인된 각 부위의 주변 염기서열을 대상으로 결손 및 삽입이 일어난 부위를 포함하는 DNA 절편을 증폭할 수 있는 PCR primer들을 고안하였다. Primer는 primer3 program (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3) 등을 이용하여 결정하였고, 각 부위에 대하여 한 쌍 이상의 primer들을 고안하여 합성하였다. 이때 모든 PCR을 동일한 조건에서 실시할 수 있도록 각 primer들의 G+C 함량을 유사하게 조정하였다. 실험에 사용된

primer는 표 1에 정리되었다.

라. PCR

10 ng의 결핵균 genomic DNA, 12.5 mM dNTP, 각 20 uM의 forward와 reverse primer, 0.5×~2.0× BandDoctor (SolGent, Daejeon, Korea), *Taq* buffer [75 mM Tris-HCl, 15 mM$(NH_4)_2SO_4$, 2.5 mM $MgCl_2$, 0.1 ㎎/㎖ BSA, 0.5% DMSO, pH 9.0], 2.5 U *Taq* polymerase (SolGent, Daejeon, Korea)를 넣고, 증류수로 최종반응 용량을 50 ㎕로 맞춘 후 다음 조건으로 PCR을 수행하였다. 95℃에서 5분간 pre-denaturation하여 DNA를 완전히 해리시킨 후, 95℃에서 30초간 denaturation, 각 primer의 해당 Tm에서 30초간 annealing, 72℃에서 3분간 extension을 30 cycles 수행한 후, 72℃에서 10분간 extension을 실시하였다. PCR이 끝난 후에는 PCR 산물을 1% agarose gel 또는 3% metaphor agarose에 전기영동하여 합성된 DNA단편을 확인하고, QIAquick PCR purification kit (Qiagen, Germany)을 이용하여 정제하였다.

마. 염기서열 확인

다형성을 보이는 K 균주와 임상분리 결핵 균주들의 삽입이나 결손이 일어난 부위의 염기서열을 확인하기 위하여, PCR 산물을 정제한 후, 외부기관에 의뢰하여 sequencing을 수행하였다.

바. 염기서열 분석

먼저 각 연구대상 부위에서 얻어진 염기서열들을 clustalX 1.83 program을 이용하여 multiple alignment를 시키고, 그 결과를 GeneDoc program을 이용하여 다시 정리하여 결손 및 삽입이 일어난 염기서열을 확인하였다. 결손 및 삽입이 일어난 위치에 따라, 유전자 (ORF) 내부 혹은 intergenic sequence에 결손 혹은 삽입이 일

어난 두 그룹으로 분류하였다. 알려진 ORF에 변이가 유발된 경우에는 각 ORF들의 아미노산 서열을 비교하여 변이 단백질의 기능성의 변화를 추론하였다. Intergenic sequence에 유발된 변이에 대해서는 각 부위의 염기서열을 이미 알려진 결핵균의 promoter 염기 서열들과 비교하여 발견된 유전적 변이가 downstream에 위치하는 특정 유전자의 발현에 미칠 영향을 추론하였다. 일정 크기 이상의 삽입된 염기서열에 대해서는 Amigene program을 이용하여 가능한 ORF를 탐색한 후, 각 ORF의 염기서열 및 아미노산 서열을 BLAST algorithm을 이용하여 GeneBank database를 검색하여 각각의 기능을 유추하였다.

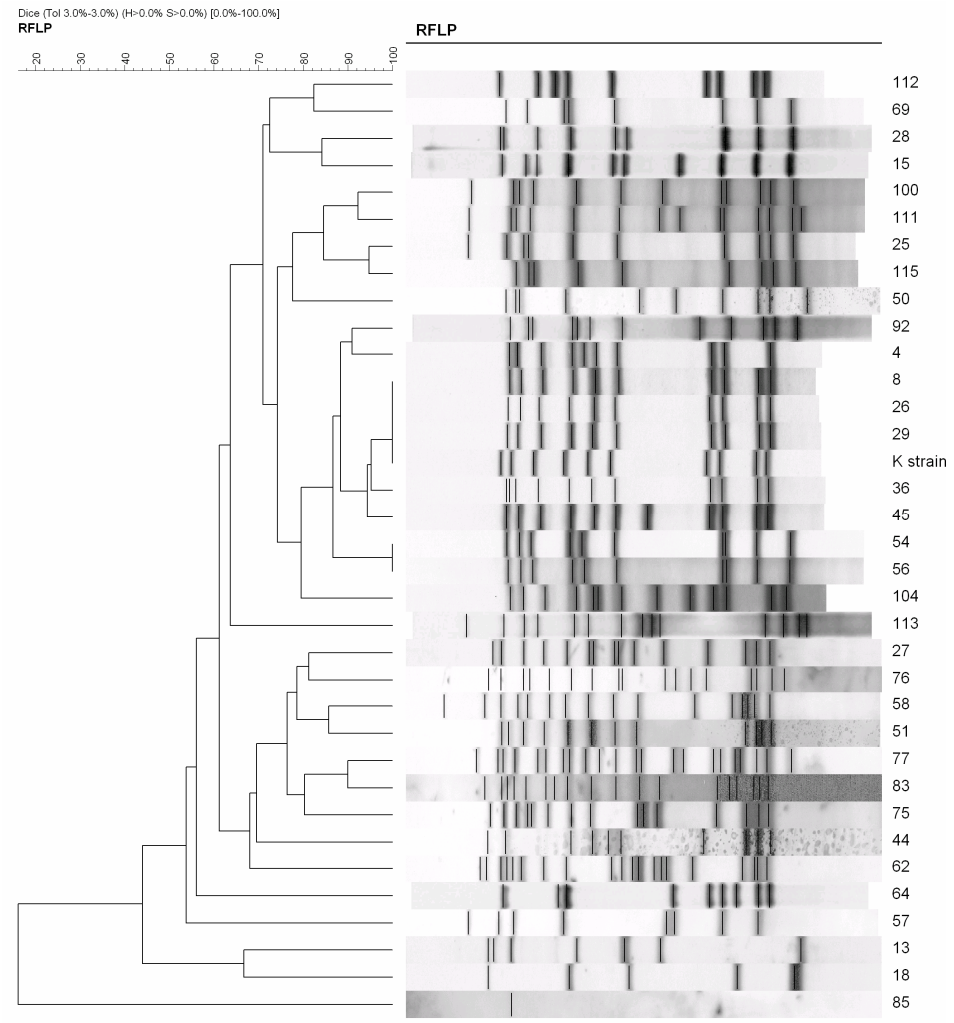

그림 1. 실험에서 사용된 임상분리 균주의 RFLP 결과.

표 1. PCR에 사용된 primer들

InDel	Forward primer (5' – 3')	Reverse primer (5' –3')
Del_1	AGAGGGCACATAGCTGCCCT	TTTGGCGCAGAACCAGGAGT
Del_2	TTCGTGCACAGTTGGGTGTT	TGGCCCAGAAACCACCAGTT
Del_3	TCCTTGCCGCGTGCATCGAT	TACCAGCAGCACCGCAAGCA
Del_8	TGGTAGTCACCTCGCCAGAA	TGCCAATATCGGCTTCGCCA
Del_9~10	TTGAAACTGCCCGAGTTGGT	TAACCTCTCGGGACTCTCCA
Del_12	TTGATCGAGGCCTATCACGA	GTCACCACCGTGGCCTCAAA
Del_13	AGCCGCCTTGCTCTTCGGCT	CCGTTGCCAGAGTTGCTGAA
Del_16	AACCGTGAAGCCGTGTGCGA	AACAGCGGCTTCACCCAGAT
Del_17	AATCAACGGCGATGGCGGTA	TTGGCTTGGGCATTGAAGGT
Del_18	CTTGGGTGGTTCTATCCGAG	ACTTCGATCTCGTAGCGTTC
Del_19	TGATCGGCAACGGTGGGTTC	CTACTGAAGGTGACGCTCTG
Del_20	TGATGGAGGAGACCATCTCG	GTGGCAATGTCGTTGACCGT
Del_21	CGAATAATGACTGCCGCTCT	TGCAAATGCTGGCAACAGCA
Del_22	TCAAAGACATGCCGAGCGTG	ACTCAGCGCCAAACCGACAC
Del_23	AGTTGGTTACCGGTGCCAGT	CTGCTCAACATTGTGAACAC
Del_25	TTGGCACCGCTGTTAGGACT	GATGCCGGAATGTTCGGTAA
Del_29	AGGGAGGACTGGTTGGACTT	TGCTGTGCGATCGAGGCGAA
Del_30	ATTCCCGACGAAAGTGACGA	ATCGCGCTCTTGTCACTAGC
Del_31	ATCAACGTCTCCGCATCGAT	TCTGCGAGCCAATTCAGCTT
Del_32	AGTCACCACCAGTTCGGGCA	TGGTCTTCGGGTCCAACTCG
Del_36	AACGCGGTGTCGATCCACGA	TACTCACCGTGGTAGCCGAC
Del_39	AACACCAGGACAGCATCATG	ACAGTCGTGTTGACCAGTGC
Del_40	TCAGCGATCTCCATGCTGGT	CACCGTCTGACTCATGGTGT
Del_42~60	TGCTGCCGATGTTGCCGCTA	TGCTGAGCCGCGAATCGTGA
Del_63	TGCTTAGCGTGGTTCCACTC	TGAACGCATCAACGGCGTTC
Del_131	TAGTCACGCACCGACAAGGT	AAGACGTGGACGACGTGCGA
Del_132	AGTGTCAACCACAGGTACTT	TGGATTGATGGTGATGACCT
Del_141	ATCTGACAGGATCTGGCAAC	CTCCATCGGCTTTGCTGGTA
Del_144	AGTACGAGGACACACTGGTC	AGTGTGATGCGCTGCCAACT
Del_154~5	TGGTAGCGCCGAAGCTCCCA	CAGCCTGAGCACAGGCCAGT
Del_177	TGGTGACGGCCAGAACGGAA	TTGCCGTTGGTGCCGCCTGT
Del_180	AAACGGTGGGATTGGCGGAA	TTACCGCCATTGCCACCTGT
Del_186	AAACACCGACGGACTCAGCA	TACCACCGTTACCACCGTTA
Del_188	AGGGAACAACCTTCGACCAC	CAACAGACACGTTGAGTCCT
Ins_2	ATCGACAAGGCCGAAGCCGA	TCCGCCGAGGACTTCTGCTT
Ins_3	TGCTGGTCAGCGTGCGTGAA	CATATACGTCTGCAGCGACT
Ins_5	AGGGCACCTCTTCGGCAATG	AGATCGCCAGCTCAGCGGAT
Ins_6	AGGTTGCAAGAGATGGCTGA	TACAACACGAGCATGCTGGA
Ins_7	TGTTGTACGGCAACGGCGGT	ATTACCAAGCAACCACCCAC
Ins_8	TGTGTGGTGGAACCCATCTG	ACCAGGTTTCGGATGGTGGT
Ins_9	TCAAGGTCGCATCGGTGAAC	TCGTGGGCTTGTACTTGCGA
Ins_10~1	AGGTGGCTGGTTGTTGGGCA	ACCGATCAGCAACCCGGCCT
Ins_12	ACCGGCGGGCTCAGCACTAA	TGCAGCGCTTCCACACGTTT

