

DNA 해프로트리와 유전적거리에 의한 가계족보의 계층화

류광렬*

Genealogical Stratification by Genetic Distance and DNA Haplotrees

Kwang Ryol Ryu *

*Professor, Department of Electronic Engineering, Mokwon University, Daejeon 35349, Korea

요 약

본 논문은 유전자가계족보 작성을 위해 인체세포의 Y염색체와 X염색체의 DNA에서 해프로코드로부터 해프로 그룹의 해프로트리, 유전적거리 및 전통가계족보를 결합하고 계층구조화한다. 적용지역은 충청지방으로 부계의 Y-DNA는 O그룹에서 O3a* 및 O2b* 그룹의 빈도가 높고, 모계의 mtDNA는 L3그룹의 D*와 M* 그룹의 빈도가 비교적 높다. 유전자 거리 산정은 니의 표준유전적거리법을 적용한다. 실험결과 0.1는 바로 직계가족, 0.1에서 0.8은 근친이며, 1.0 이상은 친족으로 추정하기 어렵다. DNA의 해프로그룹과 해프로트리에 의해 STR은 친족 확인에 적합하고 SNP는 개인유전적식별이 정확하므로 한국전통 족보는 3가지 인수를 추가하므로 더 과학적인 계층화가 구현된다.

ABSTRACT

This paper describes hierarchically stratifying and analyzing haplotrees of haplogroups from haplotypes on the Y and X chromosomes of human cells for genetic and Korean traditional and genealogical trees. The specific region is Chungcheong province, and the Y-DNA of the paternal lines has high frequency of O3a* and O2b* in the O group, and the mtDNA of the maternal line has a relatively high frequency of D* and M* in the L3 group. Each combination of these constructs the family tree of the father lines and mother lines. Genetic distances using Nei's standard genetic distance, are very close relatives of less than 0.1 and close relatives of 0.1 to 0.8. Provided, the distance is more than 1.0, it is difficult to estimate relatives. STR has the identified kinship, and SNP has the personal genetic identification. A scientific stratification of the Korean genealogical tree is created by the three factors.

키워드 : 족보계층화, 유전적거리, 해프로트리, 해프로그룹, DNA

Key word : Genealogical stratification, Genetic distance, Haplotype, Haplogroup, DNA

Received 30 September 2019, Revised 2 October 2019, Accepted 16 October 2019

* Corresponding Author Kwang Ryol Ryu (E-mail: conan@mwu.ac.kr, Tel: +82-42-829-7651)
Professor, Department of Electronic Engineering, Mokwon University, Daejeon 35349, Korea

Open Access <http://doi.org/10.6109/jkiice.2020.24.1.65>

print ISSN: 2234-4772 online ISSN: 2288-4165

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

유전자(gene)는 DNA(Deoxyribo Nucleic Acid)에 저장된 유전정보이다. DNA는 핵산(nucleic acid)의 일종이며, 지놈(genome) 또는 유전염색체는 DNA의 염기서열(base sequence) 전체이다. 유전자의 유전정보는 개인의 생화학적 고유특성을 나타내기 때문에 법의학, 친자확인, 유전질환, 개인고유식별코드 및 생체인식(biometrics) 등에 활용된다.[1-3]

유전정보는 1961년 니런버그(Nirenberg)와 코라나(Khorana)의 염기서열 해독에 의해 유전부호가 해독되었으며, 2003년 지놈 프로젝트에 의해 인간의 유전자정보체계가 99% 완성되었다.[4-5] 따라서 바이오 컴퓨팅, 생명공학, 유전통계학 분야에서 연구 및 응용되고[6-7], 공공기관, 기업, 연구소, 대학 등에서 DNA 연구 및 표본 검사 확대와 데이터베이스도 구축되고 있다.[8-9]

인체의 유전부호화는 세포분열시 세포핵 내에 염색체의 유전자좌(locus)의 분석과 DNA 복제(genetic copying)시 오류로 발생하는 돌연변이 (mutation)의 염기서열(base sequence)을 기반으로 이루어진다. 돌연변이는 주로 단연쇄반복유전자 STR(Short Tandem Repeat)과 단일염기다형성 SNP(Single Nucleotide Polymorphism) 분석에 의해 부호화되어 해프로타입(haplotype)과 해프로그룹(haplogroup)이 작성된다. STR은 친자확인, SNP는 개인식별 및 연관성 연구 등에 활용된다. 이 결과는 성염색체로부터 부계(parental line)와 미토콘드리아에서 모계(maternal line) mtDNA 또한 상염색체에서 atDNA (autosome) 등이 구축된다.[10-12]

