

Sample No.	생년	성별	
HD1	1980	M	
HD2	1966	F	
HD3	1991	M	
HD4	1993	F	
HD5	1995	F	
HD6		M	
HD7	1972	F	
HD8	1958	F	
HD9	1988	F	
HD10	1999	F	
HD11	1957	F	
HD12	1980	M	
HD13	1994	M	
HD14	1989	F	
HD15	1968	F	
HD16	1990	F	
HD17	1969	F	
HD18	1973	F	
HD19			
HD20	1959	M	
HD21	1932	F	
HD22	1996	M	
HD23	1989	M	
HD24	1990	M	
HE1	통구스족 중 나나이족 SAMPLE은 수집해 온 뒤 분석과정에서 확실한 자료적 특징이 추출 되지 않아 이번 결과에 포함시키지 않음.		
HE2			
HE3			
HE4			
HE5			
HE6			
HE7			
HE8			

3.1.2. mitochondrial DNA sequencing

아래에 별도로 제시.

3.1.3. 동일 sequence 제거 및 불분명한 sequence 제거

Phydit을 통해 sequence의 similarity를 조사하여 100% 동일한 sequence는 계통분석 자료에서 제외 시켰다. (8)에 있는 한국인의 같은 sequence 19개, 일본인의 같은 sequence 16개, 몽골인의 같은 sequence 7개, 터키인의 같은 sequence 4개를 sequence alignment에서 제외시켰다. 또한 sequencing 오차 등으

로 인하여 발생한 오류 sequence를 제외시켰다.

3.1.4. Africa-American sequence와 나머지 sequence alignment

동일, 오류 sequence를 제외하고 한국 111개, 일본 82개, 몽골 81개, 터키 63개, 통구스 51개, 총 388개의 sequence와 NCBI의 Africa-American(M76334) sequence를 이용하여 sequence alignment를 하였다. ClustalW - multiple sequence alignment(DNA)를 이용하여 align하였으며 (10)에 전체 sequence의 align 결과를 나타내었다. Multiple alignment란 두개 이상의 sequence의 similarity를 한꺼번에 비교할 수 있는 방법을 말한다. 원래 Chimpanzee(*Pan troglodytes*)를 outgroup으로 해 보기도 했으나 워낙 거리가 멀어 비교할 만한 결과가 나오지 않으므로 Africa-American으로 바꾸었다.

3.1.5. 계통 분석

Multiple alignment된 결과를 이용하여 PAUP4.0으로 neighbor-joining tree를 그렸다. Phylit을 통해 nex file을 만들고, 이를 이용하여 PAUP으로 neighbor joining tree를 구성하였다. 분자생물학적 데이터를 토대로 phylogeny tree를 만들어내는 방법으로는 일반적으로 maximum likelihood (ML), maximum parsimony (MP), 그리고 neighbor-joining (NJ)을 많이 사용한다.

Maximum likelihood phylogeny tree가 얻어지는 원리는, 주어진 sequence evolution 모델 하에서 sequence들을 가장 잘 진화적으로 설명할 수 있는 생물계통체제를 통계적으로 찾아내는 것이다. 이 maximum likelihood 접근방식은 Markov chain Monte Carlo나 Bayesian inference 등 현대적인 통계적 방법들을 포함한 여러 가지 강력한 통계적 가설 및 검정이 가능하다는 점에서 매우 강력한 방법이지만 컴퓨터 능력의 한계로 계산에 오랜 시간이 걸린다는 단점이 있다.

Neighbor-joining 등으로 대표되는 distance method는 진화적인 거리를 측정함에 있어서 각 두 개의 염기서열을 비교하고 이 두 서열간의 거리를 측정함으로써 계산한다는 점에서 maximum likelihood 방법과 같은 진화적 모델을 사용한다. 실제로 염기 서열간 거리는 maximum likelihood 방법을 사용하기도 한

다. 그러나 이렇게 계산한 거리들을 모아 phylogeny tree를 바로 만들려고 하면 항상 이들 염기서열 상호간의 거리들을 모두 만족시키는 phylogeny tree는 실제로 존재할 수 없다. 따라서 이들 값을 가장 많이 만족시키는 best-fitting phylogeny를 neighbor-joining 등의 방법을 통해 얻게 된다. 이러한 방법들은 몇 가지 단점이 있는데, 대표적인 것으로, 우선 염기 서열간 차이를 거리라고 하는 숫자로 변환시킴으로써 염기서열 그 자체에 대한 정보가 사라진다는 점을 들 수 있다. 하지만 속도가 빠르고, 정확성도 비교적 높기 때문에 가장 많이 이용되고 있다.