InDel	Forward primer (5' – 3')	Reverse primer (5' –3')
Ins_13	TGGACCGTTGCCGATGGTGA	GCCTATCAGCAACGGTTCGT
Ins_14	TTGCCGCCGATACCGGGATT	CACCGCCGGTAAGGGTGGTA
Ins_15	ATCAATACCGCTGTTGCCGT	ATGGCAGCGTTTCATGACCA
Ins_16	AATGGCGGAAAAGCCGGTTT	AACCAGTCCGATGCCATCAA
Ins_17	ACTTCGGCGACCTGTGGAAT	ACCCAAATTGCCCTGGTCGT
Ins_18	TCGTGTTGTTTGCCCGGCTT	GTGGATTCTTGCTTGGCTGA
Ins_19	ATCGGTCAGCGCATCCTTGA	AGAGGAGAGTTGACGCTGAT
Ins_20	ACACCCTGGATTCCGTTCTG	TGCCAAGGGCACGTTCGATA
Ins_23	TCGCACATCTCGACGGTCGT	ATCCTCCTGATTGCCCTGAC
Ins_24	AGCATTGCCGCCGTTACCAA	TTTCACGCCGAGTTCGTGAG
Ins_25	AGCGTCAGCATGAGCTGCAA	AACCAACTGATCGCAGACGT
Ins_26	ACGCTAGACGTCAGATCCCA	ATGAGCACCGACGATGCCAA
Ins_28	AAGCCGCTTACGGTGATGTT	TGCTAACACGGGTTTCCTCA
Ins_29	AGCCTGCTTCAACCAGATCT	TGGATCAGGTCATGGAAGGT
Ins_30	AGGATCGGAATGACAGCCAG	GTGCTGATTAAGTGGTGGCT
Ins_31	GACAATCAAGCGGCTGGCAA	AGCAGGAACTCAAGGGCGTA
Ins_32	TCCATCGCCACTGTTGAAGA	TTCCGTTTAACGTCGCTGGT
Ins_33	ACGTCGATGCCCAACGGGAT	AATCCCACCGTTGGCAGCAT
Ins_35	ACTACCACCATCTGCGCTTT	ACGCTCACAACTCGCAGCGT
Ins_38	ATGAATCGGCCTTCTCCAGT	ATCAAGACGTCCAAGCAGGT
Ins_39	ATGTGATACACGCCGACGTT	ATCATCGGAAGGCGTCGACT
Ins_40	ATTCCACCTCGTAGTCCCAA	ACCAGCTTTCAGCCTCCACA
Ins_41	TTGGTGGTATCGAACCGGAT	AGTAGGGCACTAGCACCTCA
Ins_67~8	GAAGTTCGTGTCGCCGATGT	GACGTCTCGGTGACGGGTGT
Ins_71	TCGTTGGCCATCAGCCGAAT	AGCTGGGCCTGCGGTGTATT
Ins_72	TGTGCTGCCCAACTCGGTGA	AGCTTCTGACGCCACTCGTC
Ins_73	ACGTGTCGGCGGATAGGTCT	AATAGTGCCGTCCCATGCAT
Ins_83	TTCACTCCAAGGTCGGAGTT	ATGCCGATCCTGGTGTTCAA
Ins_84	ATGGTGCCGCCATGTTTCAT	AAGCTGGTCGGGCTTGGTGA
Ins_87	ATACGTGGAATGGCTGGAGC	GTGCTGGTCGAGAACTGAAT
Ins_88	AATTGACCACGACGACAGCT	TACGAGTACAGAGACGTGCA
Ins_123~5	AGAAACCCGACGACGGAGCA	CTACAGCAACGGCGCCTTCT
Ins_126	GACTGGCCTGTGCTCAGGCT	GTGTCCTTCGGCAACGCTGT
Ins_138	ACGGACACACCGTCCACCAT	TGAATGAGACGTTCACCGGT
Ins_140	TAGGTCAGGGCGACCTCCAT	TACCGGTACAAGTAGCGACA
Ins_141	AACACCCTCAACCCCGACAC	CTGTTGCCGGTGCCAGCCTT
Ins_142~3	AGCGGTGGTAATGGCGGTGA	TTGGGTGTGGGCAAGCTGGT
Ins_147	AAGGAGGTCAAGGCGGCAGC	TTACCGCCATTGCCACCTGT
Ins_149~150	TGTTATGGTGCGAGTCACGT	TGACCCGAACGCTCAGCTGT
Ins_151	AGACAGCCAAACGGGAGATC	CTCAGGGACTGCGGATAGGT
Ins_152	ACCTTGGGATTTAGCAGGCT	TGTGGTGGTCGACTTGCCGT
Ins_153	ATGGAAGACCCGGTCAAGTT	TCACGGGCGATGAGCTCGTT
Ins_154	TGTTGCTGGACCAGACGGAC	ACGCACTCGAGGATTTGCTT

III. 결 과

1. K 균주와 H37Rv 균주의 염기서열 비교

K 균주와 K family에 속하는 균주들은 한국에서 18.4%라는 높은 빈도수로 발병하고 있고, 집단 발병에도 연관이 있는 것으로 밝혀져 있어[11] 발병력이 강할 것이라고 추측되고 있다. 하지만 K 균주의 발병력에 대한 유전적인 배경에 대해서는 아직 밝혀진 바가 없다. 그래서 본 연구에서 K 균주의 전체 염기서열을 표준 균주인 H37Rv와 비교하여 10 bp 이상 차이가 나는 LSP를 확인하였다.

그 결과, H37Rv 균주와 비교했을 때 K 균주에서 153개의 결손과 124개의 삽입이 일어났음을 확인하였다 (표 2). 결손 중에서 135개는 유전자 내에서 일어났으며, 18개는 intergenic sequence에서 일어났다. 삽입 중에서는 99개는 유전자 내에서 일어났고, 25개는 intergenic sequence에서 일어난 것으로 확인되었다. 결손과 삽입의 크기가 1358 bp ~ 1359 bp인 것은 IS*6110*에 의한 것으로, 결손의 경우에는 10개, 그리고 삽입의 경우에는 3개가 IS*6110*에 의해서 일어난 것으로 밝혀졌다. 결손 부위와 삽입 부위는 각각 표 3과 표 4에 정리되었다.

결손과 삽입으로 인해서 영향을 받은 유전자는 각각 96개와 51개로 나타났는데 이들의 대부분은 PE, PPE family, transposase, hypothetical protein이었다. 결손과 삽입에 의해 영향을 받은 ORF들은 표 5와 표 6에 정리되었다.

2. K 균주와 Beijing 균주의 결손 부위 비교

동남 아시아에서 높은 빈도로 나타나고, 뉴욕과 러시아에서도

표 2. K 균주에서 일어난 결손과 삽입

	결 손	삽 입
유전자내	135	99
Intergenic sequence	18	25
합 계	153	124

표 3. K 균주에서 확인된 결손 부위

Del	Size (bp)	Locus	Product
1	18	Rv0020c	Hypothetical protein
2	3466	Rv0071	Maturase
		Rv0072~Rv0073	ABC transpoter
		Rv0074	Hypothetical protein
3	1225	Rv0193c	Hypothetical protein
		Rv0194	ABC transpoter
4	63	Rv0278c	PE-PGRS
5	18	Rv0278c	PE-PGRS
8	15	Rv0304c	PE, PPE family
9	156	Rv0355c	PE, PPE family
10	15	Rv0355c	PE, PPE family
11	155	Rv0387c	Hypothetical protein
12	24		
13	48	Rv0747	PE-PGRS
14*	1359	Rv0795~Rv0796	Transposase
16	3533	Rv0973c	Acetyl-/propionyl- CoA
		Rv0974c	carboxylase α,β
		Rv0975c	Acyl-CoA dehydrogenase
17	48	Rv0977	PE-PGRS
18	48	Rv1059	Hypothetical protein
19	48	Rv1087	PE-PGRS
20	62		
21	105	Rv1243c	PE-PGRS
22	114		

23	36	Rv1325c	PE-PGRS
24*	1358	Rv1369c~Rv1370c	Transposase
25	18	Rv1441c	PE-PGRS
26	9247	Rv1572c	Hypothetical protein
		Rv1573~Rv1587c	phiRv1
27	1064	Rv1754c	Hypothetical protein
		Rv1755c	Phospholipase C
28	9565	Rv1758	Cutinase
		Rv1759c	PE-PGRS
		Rv1760~Rv1762c	Hypothetical protein
		Rv1763~Rv1764	Transposase
		Rv1765c	Hypothetical protein
29	45	Rv1787	PE, PPE family
30	56		
31	1204	Rv1878	Glutamine synthase
		Rv1879	Hypothetical protein
32	14	Rv1907c	Hypothetical protein
33	57	Rv2090	5'-3' exonuclease
35*	1358	Rv2105~Rv2106	Transposase
36	37	Rv2129c	Oxidoreductase
38*	1358	Rv2167c~Rv2168c	Transposase
39	57		
40	53		
41*	1358	Rv2277c	Glycerolphosphodiesterase
		Rv2278~Rv2279	Transposase
42	11	Rv2354	Transposase
43	27	Rv2354	Transposase
44	15	Rv2355	Transposase

45	26	Rv2355	Transposase
46	22	Rv2355	Transposase
47	34	Rv2355	Transposase
48	12	Rv2355	Transposase
49	24	Rv2355	Transposase
50	27	Rv2355	Transposase
51	42	Rv2355	Transposase
52	42	Rv2355	Transposase
53	28	Rv2355	Transposase
54	10	Rv2355	Transposase
55	47	Rv2355	Transposase
56	47	Rv2355	Transposase
57	56	Rv2355	Transposase
58	32	Rv2355	Transposase
59	12	Rv2355	Transposase
60	11	Rv2355	Transposase
63	20	Rv2426c	Hypothetical protein
64*	1358	Rv2479c~Rv2480	Transposase
66*	1358	Rv2648~Rv2649	Transposase
68	65		
69	11		
70	16		
71	20		
72	15		
73	12		
74	75		
75	84	Rv2814c	Transposase
76	54	Rv2814c	Transposase

77	31	Rv2814c	Transposase
78	21	Rv2814c	Transposase
79	18	Rv2814c	Transposase
80	12	Rv2814c	Transposase
81	75	Rv2814c	Transposase
82	85	Rv2814c	Transposase
83	10	Rv2814c	Transposase
84	61	Rv2814c	Transposase
85	20	Rv2814c	Transposase
86	27	Rv2814c	Transposase
87	10	Rv2814c	Transposase
88	16	Rv2814c	Transposase
89	24	Rv2814c	Transposase
90	12	Rv2814c	Transposase
91	24	Rv2814c	Transposase
92	67	Rv2814c	Transposase
93	12	Rv2815c	Transposase
94	38	Rv2815c	Transposase
95	1368	Rv2815c	Transposase
96	134		
97	293		
98	63	Rv2816c	Hypothetical protein
99	11	Rv2816c	Hypothetical protein
100	12	Rv2816c	Hypothetical protein
101	30	Rv2816c	Hypothetical protein
102	12	Rv2816c	Hypothetical protein
103	37	Rv2816c	Hypothetical protein
104	20	Rv2816c	Hypothetical protein

105	28	Rv2816c~Rv2817c	Hypothetical protein
106	47	Rv2817c	Hypothetical protein
107	124	Rv2817c	Hypothetical protein
108	120	Rv2817c	Hypothetical protein
109	15	Rv2817c	Hypothetical protein
110	343	Rv2817c	Hypothetical protein
112	15	Rv2817c	Hypothetical protein
113	101	Rv2817c~Rv2818c	Hypothetical protein
114	118	Rv2818c	Hypothetical protein
115	246	Rv2818c	Hypothetical protein
116	13	Rv2818c	Hypothetical protein
117	85	Rv2818c	Hypothetical protein
118	34	Rv2818c	Hypothetical protein
119	57	Rv2818c	Hypothetical protein
120	44	Rv2818c	Hypothetical protein
121	96	Rv2818c	Hypothetical protein
122	69	Rv2818c	Hypothetical protein
123	66	Rv2818c	Hypothetical protein
124	96	Rv2818c	Hypothetical protein
125	269	Rv2818c~Rv2819c	Hypothetical protein
126	115	Rv2819c	Hypothetical protein
127	53	Rv2819c	Hypothetical protein
128	133	Rv2819c	Hypothetical protein
129	892	Rv2819c~Rv2820c	Hypothetical protein
130	274	Rv2820c	Hypothetical protein
131	108		
132	55		
141	14		

