ISSN:2790-1661 Volume-8-(2023)

Genomic Data Disclose Potential Information on Evolutionary Interactions among Different Human Populations and Novel Education Technology Development

Wei Xia[1,a], and Zhizhou Zhang[2,b]

[1] School of Language and Literature, Harbin Institute of Technology, Weihai, China.

[2] BIOX Biotechnology Center, Harbin Institute of Technology, Weihai, China;

[a] xiawei2015@hitwh.edu.cn, [b] zhangzzbiox@hitwh.edu.cn

Abstract. Study on language gene polymorphism patterns (LGPP) across different population genomes could provide incentives to develop novel education technology and important information on human evolution. In this study, as a preliminary observation, we adopted 148 single nucleotide polymorphism (SNP) sites from 13 language genes, each with 4-13 SNPs. These SNPs were screened across 112 whole genome sequences (including 59 ancient genomes ranging from 2000 BP to 120000 BP) from five continents (Africa, Asia, Europe, North America, and South America). We found that five distinct LGPPs featured across human evolution history, in which LGPP-1 may be the oldest version shared by animals and primitive hominins, though data also showed that LGPP-1 is still existing in some modern human populations. Asian and African possessed all LGPP types while European seemed lacking in the LGPP-2. Surprisingly, African samples had a relatively larger evolutionary distance from animals than other populations in LGPP1-4, while in LGPP-5 (the modern human type), some African samples had a relatively small evolutionary distance from animals than other human populations. Except for LGPP-2, all other LGPPs contained Asian, African and European, suggesting that there were vigorous interactions among these three continents all the time during human evolution. In this study, ancient American samples were only found in LGPP1-3, suggesting that either mutual migration among different continents happened much earlier than expected, or ancient Americans had little interactions with other populations after migrating into the America land.

Keywords: Language gene; genome; human evolution; polymorphism pattern.

1. Introduction

The analysis and research based on the large-scale data of human genome are crucial in many fields, including the new techniques of individualized education and teaching based on personal genome information, the extensive interaction between different populations in the history of human evolution, and the evolutionary process of language and cognitive abilities. As language ability involves a group of language genes, and each language gene has polymorphisms among different populations, multiple genes form a polymorphism pattern. Therefore, it is worth exploring how this language gene polymorphism pattern evolves itself during the human evolution process.

In this study, as a preliminary observation, we adopted 148 single nucleic polymorphism (SNP) sites from 13 language genes (Table 1, Table 2), each with 4-13 SNPs. These SNPs were screened across 112 whole genome sequences (Table 3, including 59 ancient genomes ranging from 2000 BP to 120000 BP) from five continents (Africa, Asia, Europe, North America, and South America).

2. Methods

2.1 Language genes and their SNPs

Language is an emergent phenotype of human being, though many other animals also have their own types of ‘languages’. If a gene mutation is statistically or experimentally associated with functional loss of a certain language, it would be called language gene. Language gene SNP data were selected, though randomly, still relatively evenly distributed along the genomic sequence of

each gene, in the dbSNP database: https://www.ncbi.nlm.nih.gov/snp/; Table 1 listed thirteen language genes (as a preliminary observation, only thirteen language genes were employed at the time the manuscript was written), and a total one hundred and forty-eight SNPs from these thirteen genes were selected in this study (Table 2).

Table 1 Language genes employed in this study

	Name	Compromised ability when mutated(example)	References
1	FOXP1	Expressive language	1
2	FOXP2	Speech	2
3	CNTNAP2	Early language development	3-5
4	TPK1	Syntactic and lexical ability	6-7
5	DCDC2	Reading, dyslexia	8-10
6	KIAA0319	Reading, dyslexia	3,8,11-12
7	TM4SF20	Language delay; communication disorder	13
8	FLNC	Reading, language	14
9	ATP2C2	Memory	15
10	ROBO1	Phonological buffer	16-17
11	ROBO2	Expressive vocabulary	18
12	CMIP	Reading, memory	3,8,15
13	NFXL1	Speech	19

Table 2 Tested 148 SNPs of thirteen language genes

Abbreviation	Language gene/SNP	Abbreviation	Language gene/SNP	Abbreviation	Language gene/SNP
ROBO-10	ROBO1 rs34841026	FXP1	FOXP1 rs7638391	CMI-1	CMIP rs201316817
ROBO-1	ROBO2 rs11127602	FOXP1-1	FOXP1 rs76145927	CMI-2	CMIP rs183876152
ROBO-2	ROBO2 rs10865561	FOXP1-2	FOXP1 rs75214049	CMI-3	CMIP rs183075361
ROBO-3	ROBO2 rs5788280	FOXP1-3	FOXP1 rs17008544	CMI-4	CMIP rs114894868
ROBO-4	ROBO2 rs3923745	FOXP1-4	FOXP1 rs17008063	CMI-5	CMIP rs79979027
ROBO-5	ROBO2 rs3923744	FOXP1-5	FOXP1 rs11914627	CMI-6	CMIP rs74031247
ROBO-6	ROBO2 rs1163750	FOXP1-6	FOXP1 rs7639736	CMI-7	CMIP rs60152409
ROBO-7	ROBO2 rs1163749	FOXP1-7	FOXP1 rs1499893	CMI-8	CMIP rs57603843
ROBO-8	ROBO2 rs1163748	FOXP1-8	FOXP1 rs1053797	CMI-9	CMIP rs35429777
ROBO-9	ROBO2 rs1031377	FOXP1-9	FOXP1 rs144080925	CMI-10	CMIP rs34119643
ROBO-11	ROBO2 rs78817248	FOXP1-10	FOXP1 rs17008224	CMI-11	CMIP rs16955675
ROBO-12	ROBO2 rs144468527	FOXP1-11	FOXP1 rs147756430	CMI-12	CMIP rs2288011
ROBO-13	ROBO2 rs17525412	FLN-1	FLNC rs2291569	CMI-13	CMIP rs1187121850
ROBO-14	ROBO1 rs77350918	FLN-2	FLNC rs2291568	ATP-1	ATP2C2 rs78371901
ROBO-15	ROBO1 rs6795556	FLN-3	FLNC rs2291566	ATP-2	ATP2C2 rs74038217
ROBO-16	ROBO1 rs35456279	FLN-4	FLNC rs2291565	ATP-3	ATP2C2 rs62640935
TM1	TM4SF20 rs6724955	FLN-5	FLNC rs2291563	ATP-4	ATP2C2

