

비교언어학 및 유전학적 방법에 의한 한국어 기원의 탐구¹⁾

이상억 · 이정빈 · 김선영 · 천종식

(서울대 인문대 국어국문학과 · 의대 법의학교실 · 자연대 생명과학부)

1. 비교언어학적 연구 개관 및 유전학적 기존 연구결과 비교

한국어의 기원을 탐구하는 데에는 고고학적, 언어학적, 유전학적 방법의 세 가지가 있다. 본고에서는 후자의 두 가지 방법을 사용하여 이 문제에 접근해 보고자 한다.

한국어는 드라비디아어, 오스트로네시아어, 고아시아어, 인구어, 그리고 무엇보다 알타이어와의 가설적 연관성에 대해 다양하게 연구된 바 있다. 알타이어의 경우 한국어와 많은 공통된 특성을 공유하고 있는데, 이러한 유사점 중에 몇 가지가 실재적이고 유형론적인 것인가 하는 점은 논의의 여지가 있으며 한국어의 정확한 계통적, 유전적 친족 관계는 아직 의문으로 남아 있다.

이른바 알타이어학은 보다 엄밀하게 입증된 인도-유럽어학과 비교해 볼 때

1) 이 연구는 서울대학교 간접연구경비에서 지원되는 학제적 협력연구비로 책임 연구원 이상억 교수에 의하여 수행되었다. 초기 연구 여건 조사는 자연대 생명과학부 김선영 교수와 김봉철 박사, DNA sequencing은 의대 이정빈 교수 지도하에 법의학 교실 김기범, 장영길 연구원, DNA 분석 결과의 비교 작업은 자연대 생명과학부 천종식 교수 지도하에 허인에 연구원의 노고로 이루어졌다. 실험 자료로 쓰인 모발과 타액 채취에는 터키 유학생 술탄 웨라, 몽골 유학생 바타르수흐 앵흐툴, 외대 노어과 강덕수 교수(야쿠트와 에벤키 시료)의 협력이 지대하였다.

매우 불안정하다. 공통된 고유의 어휘부와 같은 충분한 증거가 부족한 까닭에 만주-통구스어, 몽골어, 터키어와 같이 유형론적으로 알타이어족으로 분류되는 언어들과의 친족관계와 기원을 규명하기가 어렵다. 뿐만 아니라 어떤 학자들은 한국어, 일본어, 오키나와어, 아이누어 역시 이 어족에 속한다고 주장한다. 과거 19세기에는 몇몇 학자들이 주로 공통된 유형론적 특성에 근거하여 핵심적 알타이어족이 우랄어족과 연결된다고 주장하여 우랄-알타이어 가설을 주장했으나 최근에는 대부분의 학자들이 이 가설을 포기하였다.

20세기에 Ramstedt 박사 등이 제기한 알타이어 가설은 결코 이 분야의 모든 이들에게 받아들여지는 것이 아니며 이를 둘러싸고 많은 논쟁의 여지가 남아 있다. 알타이어의 개별적 “분지(branches)” 내의 여러 언어들 이 유전적으로 서로 관련된다는 것은 사실이지만 여러 가지 분지들 자체가 유전적으로 서로 관련되는가 하는 점은 여전히 상당히 의문스럽다는 바에 대하여 대부분 동의하고 있는 듯하다.

예를 들어 이른바 터키어라고 불리는 모든 언어들 이 서로 관련되어 있다고 하는 것이나 다양한 몽골어들이 서로 관련되어 있다고 하는 것에는 의문의 여지가 없다. 그러나 몽골어와 터키어가 유전적으로 서로 관련되는가 하는 점은 언어학자들 사이에서 상당히 의문시되는 것이다. Sir Gerald Clauson과 같은 학자는 터키어와 몽골어가 오랜 시간동안 서로 지속적으로 접촉해 왔으며 그것은 매우 오래 전의 선사 시대로 거슬러 올라갈 수 있는 것이라는 점을 지적한다. 뿐만 아니라 이 두 언어 집단이 많은 어휘적 요소와 심지어 일부 형태론적 요소까지도 서로 차용해 왔다는 것도 명백하다. 그러나 이들 언어가 하나의 조어/원시어에서 파생해 나온 것이라면 조어에 가까워질수록 더 많은 유사점이 있을 것으로 기대되지만 이 두 집단의 초기문자 기록을 비교해 보면 실제로는 이들 사이에 그 정도의 유사점이 존재하지 않는다. 곧 이 두 가지 언어 집단 사이에는 오랜 시간 동안의 밀접한 접촉에 기인한 언어 수렴(convergence)이 있었지만 두 집단이 유전적으로 관련되는 것은 아니라는 것이다. 이는 모든 알타이어 집단 사이에 공유된 기본적 어휘를 포함하는 동계어(cognate)가 왜 그렇게 적은가 하는 점을 설명해 줄 수 있다.

다른 한편으로 Nicholas Poppe나 Roy Andrew Miller와 같은 학자들은 다른 학자들이 다소 공인된 어족에 대해 적용하였던 것과 거의 동일한 방법으로 알타이어들 사이에 순환적 음운 대응관계를 설정하고 원형적 알타이어의 음운 체계와 어휘적, 형태론적 요소 등을 재구하였다. (최근 Starostin (1991)은 원형적 알타이어가 기원적 6세기에 멸망하였다고 주장하였다.) 그들은 단지 세 가지 핵심적 알타이어 모두에 있어서 적어도 몇 가지의 차용이 일어났다는 증거 때문에 그들의 작업이 폐기되는 것은 불합리하며 다른 언어에 비해 알타이어들 사이의 계통적 관계에 대한 증거에 대해 너무 엄격한 기준을 부가하는 것은 부당하다고 주장한다.