142*	1358	Rv3184~Rv3185	Transposase
143*	1358	Rv3186~Rv3187	Transposase
144	135	Rv3281	Hypothetical protein
145	36	Rv3325	Transposase
146	19	Rv3325	Transposase
147	15	Rv3325	Transposase
148	17	Rv3325	Transposase
149	23	Rv3325~Rv3326	Transposase
150	11	Rv3326	Transposase
151	256	Rv3326	Transposase
154	138	Rv3344c	PE-PGRS
155	27	Rv3344c	PE-PGRS
156	501	Rv3345c	PE-PGRS
158	26	Rv3380c	Transposase
159	27	Rv3380c	Transposase
160	30	Rv3380c	Transposase
161	13	Rv3380c	Transposase
162	28	Rv3380c	Transposase
164	12	Rv3381c	Transposase
165	1376	Rv3388	PE-PGRS
168*	1358	Rv3474~Rv3475	Transposase
170	72	Rv3508	PE-PGRS
174	14	Rv3508	PE-PGRS
175	12	Rv3508	PE-PGRS
177	30	Rv3508	PE-PGRS
180	18	Rv3512	PE-PGRS
184	126	Rv3514	PE-PGRS
185	351	Rv3514	PE-PGRS

186	603	Rv3514	PE-PGRS
188	89	Rv3785	Hypothetical protein

* IS*6110*으로 인해 결손이 일어난 것

표 4. K 균주에서 확인된 삽입 부위

Ins	Size (bp)	Locus	Product
1*	1358		
2	112		
3	37		
5	162	Rv0124	PE-PGRS
6	51		
7	84	Rv0532	PE-PGRS
8	533		
9	54	Rv0703	50S ribosomal protein
10	62	Rv0746	PE-PGRS
11	26	Rv0746	PE-PGRS
12	136	Rv0746	PE-PGRS
13	12	Rv0978c	PE-PGRS
14	18	Rv1067c	PE-PGRS
15	193	Rv1068c	PE-PGRS
16	45	Rv1087	PE-PGRS
17	56	Rv1290c	Hypothetical protein
18	1674	Rv1319c	Adenylate cyclase
19	21	Rv1435c	P, G ,V-rich secreted protein
20	27	Rv1450c	PE-PGRS
21	207	Rv1450c	PE-PGRS
23	53	Rv1458c	ABC transporter
24	21	Rv1651c	PE-PGRS
25	10	Rv1668c	ABC transporter
26	171		

28	234	Rv1753c	PE, PPE family
29	15	Rv1883c	Hypothetical protein
30	10	Rv1888c	Transmembrane protein
31	45	Rv1899c	Lipoprotein
32	414	Rv1917c	PE, PPE family
33	75	Rv1917c	PE, PPE family
35	4495		
36*	1358	Rv2016	Hypothetical protein
37*	1358	Rv2019	Hypothetical protein
38	4999	Rv2024c	Hypothetical protein
39	11	Rv2084	Hypothetical protein
40	57	Rv2112c	Hypothetical protein
41	116		
56	2084		
57	10	Rv2354	Transposase
58	13	Rv2354	Transposase
59	99	Rv2356c	PE, PPE family
60	13	Rv2356c	PE, PPE family
61	43		
64	16		
65	35		
66	29		
67	21	Rv2357c	Glycyl-tRNAsynthetase
68	55	Rv2357c	Glycyl-tRNAsynthetase
69	42	Rv2357c	Glycyl-tRNAsynthetase
70	21	Rv2357c	Glycyl-tRNAsynthetase
71	21	Rv2407	Hypothetical protein
72	55	Rv2673	Integral membrane protein

73	153	Rv2680	Hypothetical protein
74	19		
75	12	Rv2814c	Transposase
76	67	Rv2815c	Transposase
77	24	Rv2815c	Transposase
78	12	Rv2815c	Transposase
79	24	Rv2815c	Transposase
80	16	Rv2815c	Transposase
81	10	Rv2815c	Transposase
82	27	Rv2815c	Transposase
83	106	Rv2882c	Ribosome recycling factor
84	336		
85	58		
86	234		
87	1337	Rv3135	PE, PPE family
88	2146	Rv3159c	PE, PPE family
89	58		
90	18		
91	11	Rv3325	Transposase
92	12	Rv3325	Transposase
93	22	Rv3325	Transposase
94	12	Rv3325	Transposase
95	18	Rv3325	Transposase
96	17	Rv3325	Transposase
97	13	Rv3325	Transposase
98	75	Rv3326	Transposase
99	15	Rv3326	Transposase
100	48	Rv3326	Transposase

101	106	Rv3326	Transposase
104	67	Rv3326	Transposase
105	77	Rv3326	Transposase
106	36	Rv3326	Transposase
107	44	Rv3326	Transposase
108	18	Rv3326	Transposase
109	23	Rv3326	Transposase
110	50	Rv3326	Transposase
111	26	Rv3326	Transposase
112	17	Rv3326	Transposase
115	86	Rv3326	Transposase
116	184	Rv3326	Transposase
117	82	Rv3326	Transposase
118	14	Rv3326	Transposase
119	17	Rv3326	Transposase
120	199	Rv3326	Transposase
121	53		
123	23		
124	10	Rv3343c	PE, PPE family
125	10	Rv3343c	PE, PPE family
126	27	Rv3345c	PE-PGRS
127	402	Rv3350c	PE, PPE family
128	12	Rv3380c	Transposase
129	10	Rv3380c	Transposase
130	12	Rv3380c	Transposase
132	28	Rv3381c	Transposase
133	13	Rv3381c	Transposase
134	23		

135	11	Rv3382c	LYTB-related protein
136	11	Rv3382c	LYTB-related protein
137	16	Rv3382c	LYTB-related protein
138	232		
140	57	Rv3464	dTDP-glucose rehydratase
141	18	Rv3507	PE-PGRS
142	18	Rv3507	PE-PGRS
143	81	Rv3507	PE-PGRS
147	135	Rv3512	PE-PGRS
148	26	Rv3512	PE-PGRS
149	111	Rv3611	R, P-rich protein
150	333	Rv3611	R, P-rich protein
151	342		
152	59		
153	53		
154	36	Rv3879c	A, P-rich protein

* IS*6110*으로 인해 삽입이 일어난 것

표 5. 결손에 의해 영향을 받은 ORF들

	ORF name	No.
Persisted ORFs[1]	Rv0020c, Rv0278, Rv0304c, Rv0355c, Rv0977, Rv1059, Rv1087, Rv1243c, Rv1754c, Rv1787, Rv1879, Rv2090, Rv2426c, Rv3281	14
3' changed ORFs[2]	Rv0387c, Rv0747, Rv1441c, Rv2129c, Rv2277c, Rv3785	6
Disrupted ORFs[3]	Rv0071, Rv0074, Rv0193c, Rv0194, Rv0973c, Rv0975c, Rv1325c, Rv1587c, Rv1878, Rv1907c, Rv2354, Rv2355, Rv2814c, Rv2815c, Rv2816c, Rv2817c, Rv2818c, Rv2819c, Rv2820c, Rv3325, Rv3326, Rv3344c, Rv3345c, Rv3380c, Rv3381c, Rv3388, Rv3508, Rv3514	28
Deleted ORFs[4]	Rv0072, Rv0073, Rv0795, Rv0796, Rv0974c, Rv1369c, Rv1370c, Rv1572c, Rv1573, Rv1574, Rv1575, Rv1576c, Rv1577c, Rv1578c, Rv1579c, Rv1580c, Rv1581c, Rv1582c, Rv1583c, Rv1584c, Rv1585c, Rv1586c, Rv1755c, Rv1758, Rv1759c, Rv1760, Rv1761, Rv1762, Rv1673, Rv1764, Rv1765c, Rv2105, Rv2106, Rv2167c, Rv2168c, Rv2278, Rv2279, Rv2479c, Rv2480, Rv2648, Rv2649, Rv3184, Rv3185, Rv3186, Rv3187, Rv3474, Rv3475, Rv3512	48
Total		96

[1] 결손이 일어났지만 ORF frame에는 영향을 주지 않은 경우

[2] 결손으로 인해서 ORF의 3' 부분에서만 frame이 바뀐 경우

[3] 결손으로 인해서 ORF가 완전히 망가진 경우

4 결손이 일어나면서 ORF가 완전히 삭제된 경우

표 6. 삽입에 의해 영향을 받은 ORF들

	ORF name	No.
Persisted ORFs[1]	Rv0703, Rv0978c, Rv1087, Rv1290c, Rv1435c, Rv1458c, Rv1651c, Rv1753c, Rv1883c, Rv1899c, Rv1917c, Rv2023A, Rv2112c, Rv2354, Rv2407, Rv2673, Rv2680, Rv2882c, Rv3343c, Rv3464, Rv3611	21
3' changed ORFs[2]	Rv0124, Rv0354c, Rv0532, Rv1067c, Rv1068c, Rv1319c, Rv1668c, Rv1888c, Rv2016, Rv2084, Rv2357c, Rv3135, Rv3382c, Rv3879c	14
Disrupted ORFs[3]	Rv0746, Rv1450c, Rv2019, Rv2024c, Rv2356, Rv2814c, Rv2815c, Rv3159c, Rv3325, Rv3326, Rv3345c, Rv3350c, Rv3380c, Rv3381c, Rv3507, Rv3512	16
Total		51

[1] 삽입이 일어났지만 ORF frame에는 영향을 주지 않은 경우

[2] 삽입으로 인해서 ORF의 3' 부분에서만 frame이 바뀐 경우

[3] 삽입으로 인해서 ORF가 완전히 망가진 경우

집단 발병을 일으켰던 Beijing family는 1995년에 처음 보고되었다[19]. 이 family에 속하는 균주들은 동물 모델에서 강한 발병력을 나타내고, 숙주의 사이토카인 분비를 억제시켜 Th1 타입의 면역반응의 유도를 억제하는 것으로 알려져 있다[20,21]. 이 균주가 강한 발병력을 나타낼 수 있는 유전적인 요인에 대하여 알려져 있는 바는 없지만, BCG 백신에 대한 저항성이 selective pressure로 작용하여 Beijing family의 전파력이 증대되고, 병원성이 증가하리라는 가설이 유력하다[19]. 한국 또한 BCG 예방접종을 하기 때문에 K 균주가 Beijing family와 연관되어 있을 가능성이 클 것으로 예상되어 K 균주와 Beijing 균주의 결손 부위를 비교해 보았다.

Affymetrix GeneChip microarray를 이용하여 H37Rv 균주와 Beijing 균주간의 다형성을 비교한 실험에서, Beijing 균주에서 21 부위의 LSP이 존재한다는 것이 밝혀졌다. 그런데 이중에서 RD105, RD149, RD152, RD207 부위에서 일어난 결손은 모든 Beijing family에서 일어나는 것으로 판명되었다[22]. 특히, RD105는 Beijing family가 아닌 다른 균주에서는 일어나지 않는 것으로 밝혀졌다.

이 4 부위에 대해서 K 균주에서도 결손이 일어났는지 비교 해 본 결과, Beijing 균주와 마찬가지로 K 균주에서도 결손이 일어났음을 확인할 수 있었다. 이 부위에 해당하는 결손부위는 K 균주에서 Del_2, Del_26, Del_27~8, Del_98~130으로 나타났다. 그러나, Beijing family에서 균주에 따라 특이적으로 나타나는 것으로 밝혀진 나머지 17개의 결손 부위에 대해서는 K 균주에서 결손이 일어나지 않았다.

3. K 균주와 CDC1551 균주의 결손 부위 비교

CDC1551 균주는 1995년에 미국에서 집단 발병을 일으켜 처음 분리되었다[13]. 이 균주는 성장속도가 빠르고, 환자와 짧은 접촉

을 가졌던 접촉자들을 높은 빈도로 sero-conversion시키는 것으로 확인되었다. 또한 감염시킨 쥐의 폐에서 Erdman 균주보다 100배나 많은 bacilli가 분리되었기 때문에, 처음에는 발병력이 높은 균주로 간주되었다. 그래서 임상분리 균주로는 처음으로 전체 염기서열 분석이 진행되었고, 2002년에 염기서열 분석이 완료되었다[12].

CDC1551 균주는 전체 4,403,765 염기쌍으로 이루어져 있고, H37Rv 균주와 비교했을 때 37 부위에서 결손이 일어났고 49 부위에서 삽입이 일어났음이 확인되었다. 그러나 최근 동물 실험 결과 CDC1551 균주가 다른 결핵 균주들에 비해 숙주에서 높은 면역 반응을 유도시키지만 다른 임상분리 균주들보다 발병력은 높지 않은 것으로 밝혀졌다[8].