Maximum parsimony 방법은 가장 적은 진화적 변화를 갖는 계통수를 찾는 것으로, sequence position에 따른 변이의 정도가 적을 경우 maximum parsimony tree는 maximum likelihood tree와 매우 유사한 결과를 가져 온다. Neighbor-joining 방법과 달리 가능성이 있는 여러 가지 tree를 만들기 때문에 여러 결과를 이용하여 다양한 분석이 가능하다. 그러나 여러 가지 homoplasy가 심할 경우 가장 좋은 tree를 결정하기 어려우며 Felsenstein zone에서 발생하는 inconsistency문제가 있다. Neighbor-joining 방법을 이용한 인종간 tree 결과는 저 밑에 제시할 그림 5와 같다.

그러면 여기에서부터는 각국어별 시료의 분류 내용과 그 가운데 동일 sequence를 우선 2개의 표로 보이겠다.

(8) 시 료

한국인 (130명)	P001, P002, P003, P004, P005, P006, P007, P008, P009, P010, P011, P012, P013, P014, P015, P016, P017, P018, P019, P020, P021, P022, P023, P024, P025, P026, P027, P028, P029, P030, P031, P032, P033, P034, P035, P036, P037, P038, P039, P040, P041, P042, P043, P044, P045, P046, P047, P048, P049, P050, P051, P052, P053, P054, P055, P056, P057, P058, P059, P060, P061, P062, P063, P064, P065, P066, P067, P068, P069, P070, P071, P072, P073, P074, P075, P076, P077, P078, P079, P080, P081, P082, P083, P084, P085, P086, P087, P088, P089, P090, P091, P092, P093, P094, P095, P096, P097, P098, P099, P100, P101, P102, P103, P104, P105, P106, P107, P108, P109, P110, P111, P112, P113, P114, P115, P116, P117, P118, P119, P120, P121, P122, P123, P124, P125, P126, P127, P128, P129, P130
---------------	--

일본인 (98명)	JP001, JP002, JP003, JP004, JP005, JP006, JP007, JP008, JP009, JP010, JP011, JP013, JP014, JP015, JP016, JP017, JP018, JP019, JP020, JP021, JP022, JP023, JP024, JP025, JP026, JP027, JP028, JP029, JP031, JP032, JP033, JP034, JP035, JP036, JP037, JP038, JP039, JP040, JP041, JP042, JP043, JP044, JP045, JP046, JP047, JP048, JP050, JP049, JP052, JP051, JP053, JP054, JP055, JP056, JP057, JP059, JP060, JP061, JP062, JP063, JP064, JP065, JP066, JP067, JP069, JP071, JP072, JP073, JP074, JP076, JP077, JP078, JP079, JP080, JP081, JP082, JP083, JP084, JP085, JP087, JP088, JP089, JP090, JP091, JP092, JP096, JP097, JP098, JP099, JP100
터키인 (67명)	S009, H10.1, H10.2, H10.3, S11.1, S11.2, S11.3, S11.4, S11.5, S11.6, S092, S093, S094, S095, S096, S097, S098, S099, S100, S101, S102, S103, S104, S105, S106, S107, S108, S109, S110, S111, S112, S113, S114, S115, S116, S117, S118, S119, S120, S121, S122, S123, S124, S125, S126, S127, S128, S129, S130, S131, S132, S133, S134, S135, S136, S137, S138, S139, S140, S141, S142, S143, S144, S145, S146, S147, S148
몽골인 (88명)	S001, S002, S003, S004, H.5, S006, H.7, S008, S012, S013, S014, S015, S016, S017, S018, S019, S020, S021, S022, S023, S024, S025, S026, S027, S028, S029, S030, S031, S032, S033, S034, S035, S036, S037, S038, S039, S040, S041, S042, S043, S044, S045, S046, S047, S048, S049, S050, S051, S052, S053, S054, S055, S056, S057, S058, S059, S060, S061, S062, S063, S064, S065, S066, S067, S068, S069, S070, S071, S072, S073, S074, S075, S076, S077, S078, S079, S080, S081, S082, S083, S084, S085, S086, S087, S088, S089, S090, S091
통구스인 (51명)	HA01, HA02, HA03, HA04, HA05, HA06, HB01, HB02, HB03, HB04, HB05, HB06, HB08, HB011, HC01, HC02, HC03, HC04, HC05, HC06, HC07, HC08, HC09, HC10, HC11, HC12, HC13, HC14, HC15, HC16, HC17, HC18, HC19, HC20, HC21, HC22, HC23, HC24, HD01, HD02, HD03, HD04, HD05, HD06, HD08, HD09, HD11, HD12, HD13, HD14, HD15
Outgroup	Africa-American(AfriameriI)