한국어, 일본어, 오키나와어와 아이누어를 알타이어족에 포함시키는 것은 보다 더 논쟁의 여지가 있다. 왜냐하면 알타이어 가설을 지지하는 사람들조차도 이들 언어를 알타이어족에 포함시키려는 주장에서 제시된 증거를 수용하지 않기 때문이다. 일본어와 알타이어 사이의 관계와 관련하여 얼마 전 Susumu Ono (1970)라는 일본학자가 일본어는 알타이어를 상층언어로 하고 오스트로네시아어를 하층언어로 하여 이루어진 것이라고 주장했던 것을 언급할 필요가 있다.

한국어를 일본어와 비교해 보면 형태론적으로나 통사론적으로 놀라울 정도의 유사성이 있음을 발견하게 된다. 그러나 이 두 언어는 하나의 공통된 계통적 근원으로부터 파생되었다기보다는 평행하게 발전해 온 것으로 보인다. 비록 Samuel Martin (1991)과 같은 학자들은 한국어와 일본어, 오키나와어가 계통적으로 관련되어 있다는 상당히 흥미로운 증거를 제시하였으나 이러한 계통적 연관성에 대해 학계는 여전히 회의적이다. 문제가 되는 언어들은 모두 전형적인 SOV언어이며 상당수의 문법적 특징을 공유하고 있다. 그러나 한국어와 일본어 사이의 동족관계를 밝히기는 쉽지 않으며 이것은 설사 한국어와 일본어가 정말로 관련되어 있다고 하더라도 그 관계가 생각보다 먼 것임을 짐작하게 한다.

모든 비교 언어학자들이 여러 가지 어족의 개체군을 분류하고 그들의 계통적 친족관계를 고찰하는 것처럼, 필자 또한 한국어와 언어학적으로 관련됐으

리라 가정되어온 언어들, 주로 알타이어를 함께 살펴 보고자 한다. 우리는 중국, 시베리아, 중앙 아시아에 분포하는 만주-통구스어[Man (=Manchu(rian)), †Juchen; Nanaj (=Nanay, Hezhe/Ulchi, Ulcha), Akani, Birar, Gold(i), Kile, Olcha, Orok(i), Samagir; Udihe (=Udehe), Oroch; Orochon (=Oroqen), Evenki (=Ewenke), Manegir, Solon; Even (=Lamut), Negidal 등]와 몽골어[Khalkha, Khorchin, Ordos, Chakhar, Urat, Kharchin-Tumut, Ujumuchin, Bargu; Oirat, Kalmyk; Buriat (=Buryat); Da(g)ur; Monguor, Yellow Uighur (=Hui), Santa, Pao-(ng)an; Moghol 등] 등의 알타이 언어들에서 가능한 한 여러 가지 개별 데이터와 종족별 데이터를 수집해야 할 것이다. 뿐만 아니라 중앙아시아, 코카서스, 소아시아 등지의 터키어[Turkish, Turkmen(ian), Azerbaidjani, Khalaj, Gagauz, Qashqay; Uighur (=Uyghur), Sala, Uzbek; Bashkir, Karachay, Karaim, Kumyk, Tatar, Crimean Tatar, Baraba; Nogai, Kara-Kalpak, Kazakh, Kirg(h)iz (=Kyrgyz); Yakut, Dolgan; Khakas, Altai, Northern Altai, Chulym, Shor, Tuva, Karagas and Chuvash]를 고려해야 한다. 한편 일본어, 오키나와어(=Ryukyuan), 아이누어에 대해서도 참작을 해야 한다. 나아가 고아시아어(=Paleosiberian: Ket, Yukaghir, Nivkh (=Gilyak), Koryak, Kamchadal (=Itelman), Chukchi)와 시베리아 에스키모어에 대해서도 언급할 수 있을 것이다.

마지막으로 최근 문제가 되는 것은 산둥지방의 동이어(Tongyi)의 기원이 무엇이며 그들이 어떻게 한국어와 관련되는가 하는 점이다. 그러나 언어학계 내부의 논쟁이 아직 어떠한 명백한 증거도 제시해 주지 않는 까닭에 우리는 다른 방법, 곧 유전학적 방법을 사용하여 이러한 가설들을 평가할 수 있다.

인류의 기원 연구에서 여성, 특히 최초의 '이브'에 대한 강조는 버클리 과학자들인 앨런 윌슨과 마크 스톤킹 그리고 레케바 칸(1987)이 사용해 온 유전학을 바탕으로 시작되었다. 유전학 연구대상의 하나인 인류와 유인원이 언제 갈라졌는지는 수백만 년이라는 연대를 필요로 하나, 현생인류의 기원은 1백만년 이하의 연대를 필요로 할 뿐이다. 이렇게 시간의 범위가 좁기 때문에 정밀한 분자시계를 필요로 하였고, 이 시계를 미토콘드리아라는 동물의 각 세포조직에서 세포 에너지를 만들어내는 작은 조직 안에서 찾았다. 미토콘드

리아는 유전물질인 핵산(DNA)을 가지고 있는데 이 핵산은 세포핵에 있는 핵산보다 10배나 빠른 속도로 돌연변이를 일으킨다. 이 신속하게 변이를 일으키는 핵산이 대단히 짧은 시간을 알려 준다.