발병력의 차이를 나타내는 임상분리 균주간에는 유전적으로 어떤 차이가 있는지를 알아보기 위해서, K 균주와 CDC1551 균주의 전체 염기서열을 비교해 보았다. 그 결과, CDC1551 균주에서 일어난 37 부위의 결손 중에서 5개 (Del_1, Del_10, Del_13, Del_19, Del_144), 49 부위의 삽입 중에서 9개 (Ins_8, Ins_18, Ins_19, Ins_20, Ins_38, Ins_71, Ins_88, Ins_89, Ins_142)가 K 균주에서도 일어났음을 확인하였다.

4. 임상분리 균주를 대상으로 PCR 수행

가. PCR primer의 고안

153개의 결손 부위와 124개의 삽입 부위 중에서 IS*6110*으로 인해 결손이 일어난 10개와, 3개의 삽입 부위에 대해서는 primer를 제작하지 않았다. 결손이나 삽입이 한 부위에 몰려서 일어나는 경우, 그리고 근접한 거리에서 일어난 경우에는 각 결손 혹은 삽입당 한 쌍의 primer를 제작할 수가 없었기 때문에, 여러 개의 결손 혹은 삽입을 한 쌍의 primer로 PCR을 하는 방법을 택하였다.

IS*6110*으로 인한 결손 혹은 삽입을 제외하고, 143개의 결손 부위를 확인하기 위해서 48쌍의 primer를 제작하였고, 121개의 삽입 부위를 확인하기 위해서 62쌍의 primer를 제작하였다. 이중에서 여러 개의 결손 부위를 하나의 primer로 PCR하는 경우는 3쌍 (Del_9~10, Del_42~60, Del_154~155)이었고, 여러 개의 삽입 부위를 하나의 primer로 PCR하는 경우는 5쌍 (Ins_10~11, Ins_57~58, Ins_123~125, Ins142~143, Ins_149~150)이었다.

먼저 H37Rv와 K 균주의 genomic DNA를 template로 하여 각 primer에 대한 PCR을 수행하였다. 이때, annealing temperature와 extension time을 조절하여 PCR 조건을 정했다. 조건이 정해진 primer에 대해서는 나머지 임상분리 균주의 genomic DNA도 포함시켜 다시 PCR을 수행하였다. 그리고 조건이 잡히지 않은 경우에는 좀 더 바깥쪽으로 primer를 다시 제작하여 PCR을 수행해 조건을 잡은 다음 임상분리 균주들을 대상으로 PCR을 수행하였다.

나. PCR template의 선정

임상분리 균주간의 유전적 다형성을 알아보기 위해 앞서 결과에서 밝혀낸 결손 혹은 삽입 부위를 대상으로 하여 primer를 제작하고, H37Rv 균주와 K 균주, 그리고 국내외에서 분리된 38개의 임상분리 균주를 template로 선정하였다.

각 임상분리 균주는 IS*6110* 탐침자를 이용한 RFLP에서 독특한 결과를 나타내는데, 그 결과에 따라 10개의 band를 가지면서 K 균주와 유사하게 나타나는 경우는 K family로, band는 10개이지만 pattern이 다를 경우에는 M family로, 그 외 band의 개수와 pattern이 다양하게 나타나는 균주에 대해서는 unique 균주로 표시하였다. 본 실험에서 사용한 임상분리 균주들의 RFLP 결과는 그림 1에 정리되었다.

다. 결손 혹은 삽입 부위에서 임상분리 균주들의 PCR 수행

본 실험에서는 H37Rv 균주와 K 균주에서 조건이 잡혔던 결손 54 부위 (primer 34쌍), 삽입 58 부위 (primer 52쌍)에 대해 PCR을 수행하였다. PCR 결과는 H37Rv 균주를 상단에 놓고, K 균주를 하단에 배치한 후, 나머지 임상분리 균주들은 각 결손 혹은 삽입 부위를 PCR한 산물의 크기에 따라 H37Rv 균주와 같은 크기를 보이는 것에서부터 K 균주와 같은 크기를 보이는 것의 순으로 정리하였다. 그리고 PCR 산물의 크기가 다양하게 나오는 것은 H37Rv 균주와 K 균주 사이에 배치하였다 (그림 2, 그림 3).

Erdman 균주, CDC1551 균주, 그리고 A, a, b, d, 13, 18 균주들은 각 결손 혹은 삽입 부위에서 H37Rv와 대체로 유사하게 나타났다. 그러나 RFLP 결과에서 K family와 M family로 판명이 났던 임상분리 균주들은 PCR 결과에서 unique 균주들과 확연한 차이를 보이지 않았다.

라. 결손 부위에서 PCR 결과 분석

K 균주에서만 결손이 일어난 부위는 Del_3, Del_16, 그리고 Del_18 부위였다. 이 부위에서 일어난 결손으로 인해서 영향을 받는 유전자는 Rv0193c, Rv0194, Rv0973c, Rv0974c, Rv1059였는데, 이들은 각각 hypothetical protein, ABC transporter, acetyl-/propionyl-CoA carboxylase를 coding하고 있었다. 같은 K family 중에서도 K 균주에서만 이 부위에서 결손이 일어났기 때문에, 이 유전자가 K 균주에 특이적으로 작용하여 표현형에 변이를 일으켰을 가능성이 있을 수 있다.

K family와 M family에서만 결손이 일어난 부위는 Del_29, Del_36이었다. 이 부위에는 Rv1787과 Rv2129c가 위치하고 있으며, 각각 PE, PPE family와 oxidoreductase를 coding하고 있다.

77번 균주를 경계로 하여 H37Rv 균주와 같이 PCR 결과에서

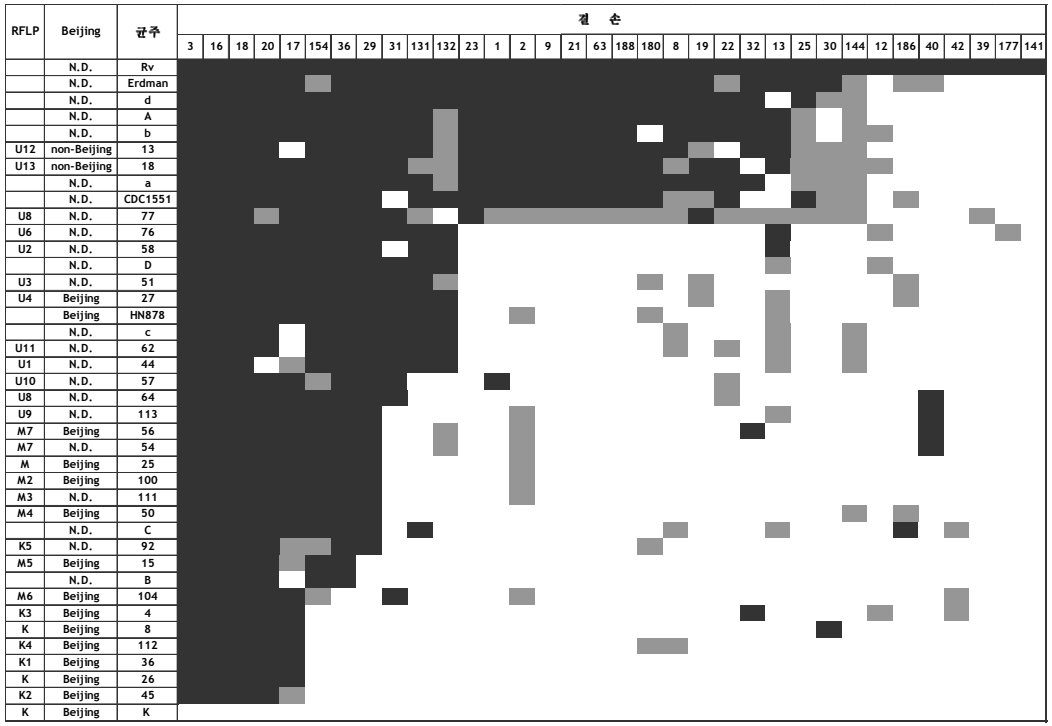

그림 2. 결손 부위에서 임상분리 균주간의 다형성. 검은색은 PCR 산물의 크기가 H37Rv와 같게 나온 것, 흰색은 K 균주와 같게 나온 것, 회색은 다양한 크기의 PCR 산물이 나온 것이다.

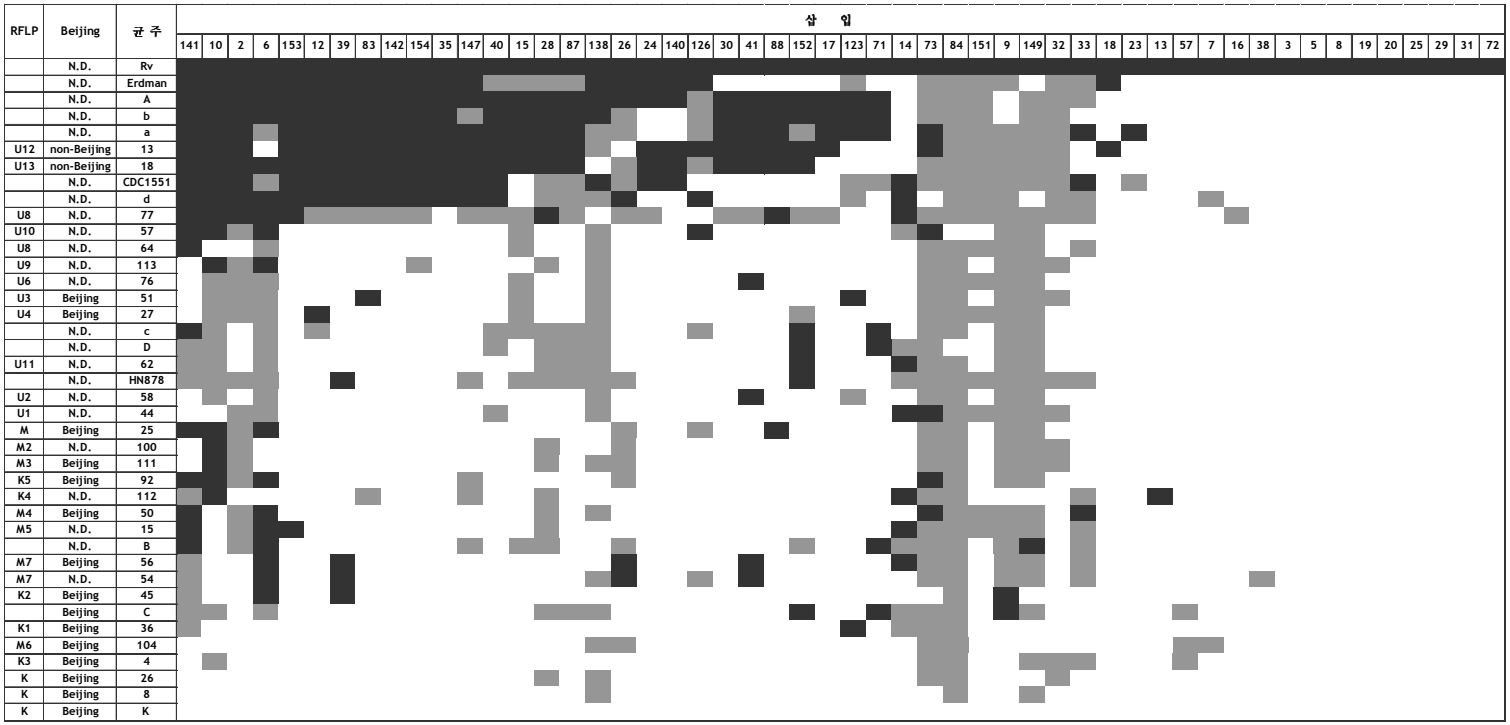

그림 3. 삽입 부위에서 임상분리 균주간의 유전적 다형성. 검은색은 PCR 산물의 크기가 H37Rv와 같게 나온 것, 흰색은 K 균주와 같게 나온 것, 회색은 다양한 크기의 PCR 산물이 나온 것이다.