(9) 동일 sequence

원 sequence	동일한 sequence
P009	P062
P010	P011
P018	P069, P114
P030	P047, P106
P048	P127
P050	P086
P068	P078
P090	P098
P123	P126
JP005	JP100, JP093, JP068, JP058
JP015	JP030
JP010	JP070, JP012, JP099
JP032	JP086
JP042	JP075
H-10-1	H-5
S022	S034
S029	S036
S054	S055
S060	S075
S073	S074, S080
S076	S077
S104	S105
S113	S119, S120
S114	S116
S125	S128, S140, S127, S139, S137, S138
S126	S131, S130, S132

그림 3. chromas program mitochondrial DNA 염기서열 확인 과정

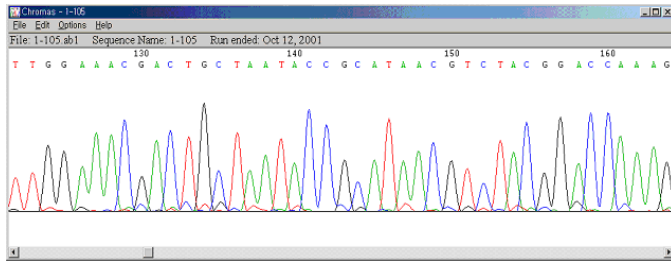


그림 4. Phytid을 이용한 mitochondrial DNA 염기서열 분석과정

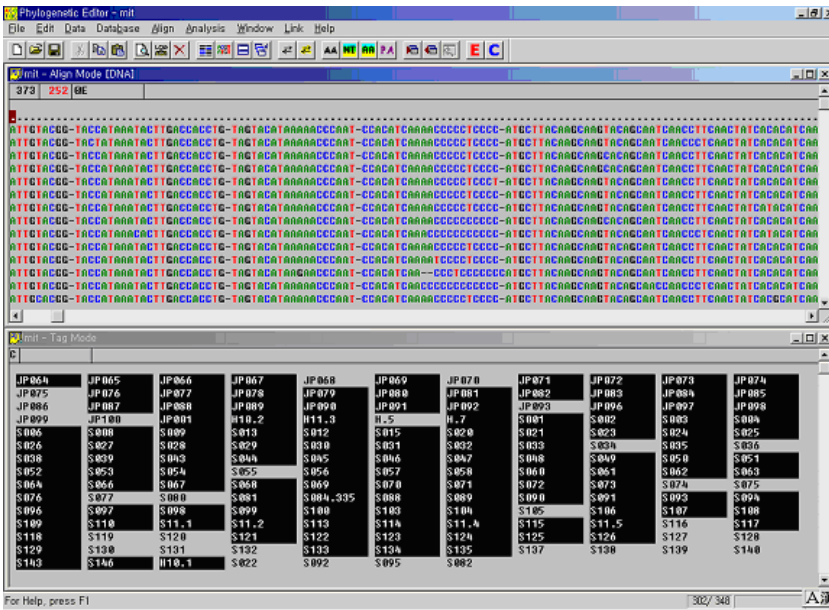


그림 5. mitochondrial DNA 염기서열에 기반한 계통수

4. 연구 결과

실험의 결과 그림 5의 계통발생학(phylogeny) tree에서 볼 수 있듯이 한국과 일본 인종사이는 굉장히 가까운 거리임을 알 수 있다.