유전적거리(genetic distance)는 성씨 또는 집단간 유사성과 이질성의 평가를 위해 DNA 연구가 기반이 된다. 이것은 성씨 집단의 변천역사에 대한 과학적 평가이며, 유사성의 차이를 정량화하여 개인 또는 그룹간 거리 정도를 판정한다. 또한, 광의로 인구 분기 이동이 일어날 때 그 시대의 이주루트를 추정할 수 있다. 유전자거리 계산은 개인간 거리에 유클리드(Euclidean) 거리기법, 집단간 거리에 나이(Nei)의 유전적표준거리법 등이 있다.[13-14]

한국의 가계족보체계는 약 천 년 이전부터 부계중심으로 작성되어왔으며, 현재 문화류씨 세보가 최초의 서적이다. 이 족보는 고려시대의 시조로부터 세대별로 부모영역, 자매영역, 자녀영역을 중심으로 계층화하였다.

[15-17] 주로 부계 Y-DNA의 계층 구조이다.

본 연구는 유전자정보코드로 부계와 모계의 DNA의 염기서열의 해프로타입과 해프로그룹에 의한 해프로트리 구조와 STR, SNP 및 유전적거리 등 3 요소를 적용하여 한국의 전통족보체계와 결합하여 고유의 과학적 유전자가계족보의 계층화를 구현한다.

II. DNA 해프로트리

2.1. DNA의 STR과 SNP

DNA는 염기단위인 뉴클레오티드(nucleotide)가 결합되어 이루어져 있고, 인체의 경우 총 30억의 염기쌍(base pair)이 있다. 염기쌍 사이의 거리가 0.34nm이므로 총 길이는 약 1m 이다. 염색체가 23쌍이므로 DNA도 23개이다. 그러므로 DNA의 평균 길이는 4.4cm 이지만 긴 것은 12 cm에 달한다. B-DNA 직경은 2 nm, 1회전 당 길이는 3.4nm이다. 이중나선(double helix)의 회전방향은 반시계방향이고 나선의 1회전 당 염기쌍은 10.5개이다.

DNA는 세포분열시 자신과 동일한 DNA를 복제하여 유전자 정보를 보존하는 기능을 하고 동시에 유전자 정보를 발현시킬 수 있는 단백질 합성을 지시하고 조절한다. 단백질을 구성하는 기본물질은 20 종류의 아미노산이다. DNA는 유전정보만을 저장하고, RNA는 유전정보를 전달하고 단백질을 합성한다. DNA 염기서열이 전사(transcription)되어 단일 나선의 mRNA를 만들어 단백질 조립 장소인 리보솜(ribosome)으로 운반한다. mRNA는 그림 1과 같이 tRNA로 전달되고 번역(translation) 된다. 세쌍(triplet) 뉴클레오타이드 코드인 코돈(codon: triplet code) 단위로 아미노산 사슬이 결합되어 새로운 단백질이 복제된다. 여기서 DNA 염기쌍 bp(base pair)는 A-T (T-A).G-C(C-G)이고, RNA 염기쌍은 T가 U로 대체된다.

DNA의 유전적 염기서열은 99.9%가 동일하고 0.1%가 상이하다. 상이한 특성을 분석하면 개인의 유전적 차이가 판별된다. 유전적 변이(variation)는 복제과정의 돌연변이에 기인한다. 유전자 좌의 마커(marker)는 염색체 내에서 염색체 위치 또는 DNA 염기서열로, 한 사람의 DNA를 다른 사람과 비교하는 데 유용하다.

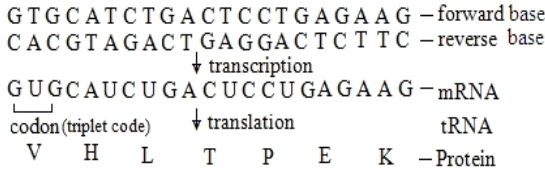


Fig. 1 Genetic copying

주로 STR과 SNP분석 기법이 사용된다. STR은 2~6bp의 짧은 염기서열이 반복적으로 나타난다. STR-DYS439의 경우 DNA서열반복코드는 AGAT, 대립유전자 반복수 12, 돌연변이비 0.00477이다. 검사 결과는 해프로타입으로 표시된다. STR 마커는 DkSnnn, DkSnnn=m 등으로 표기한다. 여기서 D:DNA, k:염색체번호 또는 Y; X 성염색체명, S:세그먼트(segment), n:발견된 순서의 숫자, m=변이 세그먼트의 반복인수이다. 예로 DYS393=12로 표기된다. 마커는 염색좌와 연관된다. 반복요소, 대립유전자(allele), 돌연변이율 등이 포함된다. 마커 레벨은 1-111이며 12, 25, 37, 67, 111 등의 5 그룹으로 구분하며 레벨이 클수록 정확성이 기대된다.