결손이 일어나지 않았던 Erdman, CDC1551, A, a, b, d, 13, 18 균주들을 제외한 다른 균주들에서는 모두 결손이 일어난 부위는 Del_1, Del_2, Del_9, Del_21, Del_23, Del_63, Del_188이었다. 이 부분 또한 PE, PPE family, hypothetical protein을 coding하고 있는 것으로 밝혀졌다.

나머지 언급되지 않은 결손 부위들은 균주에 따라 PCR 산물의 크기가 다양하게 나타나서 위의 분류에 포함시킬 수 없었다.

마. 삽입 부위에서 PCR 결과 분석

결손에서와 마찬가지로 변이가 H37Rv 균주에서부터 unique 균주, K 균주의 방향으로 진행되고 있는 것으로 보였지만, K와 연관된 균주들에서 특이적으로 삽입이 일어난 것은 확인 할 수 없었다.

Ins_2, Ins_15, Ins_28, Ins_35, Ins_40, Ins_87, Ins_142, Ins_147, Ins_154 부위에서 Erdman, CDC1551, A, a, b, d, 13, 18 균주를 제외한 모든 균주들에서 삽입이 일어난 것으로 확인이 되었다. 이 부위에 위치하고 있는 유전자들 또한 PE, PPE family, hypothetical protein들을 coding하고 있었다.

그리고, Ins_9, Ins_32, Ins_73, Ins_84, Ins_138, Ins_149, Ins_151 부위에서는 H37Rv를 제외한 다른 균주에서 다양한 크기의 삽입이 이루어진 것으로 나타났다.

5. 결손 혹은 삽입 부위의 염기서열 분석

PCR 산물의 크기가 같을 경우에 다른 임상분리 균주에서도 같은 부위에 같은 크기의 결손과 삽입이 일어났는지 확인하고, 또 PCR 산물의 크기가 다를 경우에 각 균주에 따라 염기서열 상에 어떠한 차이가 있는지를 확인하기 위하여 임상분리 균주에서 결손

8 부위, 삽입 13 부위에 대한 염기서열 분석을 수행하였다.

그 결과, PCR 산물의 크기가 같은 경우에 각 임상분리 균주에서도 동일하게 결손이나 삽입이 일어났음을 확인할 수 있었다.

PCR 산물이 다양한 크기로 나타났던 경우에는 염기서열 분석 결과, 대부분 K 균주에서 결손 혹은 삽입이 일어났던 위치에서 결손 혹은 삽입이 일어났지만 균주에 따라 결손이나 삽입이 일어난 부위가 1~2 bp 정도 차이가 나기도 했다. 그리고 다양한 크기의 염기서열이 결손/삽입되면서 frame-shift mutation을 일어날 수도 있고, truncation 될 수도 있기 때문에, 기능을 잃어버릴 될 가능성을 가지고 있었다.

Del_21은 PE-PGRS 단백질을 coding하고 있는 Rv1243c 중간에 105 bp가 결손 된 것으로써, 결손이 일어났어도 ORF가 frame-shift mutation을 일으키지 않고 유지된 경우이다. PCR 결과를 보면, Erdman, A, a, b, d, 13, 18, CDC1551 균주를 제외한 다른 임상분리 균주에서 모두 결손이 일어났음을 알 수 있다 (그림 4). 그리고 77번 균주는 PCR 산물이 2 band로 나타났는데, 정확히 H37Rv와 K 균주에서의 크기와 동일하게 나오는 점으로 볼 때, 균주를 배양할 때 오염이 되었을 가능성이 있다.

PCR 산물의 크기가 같을 때 염기서열에서도 동일하게 나오는지를 확인하기 위하여, PCR 산물을 정제한 후 염기서열 분석을 수행한 후, H37Rv와 K 균주의 염기서열을 기준으로 하여 A, HN878, 25, 4, 8, 26, 45 균주의 염기서열을 비교하였다 (그림 5). PCR 산물이 H37Rv와 같았던 A 균주에서는 결손이 일어나지 않은 것을 확인하였고, 그 외 균주들에서는 K 균주와 마찬가지로 105 bp가 결손된 것을 확인하였으나, K 균주에서 결손이 일어난 부위보다 2 bp 뒤쪽에서부터 결손이 되었음을 알 수 있었다. 그러나 아미노산의 서열을 비교해 본 결과 전체 frame에는 영향을 미치지 않았음을 확인하였다.

Ins_010, 011 역시 PE-PGRS 단백질을 coding하고 있는 Rv0746의 중간에 각 62 bp, 26 bp의 삽입이 일어난 경우이다. K 균주에서는 이 부위의 삽입으로 인해서 Rv0746이 truncation되었음이 확인되었다. PCR 결과에서 보면 일부 unique 균주와 K와 연관된 균주들에서 다양한 크기의 삽입이 일어난 것을 볼 수 있었다 (그림 6).

다양한 크기의 삽입이 일어났을 때 실제로 염기서열에는 어떤 차이가 있는지를 알아보기 위해 역시 PCR 산물을 정제한 후 염기서열 분석을 수행하였다 (그림 7). PCR 결과에서 H37Rv 균주와 같은 결과를 보였던 CDC1551 균주는 삽입이 일어나지는 않았으나 21곳에서 single nucleotide polymorphism (SNP)을 보이는 것을 확인 할 수 있었다. 또, 113번 균주는 삽입 부위 앞, 뒤쪽으로 12 bp에 결손이 일어났고, 4곳에서 SNP가 일어났음이 확인되었다. 한편, PCR 결과에서 다른 크기가 나왔던 76번 균주는 K 균주와 같은 크기의 삽입이 일어났음을 확인하였다. 27번과 HN878 균주에서는 삽입이 일어나지 않고, 135 bp가 결손되어 있음을 알 수 있었다. 4번 균주는 30 bp가 삽입되었고, 8번 균주는 K와 같이 삽입이 일어났음을 확인하였다. 이런 인해 단백질에 생긴 영향을 알아보기 위해 아미노산 서열을 비교해 보았더니, 위치는 달랐지만 임상분리 균주들에서 모두 truncation이 일어났음을 알 수 있었다.

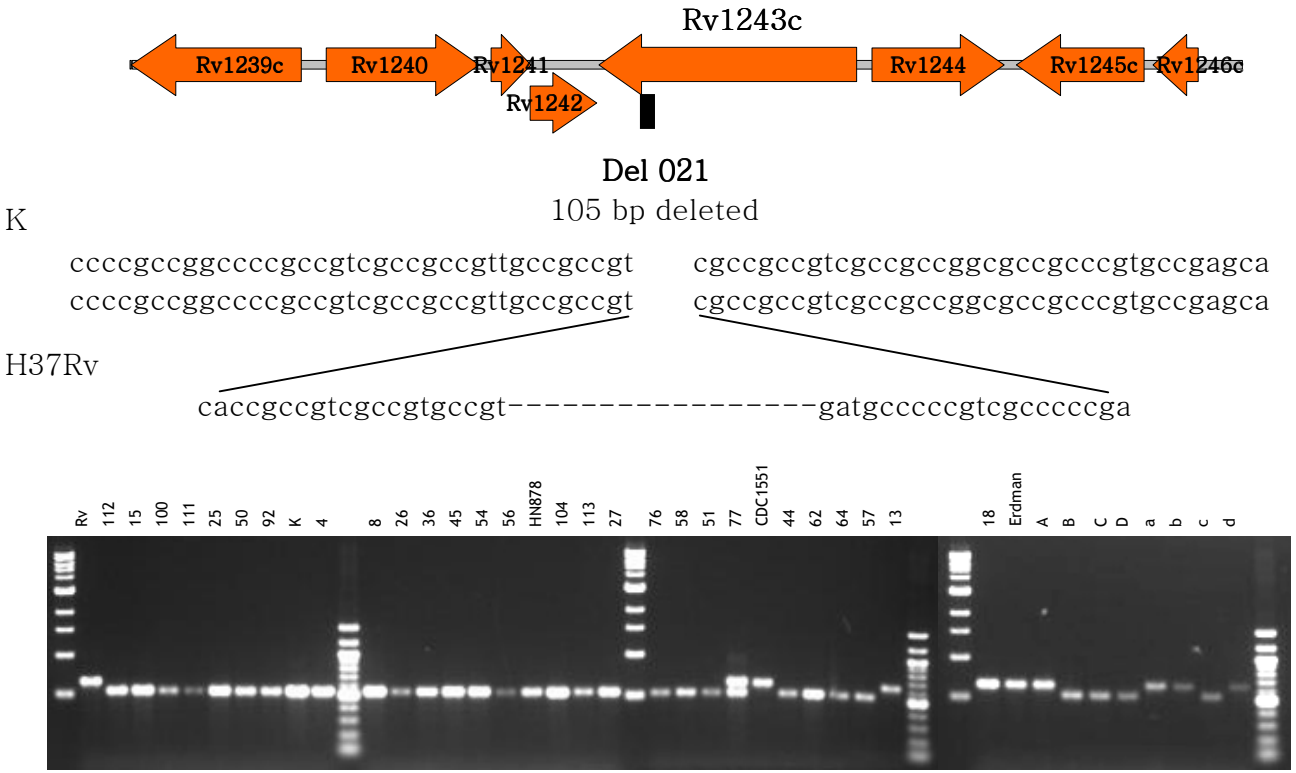

그림 4. Del_21의 PCR 결과. K 균주에서 Rv1243c의 중간에 105 bp의 염기서열이 결손되었다. 40개의 임상분리 균주를 대상으로 PCR을 수행한 결과, H37Rv 또는 K 균주와 같은 크기의 산물만 확인하였다.

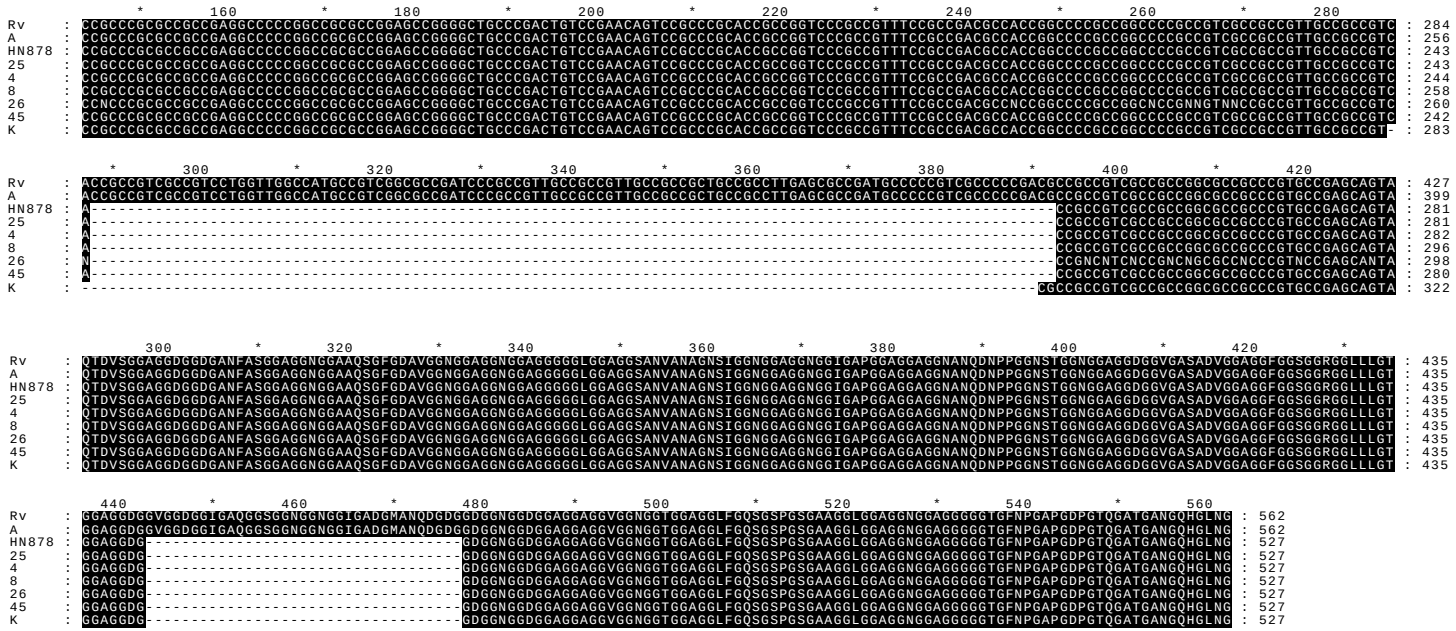

그림 5. Del_21의 염기서열 분석과 Rv1243c의 아미노산 서열 분석. K 균주와 같이 HN878, 25, 4, 8, 26, 45 균주에서도 105 bp의 염기서열이 결손되었다. 이로 인해 35개의 아미노산 또한 결손되었다. 결손된 염기서열과 아미노산은 점선으로 표시하였다.