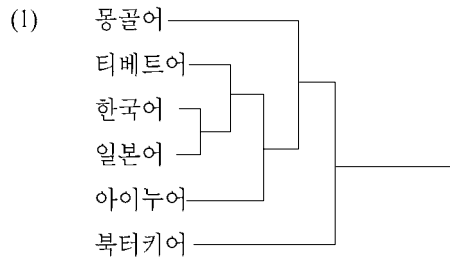
그러나 미토콘드리아의 특성에 관한 것만이 설명의 전부는 아니다. 여성의 난자가 남성의 정자와 결합하여 태아를 형성할 때 진귀한 현상이 일어난다. 정자는 난자 속에 자신의 염색체를 주입하지만, 자궁 안에서 정자가 왕성하게 움직이도록 동력을 공급하는 미토콘드리아는 뒤에 남는다. 태아가 만들어질 때는 언제나 어머니의 미토콘드리아만이 아이에게 전달된다. 따라서 우리가 지니고 있는 미토콘드리아의 핵산은 오직 여성 쪽에서 전해 받은 유전물질로서, 우리들의 어머니, 할머니, 증조할머니 등을 거쳐 완전히 우리들 자신의 역사 중 가장 먼 시기까지 거슬러 올라간다. 머나먼 과거를 들여다 보기 위해 미토콘드리아 유전학을 이용함으로써 지구상에서 누구누구가 서로 관계 있는 지를 규명할 수 있다.

세계를 지리적 구분에 따라 아프리카, 아시아, 유럽, 오스트레일리아 및 뉴기니 등으로 나누어 이 지역 출신의 150명의 여성들로부터 미토콘드리아의 핵산을 수집하여 분석한 끝에 월슨과 그의 동료들은 미토콘드리아 핵산의 유형을 찾아 내었다. 월슨은 가장 놀라운 사실은 다양한 그룹 사이에서 거의 차이를 볼 수 없다는 점이다. 차이가 거의 없다는 것은 그들이 아주 최근에 서로 갈라졌다는 것을 의미한다고 말하고 있다. 그룹들 간에 보이는 미세한 차이는 인류에게 분화가 어떻게 일어났는지를 보여준다. 전체적으로 닮아 있음에도 불구하고 분석 결과 두 개의 주요 그룹이 나타났다. 한 그룹은 아프리카를 대표하며, 또 한 그룹은 기타 지역에서 온 개인들을 포함하고 있다. 그리고 아프리카 그룹이 모든 그룹 중 가장 오래 전에 형성되었음이 분명하다고 그는 말하고 있다.

월슨과 그의 동료들은 이 자료와 다른 자료를 바탕으로, 현생인류의 공통 조상은 약 20만 년 전에 아프리카에서 살았다고 결론을 내렸다. 더 나아가 그들은 이들이 아프리카를 떠나 유럽과 아시아 대륙으로 이주하였을 때 그들은 자신들보다 더 원시적이던 그 지역 주민들과 거의 섞이지 않았다고 주장

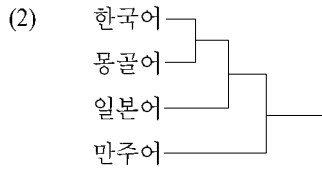
하고 있다. 다른 말로 표현하자면, 미토콘드리아 핵산 가설은 현생인류가 한 지역에서 진화한 후 다른 지역으로 이주하였으며 그 과정 중에 먼저 살던 원시인류 집단들을 교체하였다는 노아의 방주 가설을 지지한다고 할 수 있다.

한편 이러한 유전학적 실험 연구를 뒤이어 시행한 L. Luca Cavalli-Sforza, Paolo Menozzi and Alberto Piazza(1994)에 의하면, 120 대립유전자(allele)의 빈도 조사 결과가 보고 되어 있는데 42인종 중 알타이어 관계 부분만을 보이면 아래의 <계통도 1>과 같다. 본 절 끝에 제시할 <계통도 1>의 내용을 우선 골자만 보면, (1)과 같다.

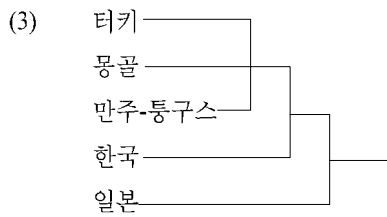


비교언어학의 일반적 추정과 달리 한-일 관계를 매우 가깝게 판정했고, 알타이어들 속에 중국어와 한 어족인 티베트어가 개재되어 있는 결과가 아주 이상하다. 그 다음에도 아이누어가 몽골어나 터키어보다 오히려 한-일어에 더 가깝다고 위치되어 있는 점이 역시 전통적 비교언어학의 일반적 추정과 달라 이채롭다. 위 표의 변두리에는 몽골어와 사모예드(우랄어), 또 북티키어와 에스키모, 척치(Chukchi, 고아시아어)를 짝 채워 붙여 놓았다. 그러나 이 조사 결과는 한국어와 티베트어가 결코 가까울 수 없고, 아이누어도 언어적으로 유사성이 없어 받아들이기 어렵다.

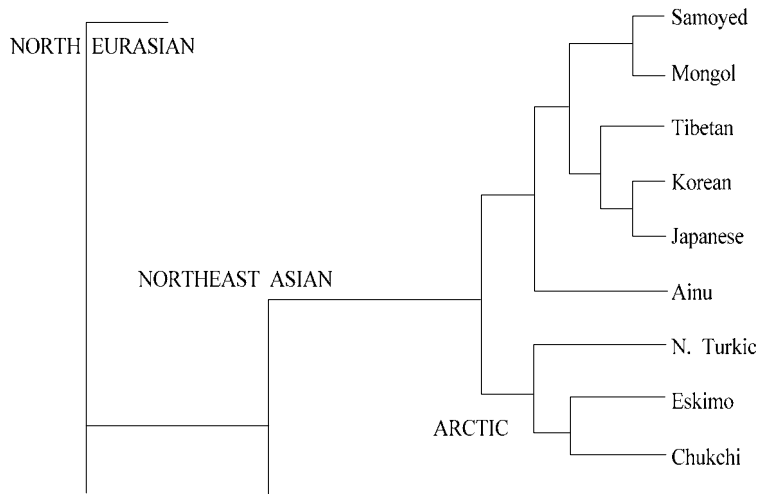
한편 Jin Feng et al.(1996)의 유전학적 방법의 연구 결과를 보면 일본어와 몽골어를 가장 가깝게 그리고 그 다음에 만주어를 인근의 옆가지로 결론짓고 있다. (아래 <계통도 2> 참조) 또는 한국어와 몽골어를 가장 가깝게 놓고 그 뒤에 일본어와 관계되는 것으로도 보니 결론은 대략 다음 (2)와 같다.