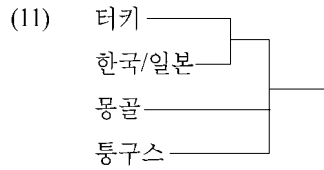
또 통구스인 HD08, HD05, HD01, HD06, HA03, HA01, HA06 및 몽골인 S064은 outgroup인 Afriameril과 비슷한 sequence를 가졌음을 알 수 있다. 그리고 가장 먼저 분화된 것으로 나타난 S061.ca는 칼카지역의 사람으로 몽골인이다. 그 후 S030.ca(몽골)가 분화되고, 계속해서 S054.ca(몽골)가 분화된다. 이는 하나의 알타이어조상에서 몽골인이 가장 먼저 분화되어 나왔음을 보여준다.

몽골인이 분화된 이후, 터키인인 S110.M, S106.W, S100M, S108M 등이 분화되고 한국인인 P048, P128, P088, P111이 분화된다. 이는 한국인과 터키인이 하나의 조상에서 분화되었음을 보여준다. (보통 알타이 어학에서 추정하는 바로는 한국인과 통구스인이 더 가까웠었다고 한다. 이 가설과 위의 결과가 반대되어 흥미롭다.)

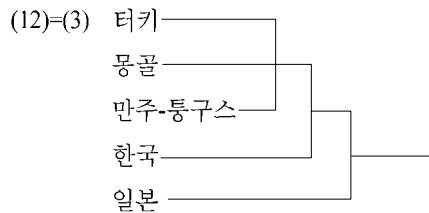
또한 한국인의 경우 다음 clade, 즉 분기군(分岐群, 계통 분기: 공통의 조상으로부터 진화된 생물 분류군)에서 일본인과 교대로 분화됨을 볼 수 있다. 즉, 이것은 한국인과 일본인의 강한 친연관계가 있음을 보여주며, 일정한 패턴이 없이 교대로 분화되는 것으로 보아 어느쪽이 먼저 분화되었다고 보기는 힘들다. 이 결과는 현대 일본인은 소수민족인 아이누와 류큐족을 제외하고는 2500년 전에 한국에서 일본으로 건너간 집단의 후예라는 Sykes 교수 (2001)의 주장과 흡사한 것이다.

또한 그 이후의 clade에서 몽골인인 S029.ca, S033.ca와 함께 통구스인인 HC07, HA04, HC24가 함께 분화되는 것으로 미루어 몽골인과 통구스인은 하나의 조상에서 분화되었음을 짐작할 수 있다. 그리고 몽골인과 통구스인이 계속해서 교대로 나타나는 것으로 보아 몽골인과 통구스인 사이에 지속적인 민족적인 유대관계가 있었음을 알 수 있다.

즉, 하나의 조상에서 몽골인, 통구스인, 그리고 제삼의 조상으로 분화되고, 제삼의 조상에서 터키인과 한국/일본인이 분화되었다고 볼 수 있다.



그러나 비교언어학적으로는 일반적으로 추정해 왔던 아래 표와 같은 가설이 유력하다. 이처럼 위로부터 터키어, 몽골어, 만주-통구스어, 한국어, 일본어 순으로 된 것이 위의 연구결과와는 다른 점이 있어 흥미롭다. 사실 서로 너무 달라 이 차이에 대해 어떤 이유를 붙여야 할지 고민거리다. 아마도 외집단(outgroup)의 설정이 비교 기준에 지대한 영향을 미치므로 이것의 선택에서도 신중을 기할 필요가 있겠다. 그러나 외집단을 남방의 인도네시아나 미국령 사모아인으로 바꾸어 봐도 결과는 대동소이하였다.