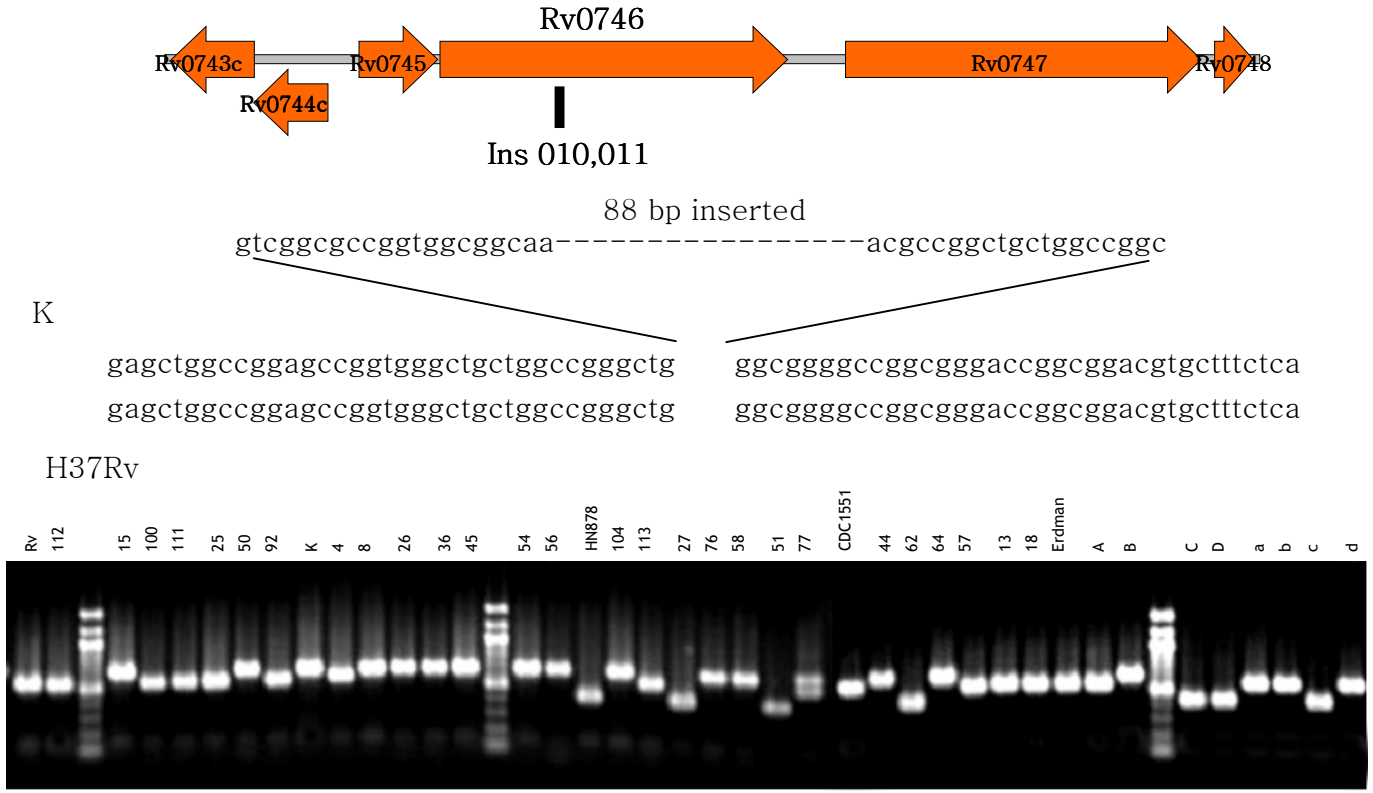

그림 6. Ins_10, Ins_11의 PCR 결과. K 균주에서 Rv0746 중간에 각 62 bp, 26 bp의 염기서열이 삽입되었다. 40개의 임상분리 균주를 대상으로 PCR을 수행한 결과, 균주에 따라 다양한 산물을 얻을 수 있었다.

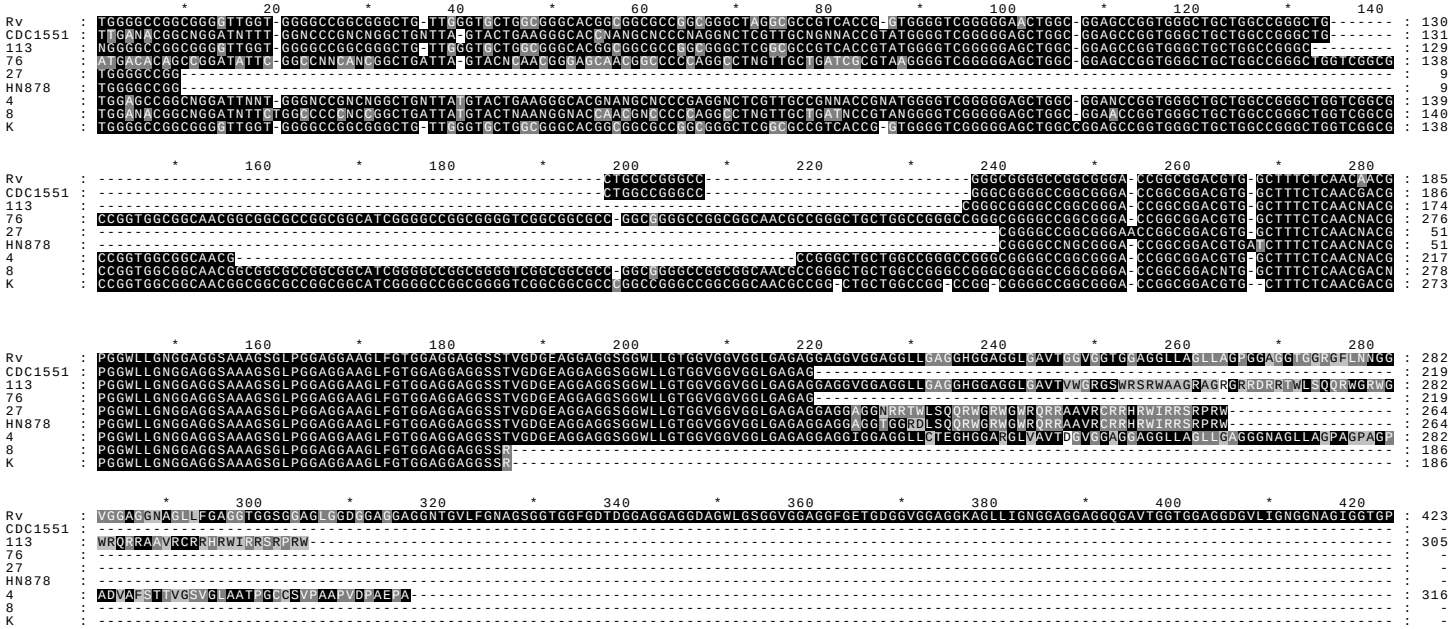

그림 7. Ins_10, Ins_11의 염기서열 분석과 Rv0746의 아미노산 서열 분석. 삽입 부위에서 균주에 따라 다양한 염기서열을 가지고 있고, 이로 인해 아미노산 서열도 다양하게 나타났음을 알 수 있었다. 결손된 염기서열과 아미노산은 점선으로 표시하고, 다르게 나타난 염기서열과 아미노산은 회색으로 표시하였다.

IV. 고 찰

균주간의 유전적인 다양성에 대한 연구는 진화와 발병력 연구에 있어 유용한 접근 방법이다. 이제까지 결핵균의 비교유전체 연구를 위해 gene-chip을 이용한 연구, subtractive hybridization과 RFLP 분석을 통한 LSP 연구들이 수행되어져 왔으나, 이 방법들은 낮은 해상도의 분석으로 끝난다는 한계가 있었다. 또한, 전체 염기서열을 분석하여 유전적인 다형성을 비교하는 방법은 그 대상이 몇 균주에 국한되고 비용과 시간이 많이 소요된다는 단점이 있었다.

그래서 본 연구에서는 최근에 전체 염기서열 분석이 완료된 K 균주의 염기서열을 결핵균 표준 균주인 H37Rv의 전체 염기서열과 비교하여, 결손과 삽입이 일어난 부위를 확인하고, 그 부위를 대상으로 PCR을 수행하여 국내에서 분리된 임상분리 균주에서 일어난 유전적인 다형성을 알아보고자 하였다. 이 방법은 삽입에 대한 정보도 얻을 수 있을 뿐만 아니라, 염기서열 분석을 통하여 각 임상분리 균주에서의 유전정보까지 획득할 수 있다. 또한 발병력과 관계되어 있는 유전자를 연구하는데 있어, 균주 특이적으로 일어난 유전자 수준의 변이에 대한 정보를 제공해 줌으로써 기능 연구를 위한 목적 유전자의 발굴 가능성을 열어줄 수도 있을 것이다.

H37Rv 균주와 비교된 K 균주는 한국에서 높은 빈도수로 감염이 되며, 고등학교에서 집단발병을 일으킨 적이 있어, 발병력이 높은 균주일 것이라 예상되고 있으나, 그 유전체에 대한 연구는 전무한 실정이었다. 그래서 호흡기 감염세균 유전체 연구센터에서 K 균주의 전체 염기서열 분석을 진행하여, 2004년에 전체 4,379,571 bp, 4,089개의 유전자로 이루어져 있다는 것을 밝혀내었다.

위의 염기서열 분석 결과를 바탕으로 H37Rv 균주와 K 균주

의 전체 염기서열을 비교하여, 10 bp이상의 결손이나 삽입이 일어난 부위를 확인하였다. 그 결과, K 균주에서 10 bp~9,562 bp의 다양한 크기의 결손이 153 부위에서 일어났고, 124 부위에서는 10 bp~4,999 bp 크기의 삽입이 일어났음을 밝혀냈다. 결손 중에서 135개는 유전자 내에서 일어났고, 18개는 intergenic sequence에서 일어났으며, 삽입 중에서 99개는 유전자 내에서, 나머지 25개는 intergenic sequence에서 일어난 것으로 확인되었다.

결손 혹은 삽입에 의해서 영향을 받은 유전자들의 대부분은 PE, PPE family에 속하는 단백질과 아직 기능이 밝혀지지 않은 단백질 그리고 transposase 등이었다.

PE, PPE family는 결핵균 유전체에서 약 10%를 차지하고 있으며, 결핵균에서만 특이적으로 나타난다[3,23]. PE family는 100여개의 유전자들로 구성이 되어 있고, ~110개의 아미노산이 보존되어 있는 N-terminal domain을 공유하고 있으며, 특징적으로 8~9번째 위치에서 Pro-Glu (PE) 모티프를 가지고 있다. PPE family는 67개의 유전자들로 이루어져 있고, N-terminal에 180개의 아미노산이 보존되어 있다. PPE family에도 8~10번째 위치에서 Pro-Pro-Glu (PPE) 모티프가 존재한다. 이들 유전자군은 C-terminal domain에 따라서 각각 PE-polymorphic GC-rich sequence (PGRS)와 PPE-major polymorphic tandem repeat (MPTR)의 하위그룹으로 나누어진다.

PE, PPE family의 기능은 대부분 밝혀져 있지 않지만, 최근에 항원다양성, 숙주세포와의 결합, 대식세포 내에서 생존, 육아종 내에서 결핵균의 존속에 관여할 것이라는 보고들이 있었다[24]. 또한 signature-tagged mutagenesis (STM), micro-array 등의 연구를 통해 PE-PGRS 단백질이 발병력과 관계가 있을 것이라는 증거들이 나왔고[25], 동물 실험에서도 PE region이 세포 매개 면역반응을 유도한다는 것과, Gly-Ala-rich PGRS domain이 체액성 면역반응을

유도한다는 것이 밝혀져 왔다[26].