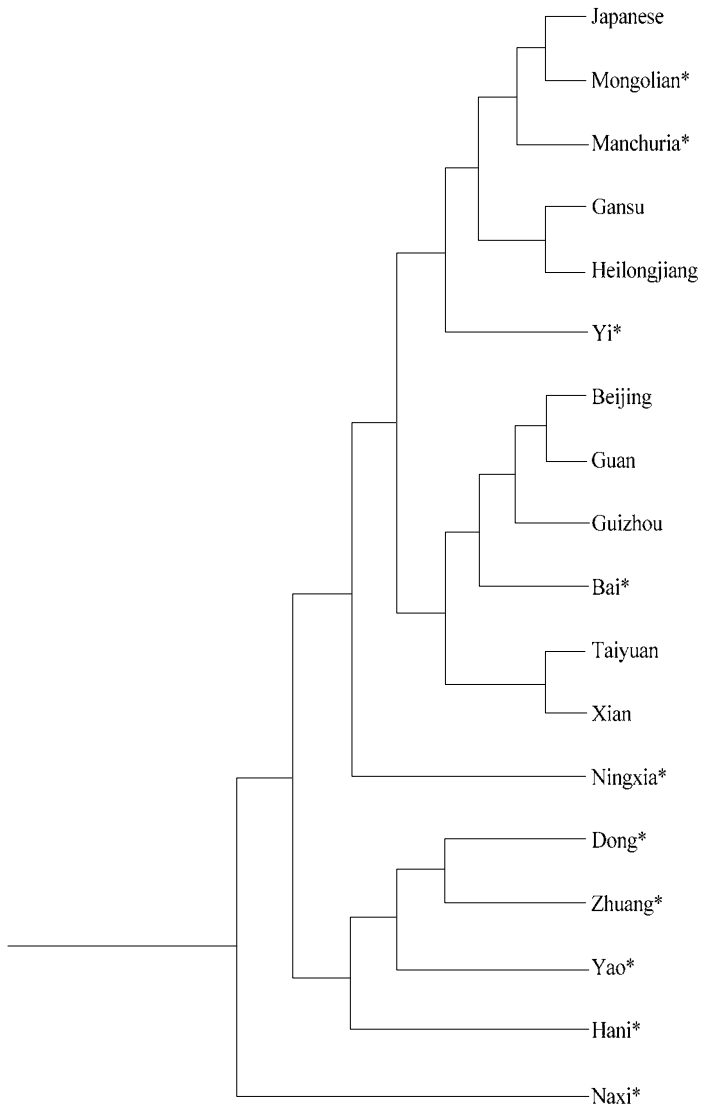
그러나 비교언어학에서 일반적으로 추정해 왔던 내용은 아래 표 (3)과 같다. (2)의 위로부터 한국어, 만주-통구스어, 몽골어, 일본어의 순으로 갈라져 위치한 것과 비슷한 결과이었으니, 위 (2)의 Jin Feng의 중국내 소수민족에 대한 연구 결과와는 다른 점이 크게 있어 흥미롭다.



<계통도 1> Altaic을 중심으로 한 Eurasiatic Super-phylum의 일부



<계통도 2> 일본어와 17개 소수민족어(*표)의 4적색 세포효소 위치에 의한 수형도(dendrogram) 일부



2. 유전학적 재검토를 위한 실험 방법론

위에서 보았듯이 기존의 유전학적 연구결과가 일반적인 비교언어학적 가설과 접근되기 어려우므로 이를 재검토하기 위하여, 우리 연구팀은 직접 한국, 일본, 몽골, 터키, 퉁구스인의 모발과 타액을 수집하여 다음과 같은 절차의 실험을 하여 보기로 한다.

유전학적 측면에서 인류집단의 역사와 진화를 이해하기 위해서는 유용한 정보를 지닌 다형적인 유전적 marker를 가능한 많이 조사해야 한다. 지금까지 인류집단의 유전적 변이에 관한 선구적인 연구는 혈청단백질(serum protein)이나 전기영동적(electrophoresis) 변이에 근거하여 시작되었다. 1970년대와 1980년대부터 개발되기 시작한 DNA 재조합 기법은 연구의 초점을 유전자 산물인 단백질에서 새롭게 보다 다양한 유전적 변이를 분석할 수 있는 restriction fragment length polymorphism(RFLP) 방향으로 전환하게 했다. 이후 DNA 염기(nucleotide) 서열 분석법과 polymerase chain reaction(PCR) 방법에 의해 DNA상에 염기치환의 정도를 정확하고도 신속히 분석할 수 있게 됨으로써 인류집단 간의 과거 진화적 계통관계에 관한 보다 많은 정보를 얻을 수 있게 되었다. 이 정보들을 언어상의 계통 규명에 이용하려는 것이 본 연구의 과제다.

아래 설명되는 내용처럼, 우리는 mtDNA (mitochondrial DNA)를 위주로 연구를 하고, 연구여건상 여력이 있으면 Y-염색체(chromosome) [또 mtDNA (mitochondrial DNA)] 위에 있는 STR marker(곧 DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385I/II)처럼 선호되어 온 유전적 marker로부터 이끌어 낸 정보를 더 고찰할 것이다. 또한 STR marker와 mtDNA 연쇄 뿐만 아니라 Y-염색체 위의 SNP marker에 대해서도 기회가 있으면 더 연구할 것이다.