참고: 외집단(outgroup)이 chimpanzee였던 경우의 결과

일차적으로 실험의 outgroup을 chimpanzee로 설정해 보았었는데 그 경우의 결과는 phylogeny tree에서 역시 한국과 일본 인종사이는 굉장히 가까운 거리임을 알 수 있었다. 그 경우 가장 먼저 분화된 것으로 나타난 S015.ca는 칼카 지역의 사람으로 몽골인이었다. 이 sequence가 outgroup인 chimpanzee와 거리가 가까운 것으로 보아 sequencing error일 것으로 추정된다. 그 후 S103이 분화되는데, 이는 터키인으로 터키인의 줄기에서 일본인과 한국인이 먼저 분화되었다고 볼 수 있었고, 그 clade에서 분화된 JP002, JP007보다 P006, P094이 약간 먼저 분화된 것을 보아 한국인이 일본인보다 먼저 분화됐을 것으로 추

측했었다. (이 결과는 보통 알타이 어학에서 추정하는 바와 반대 되어 흥미롭다.) 하지만 clade가 하나인 것으로 보아 많은 수가 이동하지 않았다는 것을 추측할 수 있다. 또한 그 이후의 clade에서 몽골인인 S091에서 터키인인 H10.2, H10.1.M이 분화되는 한편, 터키인인 S098에서 몽골인인 S002.da가 분화되는 등 계속적으로 조상과 후손의 역할이 교대로 변화하는 것으로 보아 터키인과 몽골인은 민족적인 유대관계가 있었음을 추측할 수 있다. 그리고 P072가 포함된 clade에서 통구스인인 HB03, HB05, HB08, HC05, HC09 등 많은 통구스인이 분화되는 것으로 보아 통구스인은 한국인에서 파생되었을 것이라고 생각되었다. (이 또한 알타이 어학에서는 통구스인이 한국인과 가장 가까운 친연 관계에 있다고 보지만, 그 아랫 갈래로 보지는 않고 오히려 아주 옛날에 서로 같은 조상에서 갈라졌을 것으로 본다.)

즉, 몽골인과 터키인 사이에는 밀접한 민족적 유대관계가 존재하였고, 이후 자연환경 또는 전쟁 등의 영향으로 공통 조상에서 터키와 몽골이 순서대로 먼저 갈라져 나가고 한반도로 온 한국인이 후에 일본으로도 이동하였다고 볼 수 있었다. 또한 통구스인은 일본인이 분화된 clade보다 한참 후에 형성된 clade에서 분화되는 것으로 보아 일본인이 분화된 이후 한참 후에 한국인이 분화될 때 동시에 한 갈래로 분화되었을 것이라고 생각되었다.

그러나 outgroup을 chimpanzee로 설정해 본 이 결과는, 나중에 Africa-American으로 outgroup을 바꿔 본 것 보다, 비교언어학적 가설에 더 불합치한 결과를 보여 포기하였다.

5. 향후 연구의 가능성

위에 계획되었던 유전학적 연구결과가 다 나오게 되었지만, 이를 근거로 비교언어학 및 계통론에서 얻은 과거의 가설들을 재검토하여, 보다 확고한 과학적인 결론을 도출할 수 있을지는 미지수다. 성급한 결론은 금물이겠지만, 어느 쪽이 더 실증적 자연과학이나 하는 기준으로 본다면 유전학적 연구결과가 더 신빙성이 있는 쪽이라 할 수 있다. 비교언어학 및 계통론에서 얻은 과

거의 가설들은 다분히 심증적 요소도 포함된 fiction일 수도 있기 때문에 좀더 검증이 필요하다. 결국 더 합리적인 판단에 이르도록 노력해야 할 뿐이다.

물론 Azerbaijan과 Armenian의 경우처럼 계통적 친연성이 전혀 없는 두 언어가 지리적으로 인접해 뒤섞여 살다 보니, 유전학적 조사 결과로도 유사성이 많아 판단을 흐리게 하는 경우도 있다. 그러나 우리의 역사적 상황은 어느 인접 언어와도 그 정도로 뒤섞인 것은 아님이 분명하다.