이런 점들로 볼 때 K 균주에서 결손 혹은 삽입이 주로 PE, PPE family에서 일어났다고 하는 것은 이 부위에서 유전적 변이가 일어나면서 K 균주의 발병력 표현형에 영향을 주었을 가능성을 제시한다. 따라서 후행 실험으로 이 유전자들에 대한 기능 연구를 진행하면 K 균주가 발병력을 가지게 되는 원인을 알아내는 실마리가 될 수 있으리라 기대된다.

한편, 다른 임상분리 균주와 K 균주와의 염기서열상에는 어떤 차이가 있는지를 알아보기 위해, 기존에 H37Rv 균주를 대상으로 비교하여 결손 혹은 삽입에 대한 정보가 밝혀져 있는 Beijing 균주와 CDC1551 균주를 K 균주의 결손 혹은 삽입과 비교해 보았다.

Beijing 균주는 동물실험에서 다른 균주에 비해 보다 심각한 조직학적 변화를 보였으며, 사망률 또한 높은 것으로 나타났다[20]. 사이토카인 발현에 있어서도 TNF-α 같은 경우는 초기에는 증가했으나, 빨리 감소하는 경향을 보였고, IFN-γ 발현도 다른 균주에 비해 늦어졌다. 이런 점으로 볼 때, Beijing 균주는 면역 반응 유발을 억제시킴으로써 보다 심각한 병리학적 소견을 보인다고 할 수 있다.

Beijing family가 BCG 예방접종을 실시하는 나라에서 발병률이 높은 점으로 볼 때, K 균주도 Beijing family에 속할 가능성이 있다. 또한, spoligotyping 결과에서 한국에서 분리된 임상분리 결핵균들 중 다수가 Beijing family에 속하였으므로, K 균주와 Beijing 균주가 염기서열상의 공통된 변이를 가지고 있을 것으로 예상되었다.

비교 결과, 모든 Beijing family에서 발견되는 4 부위의 결손이 K 균주에서도 존재하였으나, 그 외 다양하게 나타나는 결손들에 대해서는 발견할 수 없었다. 따라서 K 균주가 Beijing family와 같은 조상에서 유래되었으나, 결손 혹은 삽입이 일어나면서 분리되

어 나왔을 가능성이 있을 수 있었다. 하지만 어떤 시점에서 분리가 된 것인지에 대해서는 알 수 없었다. 이를 알아내기 위해서는 Beijing family에 속하는 균주들과 비교를 해서, 어떤 결손 혹은 삽입을 시점으로 해서 분리가 되는지 확인해야 할 것으로 생각된다.

Beijing 균주와 같이 일어난 결손으로 인해서 영향을 받은 ORF는 Rv0071~Rv0074, Rv1572c~Rv1587c, Rv1754c, Rv1755c, Rv1758~Rv1765c, Rv2816c~Rv2820c였다. 이중에서 Rv1754에서 Rv1765까지의 ORF는 IS*6110*으로 인한 결손의 hot spot 지역이라는 것이 밝혀졌다[27]. Rv1755c는 *plcD* 유전자로 *plcA*, *plcB*, *plcC*와 함께 phospholipase C를 coding하고 있다. 대식세포에 감염시켰을 때 *plc* 유전자의 발현이 upregulation 되었다는 점으로 볼 때 phospholipase C가 감염에 중요한 역할을 할 것이라 보여진다[28]. 또, *plc* 돌연변이체는 감염된 쥐의 폐와 신장에서 성장이 약화됨을 보이며, persistent phase에서 중요하게 작용한다는 보고가 있었다[29]. *plcABC*에 비해서 *plcD*가 균주간에 다양한 변이를 보이고, K 균주에서는 *plcD*가 완전히 결손되었다는 점으로 볼 때, 앞으로 임상분리 균주간에 이 유전자가 어떤 차이가 있는지 확인하는 것이 필요할 것이라 생각된다.

CDC1551 균주와 비교를 했을 때 5 부위에서 결손과 9 부위에서 삽입이 K 균주에서도 동일하게 일어났음을 알 수 있었다. 이들 결손 혹은 삽입에 의해 영향을 받은 ORF는 Rv0020c, Rv0355c, Rv0747, Rv1087, Rv1319c, Rv1435c, Rv1450c, Rv2024c, Rv2407, Rv3159c, Rv3281, Rv3507이었다. 이들 대부분도 PE, PPE family에 속하는 유전자들이었다. CDC1551 균주는 감염되었을 때 숙주의 면역반응을 급격하게 유발시키지만, 다른 결핵 균주와 비교했을 때 상대적으로 발병력은 높지 않은 것으로 나타났었다[8]. 그 점을 염두에 두고 K 균주와의 비교 결과를 보면, K 균주와 CDC1551 균주에서 상이하게 일어난 결손 혹은 삽입 부위

중에 K 균주의 발병력에 관계된 유전자가 있을 것이라 추측되며, 이는 후행 실험을 진행하는데 목적 부위가 될 수 있을 것으로 기대된다.

다음으로, 임상분리 균주간의 유전적인 다형성을 알아보기 위해 K 균주를 포함한 한국에서 분리된 임상분리 균주 36개와 외국에서 분리된 균주인 Erdman, CDC1551, HN878, 그리고 결핵균 표준 균주인 H37Rv의 genomic DNA를 template로 하여 각 결손 혹은 삽입 부위를 PCR하였다.

결손 54부위 (primer 34쌍), 삽입 58부위 (primer 52쌍)에 대한 PCR을 수행하여, 결과가 H37Rv와 같게 나온 것부터 K와 같게 나온 것의 순으로 정리했을 때, Erdman 균주와 CDC1551 균주, 그리고 A, a, b, d, 13, 18 균주가 대체적으로 H37Rv와 같은 결과를 보였다. RFLP 결과상으로 unique했던 균주들과 K family와 M family에 속하는 균주들에서는 각 결손 혹은 삽입 부위에 따라서 PCR 결과가 다양하게 나와 유전적인 변이가 더 다양하게 일어났음을 알 수 있었다. 이런 경향은 결손보다는 삽입에서 두드러졌다. 본 실험에서는 H37Rv와 K 균주를 기준으로 하여 결과를 해석하였기 때문에 그림 2와 3에서 보면 변이의 방향이 H37Rv 균주에서부터 K 균주로 진행되는 것으로 보이지만, 실제로는 K 균주에서 unique 균주로, 또는 다른 방향으로 변이가 진행되고 있을 가능성도 있다.

K 균주에서만 결손이 일어났던 Del_3, Del_16, Del_18 부위는 hypothetical protein과 ABC transporter를 coding하고 있는 유전자와 *accA2*, *accD2*, *fadE13* 유전자가 위치하고 있다. *accA2*와 acc*D2* 유전자는 각각 acetyl-/propionyl-CoA carboxylase α와 β를 발현시키는데, 이 단백질들은 mycolic acid biosynthesis에 관여하고 있다. 결핵균의 세포벽 구성 성분인 mycolic acid가 숙주의 면역반응을 유발하고 발병력과도 관계가 있다는 기존의 연구[30, 31]

로 미루어볼 때, 이 유전자가 결손 되었다는 것은 주목할만하며, 앞으로 유전자의 기능 연구를 수행시 목적 부위가 될 수 있을 것으로 생각된다. K family와 M family에서만 특이적으로 결손이 일어났던 부위는 Del_29와 Del_36였고, 77번 균주를 경계로 하여 H37Rv 균주와 구분되었던 균주들에서만 결손이 일어났던 부위는 Del_1, Del_2, Del_9, Del_21, Del_23, Del_63, Del_188이었다. 이 부위에 일어난 결손으로 인해 영향을 받은 유전자들은 hypothetical protein, oxidoreductase, PE, PPE family, ABC transporter, maturase를 coding하고 있었다. 이 부위들 중에서 Beijing family에서 공통적으로 일어난 Del_2는 제외하고, 나머지 결손 부위들은 어떤 시점을 계기로 해서 동시적으로 유전적 변화가 초래된 것으로 보이며, 그 영향에 대한 실험이 진행되어야 할 것이다.

삽입은 결손에 비해서 균주에 따라 다양하게 일어났지만, K 균주와 K, M family 특이적으로 일어난 삽입은 없는 것으로 판명이 되었다. 그리고 H37Rv 균주와 K 균주의 염기서열을 비교하여 삽입이 일어났다고 밝혀진 염기서열을 BLAST에서 검색해 보았으나 database에서 일치하는 유전자가 없는 것으로 나타났다. 그리고 삽입이 일어나면서 유전자가 truncation되거나 disruption 되었기 때문에, 삽입으로 인해서 새로운 기능을 가진 유전자가 삽입이 되기 보다는 결손에서처럼 기존의 유전자를 망가트리는 역할을 한 것으로 생각된다. PCR 결과에서 보면 결손에서와 마찬가지로 삽입에서도 77번 균주를 경계로 H37Rv 균주와 구분이 되었고, 이들 균주에서 Ins_2, Ins_15, Ins_28, Ins_35, Ins_40, Ins_87, Ins_142, Ins_147, Ins_154 부위의 삽입이 이루어졌음을 알 수 있었다. 이 중에서 Ins_2, Ins_35은 intergenic sequence에 삽입이 이루어진 것으로, 주변의 ORF와 promoter에는 영향을 주지 않았다. 나머지 삽입으로 인해 영향을 받은 유전자들은 PE, PPE family,

hypothetical protein을 coding하고 있었다. 이들 부위 또한 앞으로 기능 연구를 위한 목적 부위가 될 수 있을 것으로 생각한다.

8개의 결손 부위와 13개의 삽입 부위를 대상으로 염기서열 분석을 수행한 결과, PCR 산물의 크기가 동일하게 나왔던 경우에는 염기서열 분석에서도 동일한 결과를 얻을 수 있었고, PCR 산물의 크기가 다양하게 나왔던 경우는 K 균주에서 결손 혹은 삽입이 일어난 부위 근처에 다양한 크기의 결손 혹은 삽입이 일어났음을 확인하였다. 염기서열을 바탕으로 하여 아미노산의 서열을 비교했을 때, 결손 혹은 삽입에 의해 frame-shift mutation이 생기거나 truncation이 일어난 것을 확인하였고, 임상분리 균주에 따라 다른 아미노산 서열을 가질 수 있음을 알 수 있었다. 임상분리 균주에서 모든 결손 혹은 삽입 부위에 대해 염기서열 분석을 실시하지는 않았기 때문에, 앞으로 임상분리 균주에서도 위의 결과에서 기능 연구의 목적 부위로 선정한 부위들에 대한 염기서열 분석을 수행하고 K 균주와 비교해 봐야 할 것이다.

결손 혹은 삽입으로 인해 영향을 받은 ORF들을 분석하면서 야기된 문제점은 본 연구에서는 LSP을 관찰하는 것을 목적으로 하여 실험을 진행했기 때문에, SNP으로 인해 생기는 염기서열상의 변이에 대해서는 취약했다는 점이다. 그래서 결손 혹은 삽입에 의해 영향을 받았다고 분석했던 ORF 중에서도 LSP이 아닌 SNP에 의해서 단백질이 truncation되거나 disruption된 경우들이 존재했다. 이런 점을 보완하여 각 균주간의 유전적 다형성이 미친 영향을 확인하기 위해서, 위에서의 결과를 바탕으로 앞으로의 기능 연구를 위한 목적 부위를 선정하고, 그 부위를 대상으로 SNP에 대한 조사를 추가적으로 수행한 후 기능 연구를 진행해야 할 것이라고 생각한다.

V. 결 론

1. K 균주의 전체 염기서열을 결핵균 표준 균주인 H37Rv의 전체 염기서열과 비교하여 153개의 결손과 124개의 삽입이 일어났음을 확인하였다.

2. K 균주에서 일어난 결손 중 135개는 유전자 내에서 일어났고, 18개는 intergenic sequence에서 일어났다. 삽입 중에서 99개는 유전자 내에서, 25개는 intergenic sequence에서 일어난 것으로 판명되었다.