SNP는 유전자의 양은 같지만 염기서열 중 특정 위치에서 하나의 DNA 서열의 변형을 유발하여 개인차이가 발생된다, 예로, CGACGA와 CGATGA에서 T하나의 염기가 상이한 SNP이다. 단일염기가 어떤 패턴인가에 따라 개인별 특징이 결정되고, 검사 결과의 그룹이 해프로그룹이다.

SNP 마커는 Mnnn, SRYnnn (예, M175, SRY465) 등으로 표시한다. 본 논문에서는 한국인에 많은 O의 경우 O-M175(O-SRY465) 또는 D-M125 등으로 표기 한다. 여기서 O와M은 해프로그룹이고, 마커의 M 문자는 연구소명이나 연구팀명, 숫자는 발견된 순위이다.

2.2. GD 유전적 거리

유전적 거리는 DNA 결과를 개인 간 또는 집단 간 거리차로 일치 여부를 판정한다. 대상간의 거리차는 니(Nei) 유전적표준거리를 사용한다. 대립유전자 빈도수 [10] X_k 와 비교할 빈도수 Y_k 의 차이에 의해서 계산된다. 실험은 조부(GF), 부(SA)와 부의 형제 형제(SB), 자녀(GSA, GSB, GSC)와 별개인(RYU)을 대상으로 표 1의 해프로타입과 대립유전자 빈도수를 적용한다. GF와 RYU의 예는 그 결과 식 1의 분모가 0.007858, 분자 0.12669이며 니의 유전적표준거리는 1.612이다.

$$D_N = -\ln \frac{\sum_{k=1}^n (X_k Y_k)}{[\sum_{k=1}^n X_k^2 \sum_{k=1}^n Y_k^2]^{1/2}} \tag{1}$$

Table. 1 Y-STR haplotypes

name	393	390	19	391	385a	385b	426	388	439	389a	392
GF	12	23	14	10	15	17	11	17	12	13	11
SA	12	23	14	10	15	17	11	17	12	13	11
SB	12	23	14	10	15	17	11	17	12	13	11
GSA	12	23	14	10	14	17	11	17	12	13	11
GSB/GSC	12	23	14	10	14	17	11	17	12	13	11
RYU	13	23	17	10	9	18	11	12	12	14	13

389b	458	459a	459b	455	454	447	437	448	449	464a	460	YGH
29	17	8	8	11	11	25	14	20	25	14	0	0
29	17	8	8	11	11	25	14	20	25	14	0	0
29	17	8	8	11	11	25	14	20	25	14	0	0
29	17	8	9	11	11	25	14	20	25	14	0	0
29	17	8	8	11	11	25	14	20	25	14	0	0
30	17	9	10	11	11	24	14	17	29	13	11	11

GF allele frequency

0.01464 0.03877 0.08047 0.08463 0.00293 0.00193
 0.06859 0.00084 0.06761 0.12068 0.06332 0.09133
 0.04878 0.00442 0.00442 0.13169 0.00004 0.06543
 0.06173 0.0459 0.00129 0.00056

RYU allele frequency

0.11031 0.03963 0.00057 0.08665 0.00012 0.00101
 0.07023 0.00084 0.06923 0.022 0.06316 0.03807 0.0499
 0.03632 0.00344 0.13484 0.00004 0.02463 0.06321
 0.00092 0.04121 0.0005 0.08407 0.0591

2.3. 해프로그룹과 해프로타입

머리카락, 타액, 혈흔, 손톱, 체액 등에서 DNA를 추출하고 PCR(polymerase chain reaction)에 의해 DNA가 증폭되어 분석 판독한다. Y-DNA의 Y-STR 마커는 친족관계에 있는 남성간의 혈연관계를 확인하거나 부계 혈통의 검사 및 추적하는 방법으로 이용된다. 부에서 자로 98% 유전된다. Y-DNA는 부계 Y염색체를 검사하는 표준 방식이며, 이 염기서열을 검사함으로써 고조부-증