또한 미토콘드리아의 유전적 계통선을 이용한 모계의 조상찾기가 유일한 방법이 아니고, Y 염색체의 유전적 계통선을 밝힐 수 있다면 ‘아담의 아들’도 추적할 수 있을 것이다. 그러나 세포핵 속의 수많은 유전자가 역시 유사한 메시지를 후손에게 전할 수 있기 때문에, 앞서 간 수천 세대의 수없이 많은 사람의 유전자가 우리들에게 심어져 있어서, 친연성의 정도를 분간해 낸다는 일이 그리 용이한 것만은 아닐 것이다.

끝으로 outgroup선정에 따라 결과가 썩 달라지는 것은 꼭 흥미있는데, 기왕이면 알타이족(특히 몽골족)과 관련이 있을 듯하다는 American Indian (Amerindian)으로 외집단을 잡아 보았더라면 하는 아쉬움이 있다. 현재는 Amerindian의 공개된 정보를 얻을 수 없어 이번에는 수행하지 못했으나, 향후 이를 구해 유전학적 방법의 결과가 달리 나온다면, 이에 의해 비교언어학적 탐색을 심화시키도록 할 것이다.

사용 S/W

1. PAUP 4.0 beta version
2. Phytit 3.2

참 고 문 헌

- 국어연구회(1990), *국어 연구 어디까지 왔나*, 서울: 동아출판사.
 김남길(1987), “Korean,” B. Comrie ed., *The World's Major Languages*, Oxford: Oxford University Press.

- 손종영(1973), "A Study of Grammatical Case of Korean, Japanese, and Other Major Altaic Languages," Ph.D. dissertation, Indiana University.
- 손호민(1994), *Korean*, London: Routledge.
- 이기문(1963a), "A Genetic View of Japanese," *朝鮮學報* 27. 94-105/136-147.
- _____(1967), "한국어 형성사," 한국문화사 대계(大系) V, 서울: 고려대 민족문화연구소.
- _____(1972), *국어사개설*, 개정판, 서울: 민중서관(1978, 탑출판사).
- 장우진(1989), *조선사람의 기원*, 서울:사회과학출판사.
- Bodmer, W. F. & L. L. Cavalli-Sforza, 1976. *Genetics, Evolution, and Man*. Freeman and Co., San Francisco.
- Cann, R. L. M., M. Stoneking & A. C. Wilson, 1987. Mitochondrial DNA and human evolution. *Nature* 325, 31-6.
- Casanova, M., P. Leroy, C. Boucekine, J. Weissenbach, C. Bishop, M. Fellous, M. Purrello, G. Fiori, and M. Siniscalco, 1985. A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* 230: 1403-1406.
- Cavalli-Sforza, Luigi. Luca & W. F. Bodmer, 1971. *The Genetics of Human Populations*. Freeman and Co., San Francisco.
- Cavalli-Sforza, L. L. & Mark Seielstad, 2001. *Genes, Peoples, and Languages*, University of California Press, Berkeley.
- Cavalli-Sforza, Luigi. Luca, and Francesco Cavalli-Sforza, 1995. *The Great Human Diasporas: The History of Diversity and Evolution*, Addison-Wesley Publishing Co., Seoul.
- Cavalli-Sforza, L. Luca, Paolo Menozzi, and Alberto Piazza, 1994. *The History and Geography of Human Genes*. Princeton University Press, Princeton, New Jersey.
- Edwards, A., H. A. Hammond, L. Jin, C. T. Caskey, and R. Chakraborty, 1992. Genetic variation at five trimeric and tetrameric tandem loci in four human population groups. *Genomics* 12: 241-253.