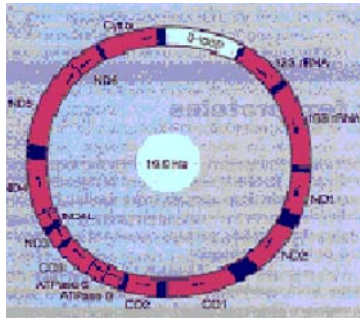
2.1. Mitochondria와 Y-chromosome의 다형성 (polymorphism)에 기초한 인종/언어의 기원연구

2.1.1. mitochondrial DNA(mtDNA)의 특성과 분석

미토콘드리아는 중요한 세포질에 존재하는 세포소 기관으로 유기물을 산화시켜 에너지 (ATP)를 생산하는 일을 한다. 자신의 DNA와 DNA복제 및 전사를 위한 여러 가지 자체 효소들을 가지며, 고유의 리보솜(일종의 단백질을 만드는 공장)을 가지고 있다.

mtDNA는 원형의 이중나선구조로 16.6kb의 길이로 되어 있다. mtDNA는 dir 20개의 tRNA와 단백질을 암호화하는 유전자로 구성되어 있다. 그런데, 약 1.7 kb 정도의 D-loop라고 부르는 조절부위를 가지고 있는데, 이곳에는 두개의 과변이 부위인 HVR (Hypervariable Region) I 및 II가 약 1 kb의 길이로 존재한다. HVR I 및 II의 염기의 변이는 다른 부위에 비해 5-10배 정도 빠른 속도로 변하기 때문에, 이 지역은 개인식별 및 혈통을 조사하기 위한 유전자 감식에 매우 유용하다.

그림 1. mitochondria DNA 지도



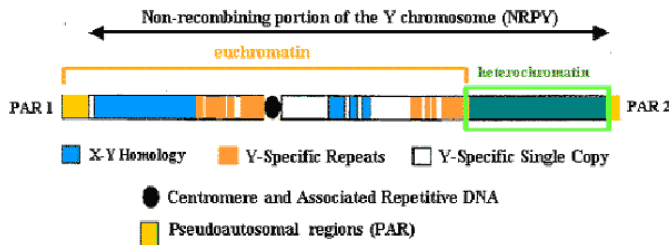
사람은 정자와 난자가 만나 융합된 하나의 수정란으로부터 출발한다. 그런데, 수정시 정자는 핵만 난자와의 융합에 참여하고, 세포질은 전혀 수정란 속으로 들어가지 못한다. 즉, 수정란의 세포질은 100% 난자로부터 유래한 것이

기 때문에, 이 속에 존재하는 미토콘드리아 역시 100% 본래 남자의 세포질에 있던 것이다. 그러므로 미토콘드리아 DNA 역시 철저히 모계유전의 방식으로 다음 세대로 전달된다. 즉 어머니의 mtDNA는 그녀의 아들이나 딸에게 그대로 전달되며, 그 딸은 또 자신의 자식에게 mtDNA를 전달한다. 그렇지만, 모계유전의 특성으로, 아버지의 mtDNA는 아쉽지만 자기 아들이나 딸 누구에게도 자신의 mtDNA를 전달해 주지 못한다. 이와 같은 mtDNA의 모계유전의 특성은 수십년간 헤어져 모녀 및 모자간의 혈육확인을 가능하게 할 뿐 아니라, 인류의 기원을 추정하는 데도 매우 유용한 정보를 제공해 주고 있다. 현재 지구상에 살고 있는 모든 인류가 중앙 아프리카의 단일 지역에서부터 기원하여 지금에 이르렀음을 밝힌 것도 mtDNA의 변이를 이용한 유연관계 추정으로부터 나왔다.

2.1.2. Y-염색체의 다형성(polymorphism)에 기초한 인종의 기원연구

인류의 Y-염색체는 극히 일부의 pseudoautosomal region을 제외하고 부계로만 유전되며 교차되지 않는 까닭에 과거에 일어난 돌연변이들이 연관상태로 그 정보를 보존하게 된다. 따라서 본 연구에서는 인류집단의 과거 역사를 이해하는데 유용한 지표로 활용되고 있는 Y-염색체의 다형적 marker와 combination haplotype을 언어학적인 측면에서 살펴보고자 한다.

그림 2.



인류의 Y-염색체는 heterochromatin 부위와 euchromatin 부위로 구분된다. Euchromatin 부위는 크게 4 종류의 서열로 구성되어 있다: (1) Y-염색체 단완

의 말단 부위를 포함하여 X 염색체와 상동성을 가진 서열, (2) 동원체와 그 주변의 반복서열, (3) Y-염색체 특이 반복서열, (4) Y-염색체 특이 single copy 서열. 특히 Y-염색체가 인류의 진화연구에 유용한 지표로 이용되는 데는 교차가 일어나지 않는, NRPY(non-recombination portion of the Y chromosome) 부위가 있기 때문이다. 즉, NRPY는 전형적으로 부계 유전되는 부위로써 X 염색체와 교차가 일어나지 않는다.

Mitochondrial DNA(mtDNA)가 모계 유전되는 것과 같이 NRPY는 부계로만 유전되기 때문에 정상적인 핵형(46, XY)을 지닌 남자는 단일 부계를 통하여 자신의 계통을 추적 확인할 수 있다. 예를 들면, Y-염색체의 특정 부위에 한 돌연변이가 일어났다면 그 유전자 좌(locus)에서 이러한 돌연변이 대립인자는 세대를 통하여 교차 없이 다른 유전자들과 연관상태를 그대로 유지하게 된다. 따라서 현재 확인되는 모든 Y-염색체는 이전에 일어났던 돌연변이의 정보를 연관상태로 보존하고 있기 때문에 Y-염색체 DNA를 통하여 인류는 부계를 중심으로 단일 조상의 추적이 가능하다[Casanova et al., 1985].