조부-조부-부-자로 전달되는 부계중심의 혈연관계가 검사된다. 충청도 인구의 표본조사[18]에서 해프로그룹의 마커는 O3a* (O3a1c, O3a2c1a, O3a2c1), O2b* (O2b1b, O2b1a, O2b1c), C2 순으로 빈도가 높고, 기타 낮은 빈도 D N J Q 등이다. 전국에 문화류 남의 인구는 2015년 통계 기준 164천여 명이며 충청도지방에 약 10%가 거주한다. 특정한 류씨 성의 검사는 O3a3b*이다. 그래서 다수의 유전자도 O3*와 O2*에 속한다고 본다.

mtDNA는 모계혈통을 기반으로 물려받는다. 고조모-증조모-조모-모-녀의 계층구조이다. 충청지방의 모계는 주로 D*(D4, D5), M*(M8), B*(B4), 기타 등의 빈도로 분포한다. 이 해프로타입 트리가 모여 모계유전자가 계보가 작성된다.

2.4. 해프로트리

해프로그룹은 동일한 해프로타입의 집단이며 주로 SNP 돌연변이의 그룹이므로 어떤 변이를 기준했는가에 따라 다양한 그룹이 형성된다. 염색체상의 패턴은 해프로타입 블록과 태그(tag) SNP를 찾아 작성된다. SNP는 DNA 염기서열(A,T,G,C)에서 하나의 염기만 다른 유전적 변이이다. 이것은 한 블록 안에는 많은 수의 SNP가 존재할 수 있지만 대표할 수 있는 몇 개의 특정 태그(tag) SNP만 사용된다. 이것은 유사한 해프로타입의 집단으로 동일한 염색체내에서 유전되는 경향이 있는 인접한 SNP집합이다. 예로, 염기서열 ACAGTCACGTCGTCACG와 ACCGTCACGAACGGACG를 각각 비교하면 SNP는 3, 10, 14번째 위치에서 이상이 보인다. 이 그룹이 유전되는 특성이 있다면 해프로타입은 A T T와 염기쌍 C A G이다. 따라서 한국인에 많은 Y-SNP 변이 마커 M-176(O2b), O-M175, O-M9는 표2와 같다.

Table. 2 Y-SNP Haplogroups

Haplogroup	SNP
O-M176	TACAGCCATGCACAGAGAG
O-M175	GATTTAAACTCTCTGAATCAGGCACAT
O-M9	CAGGACCCTGAAATACAGAACT

STR은 반복단위의 크기가 2-5개의 염기쌍이 반복되는 변이이다. 예로 STR-DYS449의 반복염기는 TTTC이며 반복수는 27이다. SNP는 STR보다 변이율이 낮지만 부계간에 동일한 유전자형을 갖기 때문에 분별력 문제가 있다.

그래서 SNP는 주로 개인특징식별분석에, STR은 부모형제 친자확인 검사 및 유전자계보 등에 적용된다.

부계 중심의 Y-DNA 해프로그룹 명은 대문자의 그룹명과 하위분기(subclade)로 이루어진다. 하위분기는 숫자와 소문자의 영숫자로 조합된다. 즉, O2b1a2와 같이 문자-숫자의 조합이다. 그룹명의 대문자는 서양 인류의 시조인 아담(Adam) A를 시작으로 A0부터 R까지이며, 인류의 이동루트에 따라 Y-DNA 그룹명이 정해진다.