- Goedde, H. W., Y. K. Paik, C. C. -Lee, H. G. Benkmann, L. Kriese, P. Bogdanski, and M. Winkler, 1987. Red cell and serum protein polymorphisms in three population groups of south Korea. *Gene Geogr.* 1: 177-188.
- Hong S. S, J. A. Suh, J. J. Chae, S. H. Goh, Y. S. Kim, U. K. Kim, Y. Namkoong, and C. C Lee, 1993. Frequency distribution of alleles at D1S80 and apo-B 3' VNTR loci in Korean population. *Mol. Cells* 3: 457-453.
- Jin Feng, Saito Naruya, and Omoto Keiichi, 1996. *Population Genetic Studies on the Chinese*. International Research Center for Japanese Studies, Kyoto.
- Jobling, M. A. and C. Tyler-Smith, 1995. Fathers and sons: the Y chromosome and human evolution. *Trends Genet.* 11: 449-456.
- Martin, Samuel, 1991. On dating changes in the phonetic rules of Korean. *Festschrift Fr Bruno Lewin*, vol.3. Bochum.
- Melton, T., S. Clifford, J. Martinson, M. Batzer, and M. Stoneking, 1998. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *Am. J. Hum. Genet.* 63: 1807-1823.
- Nei, M. and A. K. Roychoudhury, 1993. Evolutionary relationships of human populations on a global scale. *Mol. Biol. Evol.* 10: 927-943.
- Ono, Susumu, 1970. *The Origin of Japanese*, Tokyo.
- Poppe, Nicholas, 1964, *Introduction to Altaic Linguistics*, Otto Harrassowitz, Wiesbaden.
- Roewer, L., M. Kayser, P. Dieltjes, M. Nagy, E. Bakker, M. Krawczak, and P. de Knijff, 1996. Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Hum. Mol. Genet.* 5: 1029-1033.
- Ruhlen, Merritt, 1996. *The Origin of Language: Tracing the Evolution of the Mother Tongue*, John Wiley & Sons.
- Starostin, Sergej Anatolyevicz, 1991. *Altaic Problems and the Origin of the*

- Japanese Language. Moscow. [김영일 역, 1996. 알타이어 비교연구, 대일 출판.]
- Swofford, David L. 1988. Phylogenetic Analysis Using Parsimony.
- Sykes, Bryan, 2001. The Seven Daughters of Eve: The Science That Reveals Our Genetic Ancestry. W.W. Norton & Co.
- Tautz, D., 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 17: 643-647

탈고 후기

Cann, R. L. M., M. Stoneking & A. C. Wilson(1987)과 Cavalli-Sforza, L. Luca, Paolo Menozzi and Alberto Piazza(1994)는 이 연구의 촉발제가 되었다. 특히 후자 책에 내려진 결론이 비교언어학의 추정과 꽤 달랐기 때문에 재조사해 보려는 탐구심을 촉발한 것이다.

본 연구 초기에는 현재 충북대가 보유하고 있는 3만년전 홍수굴 아이 유골도 조사하려 했었다. 처음 전해 듣기로는 미이라라고 해서 상당히 기대를 하였으나 직접 확인한 결과 3만년된 유골의 상태가 그리 좋아 보이지 않았다. 뼈가 온전히 보존되어 있지 않았고 상당히 조각 조각 부서진 상태이며 또한 보존을 위해서 고정액으로 처리한 상태여서 채료를 할 수 없는 상황이었다.

충북대에는 다른 유골들도 있어서 조선시대 것이 약 30-50구, 고려시대의 것이 5구 정도가 있다. 3만년된 유골에서의 DNA추출이 불가능하더라도 고려와 조선시대의 유골에서는 가능하리라고 생각되었다. 이 시대의 DNA와 현대인의 DNA 분석 비교를 앞으로 시도해 보면 좋을 것이다. 그런데 Wilson & Cann의 논문에서도 현대 여성을 대상으로 했으니, 우리도 최소한 현대한국인을 우선 조사할 수 있고, 그런 실험 결과를 대비시켜 볼 수 있을 인접 지역(시베리아, 중국의 만주, 일본) 자료를 더 확보 연구해서, African Origin 도표 중에서 한국인이 어디에 정확히 위치하는지를 밝히려 하면 될 것이다. 본 연구는 그런 태도로 이루어진 것이다.