이와는 달리 X 염색체와 상(常)염색체(autosomal chromosome)는 거의 매 세대마다 교차에 의해 재조합 되기 때문에 연관상태가 빈번히 바뀌게 되며, Y-염색체처럼 단일 조상으로부터 전달된 것이 아니라 수많은 조상들로부터 재조합된 것이다.

2.1.3. Y-염색체의 다형적 marker와 combination haplotypes

Genome 내에서 10~50 bp 단위로 직렬 반복된 motif를 지닌 유전자 좌는 흔히 VNTRs, 또는 minisatellites라 하며, 이와는 달리 1~6 bp의 직렬 반복 motif를 나타내는 경우는 STRs, 혹은 microsatellites라 한다. 일반적으로 VNTRs는 인류의 genome 내에서 제한적으로 분포되어 있을 뿐만 아니라 다형의 정도가 비교적 낮다. 그러나 STRs는 전체 genome 내에 광범위하게 분포되어 있으며 다형의 정도가 높고, 특히 PCR 방법에 의해 유전자의 typing이 간결하고 정확하기 때문에 유전자 mapping, 개인식별, 진화 유전학 분야에 유용하게 사용된다[Edwards et al., 1992; Tautz, 1989].

지금까지 인류집단을 대상으로 가장 집중적으로 조사된 Y-염색체 microsatellites marker는 DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393와 DYS385I/II 이다.

(4) Y-linked microsatellites polymorphisms

Global Data Base Locus	Locus Name	Comments	Reference
DYS385	UT588	CAAA/GGAA repeat	Jobling & Tyler-Smith, 1995
DYS392	-	ATA repeat	Roger et al., 1996
DYS19	27H39LR	GATA repeat	..
DYS389-I	-	TCTG/TCTA repeat	..
DYS389-II	-	TCTG/TCTA repeat	..
DYS390	-	TCTG/TCTA repeat	..
DYS391	-	TCTG/TCTA repeat	..
DYS393	-	AGAT repeat	..

2.2. 미토콘드리아DNA(Mitochondria D-loop region)의 염기서열 분석 결정

mtDNA D-loop region에 특이적으로 결합하는 프라이머(primer)를 이용하여 PCR 반응으로 충분한 양의 D-loop region DNA 조각을 확보한다. PCR 산물은 아가로스 젤상에서 전기영동하여 원하는 크기의 밴드만을 잘라낸 후 시판되는 DNA정제 키트를 이용하여 정제한다. 염기서열 분석은 서울대 의대 법의학 교실에 있는 자동 염기서열 분석 장비인 ABI377 혹은 ABI3100을 이용하여 수행했다.

그리고 법의학 교실에서 이미 분석해 놓은 한국인과 일본인의 염기서열 자료도 이 연구에 이용하기로 하였다.

(5) Y-chromosome STR 분석

Primer list

Locus	motif	Repeat No	Allele No	Size	Diversity	PCR Primer
DYS19	(GATA) _n	10-19	10	174-210	0.72	Primer A: C T A C T G A G T T C T G T T A T A G T Primer B: A T G C C A T G T A G T A G A C A 6) Primer A2: G T T A T A T A T A T A G T T T A G 3) Primer B2: G T T A A G A G A G T G T C A C T A 3)
DYS389I ¹⁾	(GATA) _n (GACA) _n	9-17	9	235-263	0.61	Primer A: C C A A C T C T C A T C T G T A T T A T C T A T Primer B: T C T T A T C T C C A C C A C C A G A Primer 2B: T T A T C C C T G A G T A G T A G A A G A T 5)
DYS389II ¹⁾	(GATA) _n (GACA) _n	26-34	9	355-383	0.75	see DYS389I
DYS390	(GATA) _n (GACA) _n	17-28	11	191-227	0.61	Primer A: T A T A T T A C A C A T T T T T G G G C C Primer B: T G A C A G T A A A A T G A A C A C A T T G C
DYS391	(GATA) _n	7-14	8	271-299	0.49	Primer A: C T A T T C A A T C A T A C A C C C A Primer B: G A T T C T T T G T G G T G C T G
DYS392	(A T T) _n	6-16	11	233-263	0.52	Primer A: T C A T T A A T C T A G C T T T T A A A A C A A Primer B: A G A C C C A G T T G A T G C A A T G T
DYS393	(GATA) _n	9-16	8	108-136	0.34	Primer A: G T G G T C T T C T A C T T G T G T C A A T A C Primer B: A A C T C A A G T C C A A A A A T G A G G
DYS385	(G A A A) _n	7-23	81	352-412	0.85	Primer A: A G C A T G G G T G A C A G A G C T A Primer B: G G G A T G C T A G G T A A A G C T G Primer 2B: C C A A T T A C A T A G T C C T C C T T T C 4)
YCAII ²⁾	(C A) _n	1-9	35	144-160	0.67	Primer A: T A T A T T A A A T A G A A G T A G T G A Primer B: T A T C G A T G T A A T G T T A T A T T A

2.2.1. DNA 추출

PCR cycle 조건

pre-incubation 95도 12 분 / 28 cycles [95도 30초, 53 60도 1 분, 72도 20 초] /
final extension step 72도 10 분.

Electrophoresis conditions

ABI 310 sequencer (POP4, short capillary, filter set A): 2 microliter PCR 산물
과 14 microliter formamide, 0.5 microliter PE Rox 500. Run time 24 min.

위 조건에서 일차적으로 한국인 216명에 대해서 Phenol extraction의 방법으로 DNA를 추출하였다. 이 중 130명을 2차적으로 추리기로 한다.