3. 결손과 삽입으로 인해 disruption 되거나 truncation 되는 등의 영향을 받은 유전자는 각 96개와 58개였으며, 이들이 coding하고 있는 단백질들은 대부분 PE, PPE family, hypothetical protein, transposase이었다.

4. K 균주와 Beijing 균주를 비교하여, 모든 Beijing family에서 일어나는 것으로 보고된 4개의 결손이 K 균주에서도 일어났음을 확인하였다. 그러나 Beijing family에서 균주에 따라 다양하게 나타나는 결손은 K 균주에서 일어나지 않았다. 따라서 K 균주는 Beijing 균주와 같은 조상을 가지고 있으나, 변이가 일어나면서 분화되어 나왔을 가능성이 있을 수 있다.

5. H37Rv 균주와 비교하여 CDC1551 균주에서 일어난 결손과 삽입 중에서 5개의 결손과 9개의 삽입은 K 균주에서도 일어났음을 확인하였다. 그러나 K 균주와 CDC1551 균주간의 공통되는 경향은 찾을 수 없었다.

6. H37Rv 균주와 K 균주를 포함하여 국내외에서 분리된 임상분리 균주 40개의 genomic DNA를 template로 하여 결손 54부위, 삽입 58부위에 대하여 PCR을 수행하였다.

7. PCR 결과에서 Erdman 균주와 CDC1551 균주, 그리고 A, a, b, d, 13, 18 균주는 대체적으로 H37Rv와 같은 결과를 보였다.

8. RFLP 결과에서 unique했던 균주들과 K, M family에 속하는 균주들에서 각 결손 혹은 삽입 부위에 따라서 PCR 결과가 다양하게 나와 유전적인 변이가 다양하게 일어났음을 알 수 있었다.

9. K 균주에서만 결손이 일어났던 Del_3, Del_16, Del_18과 K family에서만 일어났던 Del_29, Del_36 부위는 앞으로 K 균주의 병독성에 기여하는 유전자를 찾기 위한 기능 연구에서 목적 부위가 될 수 있을 것으로 기대된다.

10. Del_1, Del_2, Del_9, Del_21, Del_23, Del_63, Del_188, Ins_2, Ins_15, Ins_28, Ins_35, Ins_40, Ins_87, Ins_142, Ins_147, Ins_154 부위는 PCR 결과에서 H37Rv 균주와 뚜렷한 차이를 보이므로, 각 임상분리 균주들이 표현형의 차이를 보이는 원인을 연구하는데 목적 부위가 될 수 있을 것으로 예상된다.

11. 21개의 결손 혹은 삽입 부위에 대해서 임상분리 균주들의 염기서열을 분석한 결과, PCR 산물의 크기가 같을 경우에는 임상분리 균주에서도 동일한 크기의 결손 혹은 삽입이 일어났음을 확인하였다. 그러나 PCR 산물의 크기가 다양하게 나타난 경우에는 임상분리 균주에서 일어난 결손 혹은 삽입의 크기 또한 다양하였고, 이로 인해서 아미노산의 서열에도 영향을 준다는 것을 확인하였다.

참고문헌

1. Dye C, Scheele S, Dolin P, Pathania V, Raviglione MC. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. JAMA 1999;282:677-686.

2. 대한결핵협회. 2004년 결핵현황. 2004.

3. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 1998;393:537-544.

4. Thierry D, Brisson-Noel A, Vincent-Levy-Frebault V, Nguyen S, Guesdon JL, Gicquel B. Characterization of a *Mycobacterium tuberculosis* insertion sequence, IS*6110*, and its application in diagnosis. J Clin Microbiol 1990;28:2668-2673.

5. Thierry D, Cave MD, Eisenach KD, Crawford JT, Bates JH, Gicquel B, Guesdon JL. IS*6110*, an IS-like element of *Mycobacterium tuberculosis* complex. Nucleic Acids Res 1990;18:188.

6. Otal I, Martin C, Vincent-Levy-Frebault V, Thierry D, Gicquel B. Restriction fragment length polymorphism analysis using IS*6110* as an epidemiological marker in tuberculosis. J Clin Microbiol 1991;29:1252-1254.

7. Bellamy R, Beyers N, McAdam KP, Ruwende C, Gie R, Samaai P, Bester D, Meyer M, Corrah T, Collin M, Camidge DR, Wilkinson D, Hoal-Van Helden E, Whittle HC, Amos W, van Helden P, Hill AV. Genetic susceptibility to tuberculosis in Africans: a genome-wide scan. Proc Natl Acad Sci USA 2000;97:8005-8009.

8. Manca C, Tsenova L, Barry CE 3rd, Bergtold A, Freeman S, Haslett PA, Musser JM, Freedman VH, Kaplan G. *Mycobacterium tuberculosis* CDC1551 induces a more vigorous host response *in vivo* and *in vitro*, but is not more virulent than other clinical isolates. J Immunol 1999;162:6740-6746.

9. Musser JM. Antimicrobial agent resistance in mycobacteria: molecular genetic insights. Clin Microbiol Rev 1995;8:496-514.

10. Sreevatsan S, Pan X, Stockbauer KE, Connell ND, Kreiswirth BN, Whittam TS, Musser JM. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci USA 1997;94:9869-9874.

11. Kim SJ, Bai GH, Lee H, Kim HJ, Lew WJ, Park YK, Kim Y. Transmission of Mycobacterium tuberculosis among high school students in Korea. Int J Tuberc Lung Dis 2001;5:824-830.

12. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, Gwinn M, Haft D, Hickey E, Kolonay JF, Nelson WC, Umayam LA, Ermolaeva M, Salzberg SL, Delcher A, Utterback T, Weidman J, Khouri H, Gill J, Mikula A, Bishai W, Jacobs Jr WR Jr, Venter JC, Fraser CM. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J Bacteriol 2002;184:5479-5490.

13. Valway SE, Sanchez MP, Shinnick TF, Orme I, Agerton T, Hoy D, Jones JS, Westmoreland H, Onorato IM. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. N Engl J Med 1998;338:633-639.

14. Boddinghaus B, Rogall T, Flohr T, Blocker H, Bottger EC. Detection and identification of mycobacteria by amplification

of rRNA. J Clin Microbiol 1990;28:1751-1759.

15. Gutacker MM, Smoot JC, Migliaccio CA, Ricklefs SM, Hua S, Cousins DV, Graviss EA, Shashkina E, Kreiswirth BN, Musser JM. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics 2002;162:1533-1543.

16. Kato-Maeda M, Rhee JT, Gingeras TR, Salamon H, Drenkow J, Smittipat N, Small PM. Comparing genomes within the species *Mycobacterium tuberculosis*. Genome Res 2001;11:547-554.

17. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science 1999;284:1520-1523.

18. Musser JM, Amin A, Ramaswamy S. Negligible genetic diversity of *Mycobacterium tuberculosis* host immune system protein targets: evidence of limited selective pressure. Genetics 2000;155:7-16.

19. van Soolingen D, Qian L, de Haas P, Douglas JT, Traore H, Portaels F, Qing HZ, Enkhsaikan D, Nymadawa P, van Embden J. Predominance of a single genotype of

Mycobacterium tuberculosis in countries of east Asia. J Clin Microbiol 1995;33:3234-3238.

20. Lopez B, Aguilar D, Orozco H, Burger M, Espitia C, Ritacco V, Barrera L, Kremer K, Hernandez-Pando R, Huygen K, van Soolingen D. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotype. Clin Exp Immunol 2003;133:30-37.

21. Manca C, Tsenova L, Bergtold A, Freeman S, Tovey M, Musser JM, Barry III CE, Freedman VH, Kaplan G. Virulence of a *Mycobacterium tuberculosis* clinical isolate in mice is determined by failure to induce Th1 type immunity and is associated with induction of IFN-alpha/beta. Proc Natl Acad Sci USA 2001;98:5752-5757.

22. Tsolaki AG, Gagneux S, Pym AS, de la Salmoniere Y, Kreiswirth BN, van Soolingen D, Small PM. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J Clin Microbiol 2005;43:3185-3191.

23. Cole ST. Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. Microbiology 2002;148:2919-2928.

24. Ramakrishnan L, Federspiel NA, Falkow S. Granuloma-specific expression of Mycobacterium virulence proteins

from the glycine-rich PE-PGRS family. Science 2000;288:1436-1439.

25. Camacho LR, Ensergueix D, Perez E, Gicquel B and Guilhot C. Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. Mol Microbiol 1999;34:257-267

26. Delogu G, Brennan MJ. Comparative immune response to PE and PE-PGRS antigens of *Mycobacterium tuberculosis*. Infect Immun 2001;69:5606-5611.

27. Sampson SL, Richardson M, van Helden PD, Warren RM. IS6110-mediated deletion polymorphism in isogenic strains of *Mycobacterium tuberculosis*. J Clin Microbiol 2004; 42:895-898.

28. Agaisse H, Gominet M, Økstad OA, Kolstø AB and Lereclus D. PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. Mol Microbiol 1999;32:1043-1053.

29. Raynaud C, Guilhot C, Rauzier J, Bordat Y, Pelicic V, Manganelli R, Smith I, Gicquel B, and Jackson M. Phospholipase C are involned in the virulence of *Mycobacterium tuberculosis*. Mol Microbiol 2002;45:203-217.

30. Dubnan E, Chan J, Raynaud C, Mohan VP, Laneelle MA, Yu K, Quenard A, Smith I, and Daffe M. Oxygenated mycolic acids are necessary for virulence of *Mycobacterium tuberculosis* in mice. Mol Microbiol 2000;36:630-637.

31. Glickman MS, Cox JS, and Jacobs WR. A novel mycolic acid cyclopropane synthase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*. Mol Cell 2000;5:717-727.

ABSTRACT

Genetic polymorphisms among clinical strains of
Mycobacterium tuberculosis

Jiwon Kim

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Sang-Nae Cho)

Despite the efforts to combat tuberculosis it still remains one of the major public health problems, which forced World Health Organization (WHO) to declare tuberculosis a global emergency in 1993. Mycobacterial species belonging to *Mycobacterium tuberculosis* complex show 99.9% similarity in nucleotide sequences, but they differ widely in their host tropisms and phenotypes related to pathogenicity. Recently clinical strains of *M. tuberculosis* showing diversity in virulence phenotype have been reported, but the molecular basis for the phenotypic diversity among the clinical isolates remains poorly understood.

Based on the assumption that genetic differences among clinical strains account for the phenotypic variation, genetic polymorphism among recent clinical isolates of *M. tuberculosis* was investigated. Comparison of the whole genome sequence of *M. tuberculosis* K strain, one of the most prevalent strains in Korea, with that of *M. tuberculosis* H37Rv revealed the presence of 153 deletions and 124 insertions in the K genome relative to

the H37Rv genome. Those deletions and insertions affect 96 ORFs and 58 ORFs, respectively, which encode mostly PE, PPE family proteins, hypothetical proteins and transposases. DNA fragments corresponding 54 deletions and 58 insertions were amplified by PCR with specific primers and analyzed to find genetic variations among 40 clinical isolates.

Clinical isolate A, a, b, d, 13 and 18 showed similar insertions and deletions found in H37Rv, while other isolates showed variations similar to K strain. Del_3, Del_16 and Del_18 were found to be unique in K strain and Del_29 and Del_36 were shown to be conserved among members of K family, suggesting that variations in these loci may account for the hypervirulence of K strain. Nucleotide sequences of PCR products carrying 21 selected deletions and insertions showed that PCR products with the same size have identical nucleotide sequence for each deletion or insertion, while PCR products with variations in size have altered nucleotide sequences because of insertions or deletions, which result in changes in amino acid sequences.

In this study PCR amplification of targeted regions and sequence analysis have provided information on both insertions and deletions, in contrast to the conventional comparative genomic approaches based on DNA hybridization which identify deletions only. The information presented herein could contribute to the identification of *M. tuberculosis* genes associated with virulence of the pathogen.

Key words : *Mycobacterium tuberculosis*, clinical strains, genetic polymorphism