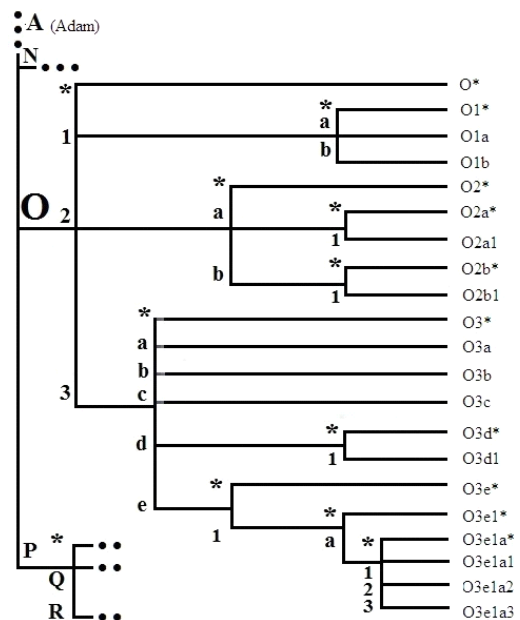


Fig. 2 Y-DNA haplotree

해프로그룹명은 단축형과 일반형(longhand)으로 구분된다. 단축형은 대문자와 -(hyphen) 및 돌연변이 STR 또는 SNP로 작명된다. 예로 O-M175의 경우 영문 대문자 O는 해프로그룹이고 M175는 SNP 마커이다. O-M175*에서 *표시는 파생의 분기(branch)가 많아 a b... 또는 1 2... 등이 추가됨을 의미한다. 이 의미는 어데 류씨 성의 무슨 파와 세손 및 항렬의 구조와 같다. 주로 한국인에게 많은 O 해프로트리(haplotree) 그림은 2와 같이 작도되고, 유전자정보코드와 해프로그룹이 모여 계층화 집단을 이루면 부계유전자계보가 작성된다.

모계 mtDNA는 이브(Eve) L을 시조로 그림 3과 같이 대문자 L0부터 Y까지 표시하여 계통발생분기(phylogenetic

tree)로 작도된다. 트리구조는 그림 2의 Y-DNA와 유사하나 간소화된 해프로트리는 그림 3과 같이 L3에서 M과 N으로 분기하고 L3M. M은 동쪽지역 A B C D F G 등으로 분기한다. 따라서 Y-DNA와 mtDNA 검사의 해프로그룹에 의해 유전자족보, 가계족보도 DNA 작성과 조상의 연관성 및 이주루트가 추정된다.

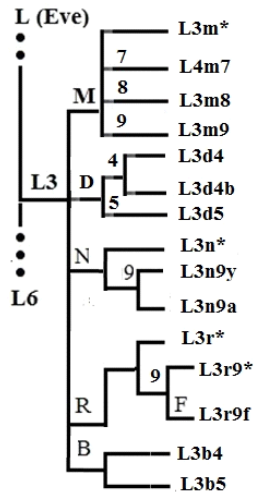


Fig. 3 mtDNA haplotree

2.5. 결과 고찰

유전적거리는 그림4와 같이 조부(GF)와 두 아들(SA, SB)은 0.0, 손자(GSA, GSB, GSC)는 해프로타입 차이로부터 0.0, 0.05, 0.8 또한 류는 1.6이다. 이 결과에서 0.1은 바로 직계가족, 0.1에서 0.8은 근친이며, 1.0 이상은 친족으로 추정하기 어렵다. 따라서 1.0 이상은 혈연관계로 볼 수 없다.

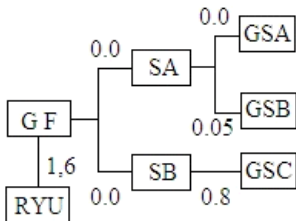


Fig. 4 Results of genetic distances

유전자는 특정한 인체형질을 결정하고 생명유지에 필요한 정보를 저장한다. DNA는 한세대에서 다음 세대로 유전하여 한 가문의 족보 계층을 구성한다. 한국인의

전통 가계족보는 시조를 중심으로 거주지역(본관)과 고 급관료가 되어 일부 분파를 만든다. 세는 조, 부, 기, 자, 손 등을 년대 순차로 계열된다. 대는 출생에서부터 30년을 1대로 시간적 영역이다.

부계중심의 전통족보는 계층구조가 유사하고 부모영역, 형제자매영역, 자녀영역과 하단에 쪽 지정영역의 단으로 분리된다. 우측 끝에는 단 별로 세가 지정된다. 기록되는 정보는 본관, 성명, 출생일, 사망일, 출생지, 성별 등이다. 여기에 XDNA, mtDNA, 유전적거리 등의 유전자정보코드를 그림 5와 같이 포함하면 과학적 유전자가계족보가 계층화된다. 그러나 유전자 검사의 결과가 족보 단 별 영역에서 혈연관계로 추정하기 어려운 경우도 발생하기 때문에 유전자정보 포함 여부는 보안 및 신중할 필요가 있다.