2.2.2. 중합효소연쇄반응에 의한 mitochondrial 유전자 증폭

미토콘드리아의 HVI과 HVII부위를 증폭시키기 위해 기존에 알려져 있는 염기 서열을 분석하여 primer을 제조하였다. PCR 반응은 1 U of Taq DNA polymerase, 250 μ M each dNTP, 50 mM Tris-HCl (pH 8.3), 40 mM KCl, 1.5 mM MgCl₂를 포함하는 AccuPower™ PCR PreMix (Bioneer)을 이용하였다. 주형 DNA 50 ng, 각 primer 20 pmol 넣고 증류수로 최종농도가 20 μ l가 되도록 첨가하여 혼합물을 만들었다. 중합효소연쇄반응은 first denaturation 95°C 5분, 30 cycle로, denaturation 96°C 10초, annealing 61°C 30초, extension 72°C 30초, final extension 72°C 10분으로 수행하였다.

2.2.3. 자동 염기 서열 분석

중합효소연쇄반응 산물을 1% agarose gel에 전기 영동한 후, QIAEX Gel Extraction Kit (QIAGEN, Hilden, Germany)로 추출한 후에 염기서열 분석을 하였다. 자동 염기 서열 분석은 주형 DNA 60 ng, primer 3.2 pmol를, 증류수에 혼합하여 12 μ l을 만든 후에 BigDye Terminator RR mix (Perkin-Elmer Applied Biosystems, Part No. 4303153) 8 μ l을 섞어 PCR을 수행하였다. PCR 반응은 Perkin Elmer Cetus 9600을 사용하여, 25 cycle로 95°C 15초, 50°C 10초, 60°C 4분으로 수행하였다. 반응이 끝난 sample은 QIAEX spin column을 이용하여 정제한 후에 5.0% acrylamide gel에서 1X TBE buffer에서 40watt로 12시간 전기 영동하였다.

2.2.4. 미토콘드리아 염기 서열의 multialignment과 계통도의 구성

216개의 한국인 유전자 염기서열 중 130개, 일본인 100명 중 98개 그리고 그외 외국인 155명의 유전자 염기서열 중 primer를 제외한 16016에서 16401 그리고 048에서 388부위의 염기서열을 바탕으로 계통도를 구성하였다. 유전자 염기서열을 DNASTAR program 중의 Meg-Align package의 multialignment algorithm으로 정렬하였다.

3. 유전학적 실험을 통한 한국인/한국어의 기원연구

모계 유전을 통해 전달된 mitochondrial DNA의 분석이 인종의 계통과 그 결과에 의해 언어 계통의 가설을 테스트하는 데에 쓰일 수 있음이 밝혀졌다. 현대 한국인으로부터 DNA를 분리하고 cytochrome B를 코드화하는 유전자와 같은 몇 가지 선택된 marker 유전자를 polymerase 연쇄 반응에 의해 분리하게 된다. 염기 서열은 계통발생론적(phylogenetic) 분석에 따라 결정될 것이다. 만일 현대 한국인으로부터 추출한 유전자 서열을 mitochondrial DNA에 기초한 현재의 가계도(genealogical tree)에 배치할 수 있다면 언어학적 유전학에 기반한 현존의 가설들을 재평가할 수 있을 것이다.

단백질이나 RFLP와 같은 classic marker 분석에 의해 볼 때, 일반적으로 한국인 집단은 몽골인 집단과 같은 동북 아시아인 집단에 가까운 것으로 분석된 바 있다[Goedde et al., 1987; Hong et al., 1993; Nei & Roychoudhury, 1993]. 그러나 Y-염색체의 marker인 YAP/DXYS5Y/SRY465에 관한 연구에 의하면 한국인 집단에서 haplotype HII-1과 같은 동북 아시아 기원으로 추정되는 gene pool외에 haplotypes HI-2a 및 HI-2b의 gene pool이 상대적으로 높게 분포하고 있다. 현재 그 기원에 관해서는 분명치 않으나 아마도 중국 중남부 지역의 벵농사를 하던 민족의 이동과 관련이 높은 것으로 추정된다, [Melton et al., 1998]. 따라서 한국인 집단의 구조는 동북 아시아인의 단일 기원이 아니라 적어도 두 계통 이상의 dual pattern을 보인다.

이에 대하여 북한 사회과학출판사에서 1989년에 장우진 집필로 출판된 조선사람의 기원을 비교하여 보는 일은 흥미롭다. 치아 측정 징표와 혈청 및 지문의 불연속징표들을 통하여 조선 사람의 기원상 특성을 찾으려 했다. 이때 혈청, 지문을 조사하여 살아있는 유기체의 유전신호를 파악하는 데 유전자형의 불연속적 분자구조와 관련시키려는 것이다. 혈청학적 유형으로 보아서 (p. 289) 조선사람은 중국의 동북부 [길림, 심양] 지역 주민들과 만주족, 솔론 그리고 연해주 일대의 니브히와 같은 유형의 특징을 가지고 [만주-통구스족과의 유관성을 암시: 필자] 하북성 일대의 중국사람이나 간토히방의 일

본사람, 몽골이나 부랴뜨, 에웬키는 좌표상에서 유전학적 거리가 상당히 멀다.

향후, 보다 다양한 유전적 marker와 광범위한 관련 집단을 대상으로 분석이 이루어진다면, 위와 같은 방법도 한국인 집단의 기원을 좀더 분명하게 이해할 수 있는 보조수단으로 기대된다.

덧붙여 언급해 둘 것은 Cavalli-Sforza와 Anthony Edwards가 1962년에 5 혈액군의 기본 체계 위에 3 대륙 15 인종의 진화 수형도를 그려 본 결과 한국어가 그 15 종 중에서는 New Guineans과 가장 가까웠고 그 다음에는 Australian 원주민과 가깝다는 보고가 Cavalli-Sforza (1995) 부자의 책 속에 보이거나 언어학적으로는 거의 무의미한 것이다.