《ABSTRACT》

A Search for Korean Origins through Genetic and
Comparative Linguistic Methods

Sang-Oak Lee, Jung-Bin Lee, Sunyoung Kim, Jongsik Chun

In exploring the origins of the Korean language, there are three windows through which we may penetrate the mysteries of this difficult question: archaeological, genetic/ genomic, and linguistic. In this paper I will try to answer this question by looking mainly through the latter two windows. As many linguists have attempted to place the Korean language into the proper language family according to linguistic affinity, I also want to classify the Korean language with linguistically related languages, mainly the Altaic languages.

The so-called Altaic linguistics is very unstable compared with the more rigorously documented Indo-European linguistics. The lack of sufficient evidence, such as a common indigenous lexicon, has made it difficult to ascertain the genetic relationships and origins of languages such as Korean, Japanese, Manchu-Tungusic (Man, Gold(i), Oroqen, Ewenke, Lamut, Nanay, etc.), Mongolian (Khalkha, Chakhar, Urat, Khorchin, Ordos, Buriat, Oirat, Kalmyk, Da(g)ur, Monguor, Yellow Uighur. etc.) and Turkish (Turkish, Turkmen(ian), Azerbaidjani, Uighur, Uzbek, Kumyk, Tatar, Kazakh, Kirg(h)iz, Yakut, Altai, and Chuvash, etc.), which are typically classified as Altaic.

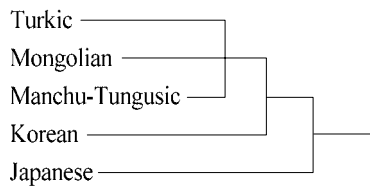
Recently Starostin (1991) claimed that Proto-Altaic had disappeared around the sixth century B.C. Other measures are needed to evaluate this hypothesis, though, since the intra-linguistic debate has not provided any clear evidence or breakthroughs. That is why we turn to genetics as an approach and incorporate the results to shed new light on the origins of the Korean language. It has been

shown that analysis of mitochondrial DNA, transmitted through maternal lineage, can be used to test this kind of hypothesis.

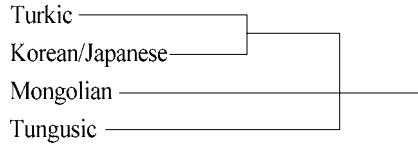
After collecting data from such ethnic groups as Korean, Japanese, Ewenke, Nanay, Khalkha, Buriat, and Turkish, we have investigated information on mtDNA (mitochondrial DNA). While we do not use the popular STR markers on Y chromosomes (namely, DYS19, DYS389I, etc.) in this study, we will type a number of Korean, Tungusic and Mongolian populations on Y chromosomes, and other nuclear markers such as the gene encoding cytochrome B will be isolated by the polymerase chain reaction, as in the next study,.

In reality, it was rather difficult to extract genes from hair-roots and saliva, especially when there were not many samples available for serious study. Yet, we could fill the void for Koreans in the map that shows the mitochondrial DNA types in Africans, Australian Aborigines, Caucasians, East Asians, Native Americans, and New Guineans. The nucleotide sequences were determined through phylogenetic analysis. If we can position the sequences from Koreans in the currently available genealogical tree based on mitochondrial DNA, we may be able to reassess the existing hypotheses on linguistic genealogy.

However, because primordial remains or ancient fragments of linguistic evidence are not readily available in Korea due to special geo-political situations, the position of the Korean language has been left unattested in the genealogical tree based on languages. Nonetheless, there are some examples similar to the Altaic languages that are estimated to have been used in Korea, and using these we may draw the linguistic tree shown below.



On the other hand, our investigation through which maternal lineage alone has well established a biological lineage can be reported as follows.



Korean is not isolated from Japanese, and the sisterhood with Turkic is also strange because the similarity between Korean and Manchu-Tungusic is more convincing according to linguistic evidences. It looks to be difficult to conclude anything at this stage of research. However, we know the genetic methods are quite scientific, while the comparative linguistic study often relies on abstract reconstruction.