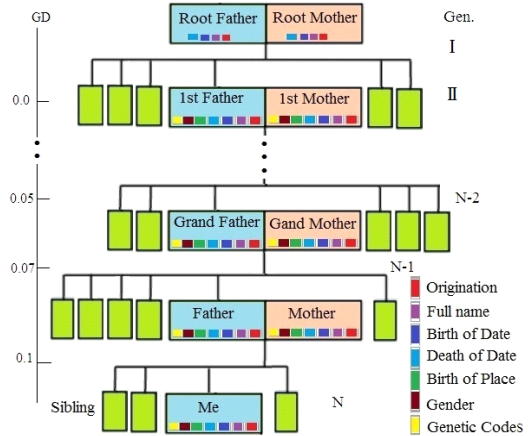


Fig. 5 Korean genealogical family tree

III. 결론

본 연구는 전통가계족보와 결합하기 위해 유전적거리와 Y염색체와 X염색체의 DNA에서 STR과 SNP를 기반으로 해프로타입과 해프로그룹 등의 요인을 계층구조화에 적용하였다. 유전적거리 산정은 나의 표준유전적거리법을 적용했다. 실험결과 0.1은 바로 직계가족, 0.1에서 0.8은 근친이며, 1.0 이상은 친족으로 추정하기 어렵다. DNA 대상은 충청도 사람이며, 한국인의 경우 부모 중에서 부계의 Y-DNA는 O그룹에서 O3a* 및 O2b*가 많이 나타나고, 특정한 류씨 성의 유전자정보코드는 O3a3b*이지만 대표성이 미흡하기 때문에 문화류

씨들의 DNA검사를 확대하고, 기타 성씨들의 해프로그
룹을 작성 할 필요가 있다. 모계의 mtDNA는 L3그룹으
로 L3D*이 비교적 빈도가 높다. 해프로그룹은 개인 고
유의 생체고유식별코드이므로 코드의 집합이 유전자족
보로 응용 될 수 있고, 3가지 유전적 요소가 전통족보와
결합하여 과학적 DNA코드의 가계족보를 구현한다.

ACKNOWLEDGEMENT

This study was financially supported by the
research year fund of Mokwon University.

References

- [1] H. B. Ma, "Paternity Testing," *Journal of American Science*, vol.2 no.4, 2006.
- [2] Y. Nurhantarl, and H. Suryadi, "Genetic Inconsistency in Paternity Investigation," *conf. of UASC Life Sciences*, vol. 2019. 2016.
- [3] T.A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, and H.M. Kang, "A Global Reference for Human Genetic Variation," *Nature* 526, pp.68-74, 2015.
- [4] H.J. Seo, S.H. Cho, M.Y. Kim, and S.d. Lee, "Forensic DNA Phenotyping: A Review in Korean Perspective," *Korean J. Leg. Med.* vol. 41, pp.23-31, 2017.
- [5] H. Sahakyan, B.H. Kashani, R. Tamang, and A. Kushniarevich, "Origin and Spread of Himan Mitochondrial DNA Haplogroup U7," *Nature Scientific Reports*, 2017.
- [8] A. H. Zhang, H. Y. Lee, S. B. Seo, H. J. Lee, and S. D. Lee, "Y Haplogroup Distribution in Korean and Other population," *Korean J Leg Med* , vol. 34, no. 1, pp.34-44, 2012.
- [9] T. Wei, F Liao, C. Pan, C. Xiao, and D. Huang, "A Novel Multiplex Assay of SNP-STR markers for forensic Purpose," *PLOS ONE*, 2018.
- [10] E. Friedman, "Y-DNA SNPs and Haplogroups," *Family Tree DNA*, 2014.
- [11] Y-Chromosome Marker Detail, [Internet]. Available :<http://www.smgf.org/chromosome/>, 2018.
- [12] R.C. Aiken, K.E. Johnson, and B.J. Voight, "Signal of Variation in Human Mutation nRate at Multiple Levels of Sequence Context," *MBE*, 2019.
- [13] O. Tal "Two Complementary Perspective on Inter-individual Genetic Distance," *Biosystems*, 2012.
- [14] Genetic distance, [Internet]. Available : https://en.wikipedia.org/wiki/genetic_distance.
- [15] C. Ryu, "Stratification of Genetic Genealogy on Specific-Regional People by DNA and Haplogroup," *proc. of KIICE*, vol.23, no.2, pp.565-568, 2019.
- [16] S.E. Wallace, E.G. Gournal, V. Nikoloval, and N.A. Sheehan, "Family Tree and Ancestry Inference," *BMC Medical Ethics*, 2015.
- [17] The Great Assembly of Moonhwa Ryu's Clan, "*The Moonhwa Ryu's Genealogy*," vol. 1, N-Korean, 2008.
- [18] 47zM128, Korean Y-DNA and mtDNA, [Internet]. Available: <https://qudro.blog.me/>.



류광렬(Kwang Ryol Ryu)

목원대학교 전자공학과 교수

※ 관심분야 : BME, DSP, Biometrics, Spiritual Science