3.1. 미토콘드리아 DNA를 이용한 인종간의 계통분석 (Phylogeny between human species by mitochondrial DNA) 연구 수행 내용

한국, 일본, 터키, 몽골, 통구스 인종간의 인종간의 친연관계를 파악하기 위해 각 인종별로 (평균치) 90명 가량의 사람으로부터 mitochondrial DNA를 추출하여 이들을 sequencing하여 Africa-American을 outgroup, 즉 (자기그룹ingroup이 아닌) 외(外)집단으로 하여 phylogeny 계통 발생론(系統發生論) tree를 구성하였다.

3.1.1. 한국인, 일본인, 몽골인, 터키인, 통구스인의 시료 채취

아래 (6), (7)에서 볼 수 있듯이, 한국인 130명, 일본인 98명 외에, 몽골인 88명, 터키인 67명, 통구스인 51명, 총 434명의 Mitochondrial DNA를 채취하였다. 터키인과 몽골인의 경우 일부는 채취한 지역에 따른 분류를 하였다. 지역이 불분명한 시료를 제외하고, 나머지 시료의 경우 지역 이름, 또는 지방을 다음과 같이 터키중부 = S000.M, 터키동부 = S000.E 등으로 시료의 번호 뒤에 영문으로 나타내었다. 통구스인의 경우 부모 모두가 에벤키(통구스)인인 경우(HC)와 야쿠트(터키계)와 에벤키의 혼혈인 경우(HB)를 고려하여 분류하였다. 통구스계의 에벤키인들과 섞여서 터키계의 야쿠트인들이 같이 살아왔기 때문에 혼혈이 많기 때문이다.

(6)

1	토르구트	45		96	터키중부
2	다리강가	46		97	터키중부
3	칼카	47		98	터키중부
4	칼카	48		99	터키중부
5	칼카	49		100	터키중부
6	칼카	50		101	불명확
7	칼카	51		102	터키북부
8	볼치온	52		103	불명확
9	터키서부	53		104	터키북서쪽
10-1	터키중부	54		105	터키북서쪽
10-2	터키중부	55		106	터키서쪽
10-3	터키중부	56	칼카-이하 동	107	불명확
11-1	터키동부	57		108	터키중부
11-2	터키동부	58		109	터키중부
11-3	터키동부	59		110	터키중부
11-4	터키동부	60		111	터키동북
11-5	터키동부	61		112	터키남쪽
11-6	터키동부	62	이하 없음	113	터키동북
12	칼카-이하 동	63		114	터키중부
13		64		115	터키중부
14		65		116	터키중부
15		66		117	터키중부
16		67		118	터키중부
17		68		119	터키중부
18		69		120	터키중부
19		70		121	터키중부
20		71		122	터키북부
21		72		123	터키북부
22		73		124	터키북부
23		74		125	터키북부
24		75		126	터키북부
25		76		127	터키북부
26		77		128	터키북부
27		78		129	터키북부
28		79		130	터키북부
29		80		131	터키북부
30		81		132	터키북부
31		82		133	터키북부
32		83		134	터키중부
33		84		135	터키북부
34		85		136	터키북부
35		86		137	터키북부
36		87		138	터키북부
37		88		139	터키북부
38		89		140	터키북부
39		90		141	터키북부
40		91		142	터키북부
41		92	터키중부	143	터키북부
42		93	터키중부	144	터키북부
43		94	터키중부	145	터키북부
44		95	터키중부	146	터키중부

(7)

Sample No.	생년	성별	
HA1	1937	F	부모?에벤키
HA2	1940	F	부모?에벤키
HA3	1961	M	모-에벤키
HA4	1979	M	부모?에벤키
HA5	1978	M	부모?에벤키
HA6	1978	M	모-에벤키
HB1	1984		부-야쿠트 모-에벤키
HB2	1969		부-야쿠트 모-에벤키
HB3	1975		부-야쿠트 모-에벤키
HB4	1985		부-에벤키 모-야쿠트
HB5	1975		부-에벤키 모-야쿠트
HB6	1983		부-에벤키 모-야쿠트
HB7	1969		부-야쿠트 모-에벤키
HB8	1982		부-러시아 모-에벤키
HB9	1971		부-구르지아 모-에벤키
HB10	1980		부-에벤키 모-야쿠트
HB11	1977		부-에벤키 모-야쿠트
HB12	1976		부-추바시 모-에벤키
HB13	1967		부-러시아 모-에벤키
HB14	1985		부-에벤키 모-야쿠트
HC1	1962		부-에벤키 모-에벤키
HC2	1946		부-에벤키 모-에벤키
HC3	1954		부-에벤키 모-에벤키
HC4	1941		부-에벤키 모-에벤키
HC5	1939		부-에벤키 모-에벤키
HC6	1939		부-에벤키 모-에벤키
HC7	1962		부-에벤키 모-에벤키
HC8	1962		부-에벤키 모-에벤키
HC9	1991		부-에벤키 모-에벤키
HC10	1944		부-에벤키 모-에벤키
HC11	1966		부-에벤키 모-에벤키
HC12	1938		부-에벤키 모-에벤키
HC13	1978		부-에벤키 모-에벤키
HC14	1958		부-에벤키 모-에벤키
HC15	1976		부-에벤키 모-에벤키
HC16	1962		부-에벤키 모-에벤키
HC17	1954		부-에벤키 모-에벤키
HC18	1963		부-에벤키 모-에벤키
HC19	1998		부-에벤키 모-에벤키
HC20	1924		부-에벤키 모-에벤키
HC21	1953		부-에벤키 모-에벤키
HC22	1964	F	부-에벤키 모-에벤키
HC23	1968		부-에벤키 모-에벤키
HC24	1950		부-에벤키 모-에벤키