					rs62640932
TM2	TM4SF20 rs44675173	FLN-6	FLNC rs2291562	ATP-5	ATP2C2 rs62640931
TM3	TM4SF20 rs4675172	FLN-7	FLNC rs2291561	ATP-6	ATP2C2 rs62050917
TM4	TM4SF20 rs4673192	FLN-8	FLNC rs2291560	ATP-7	ATP2C2 rs16973859
TM5	TM4SF20 rs4438464	FLN-9	FLNC rs2291558	ATP-8	ATP2C2 rs13334642
TM6	TM4SF20 rs4428010	FLN-10	FLNC rs2249128	ATP-9	ATP2C2 rs4782970
TM7	TM4SF20 rs4408717	FLN-11	FLNC rs117864464	ATP-10	ATP2C2 rs4782948
TM8	TM4SF20 rs13415654	FLN-12	FLNC rs35281128	ATP-11	ATP2C2 rs2435172
TM9	TM4SF20 rs80305648	FLN-13	FLNC rs371111092	ATP-12	ATP2C2 rs247885
TM10	TM4SF20 rs137891000	DCD-1	DCDC2 rs35029429	ATP-13	ATP2C2 rs247818
TPK-1	TPK1 rs113536847	DCD-2	DCDC2 rs2274305	KIA-1	KIAA0319 rs138160539
TPK-2	TPK1 rs79464600	DCD-3	DCDC2 rs34584835	KIA-2	KIAA0319 rs117692893
TPK-3	TPK1 rs77358162	DCD-4	DCDC2 rs33943110	KIA-3	KIAA0319 rs114195393
TPK-4	TPK1 rs28380423	DCD-5	DCDC2 rs33914824	KIA-4	KIAA0319 rs699461
TPK-5	TPK1rs17170295	DCD-6	DCDC2 rs9467075	KIA-5	KIAA0319 rs699462
TPK-6	TPK1 rs12333969	DCD-7	DCDC2 rs9460973	KIA-6	KIAA0319 rs699463
TPK-7	TPK1 rs6953807	DCD-8	DCDC2 rs3846827	KIA-7	KIAA0319 rs730860
TPK-8	TPK1rs17170295	DCD-9	DCDC2 rs3789219	KIA-8	KIAA0319 rs10946705
TPK-9	TPK1 rs67644764	DCD-11	DCDC2 rs33943110	KIA-9	KIAA0319 rs75674723
TPK10	TPK1 rs77358162	DCD-12	DCDC2 rs190254728	KIA-10	KIAA0319 rs75720688
NFX-1	NFXL1 rs1964425	CNTN-1	CNTNAP2 rs1637842	KIA-11	KIAA0319 rs150584710
NFX-2	NFXL1 rs1822030	CNTN-2	CNTNAP2 rs1637841	KIA-12	KIAA0319 rs115399701
NFX-3	NFXL1 rs1822029	CNTN-3	CNTNAP2 rs1479837	KIA-13	KIAA0319 rs7770041
NFX-4	NFXL1 rs1812964	CNTN-4	CNTNAP2 rs1468370	FOXP2-1	FOXP2 rs10227893
NFX-5	NFXL1 rs1545200	CNTN-5	CNTNAP2 rs1062072	FOXP2-2	FOXP2 rs10244649
NFX-6	NFXL1 rs1440228	CNTN-6	CNTNAP2 rs1062071	FOXP2-3	FOXP2 rs12705977
NFX-7	NFXL1 rs1371730	CNTN-7	CNTNAP2 rs987456	FOXP2-4	FOXP2 rs61732741
NFX-8	NFXL1 rs1036681	CNTN-8	CNTNAP2 rs700309	FOXP2-5	FOXP2 rs61758964
NFX-9	NFXL1 rs978094	CNTN-9	CNTNAP2 rs700308	FOXP2-6	FOXP2 rs62640396
NFX-10	NFXL1 rs920462	CNTN-10	CNTNAP2 rs3194	FOXP2-7	FOXP2 rs73210755
NFX-11	NFXL1 rs147017712	CNTN-11	CNTNAP2	FOXP2-8	FOXP2

			rs535454043		rs1058335
NFX-12	NFXL1 rs13152765	CNTN-12	CNTNAP2 rs2373284	FOXP2-9	FOXP2 rs61753357
NFX-13	NFXL1 rs34323060	CNTN-13	CNTNAP2 rs61732853	FOXP2-10	FOXP2 rs144807019
				FOXP2-11	FOXP2 rs182138317

2.2 Sample genome sequences.

All genome sequences (Table 3) were downloaded from ENA database (https://www.ebi.ac.uk/ena/browser/). Total one hundred and twelve whole genomes (including fifty-nine ancient genomes) from five continents (Africa, Asia, Europe, North America, and South America) were collected, among which, there are twenty-seven from EastAsia (China), ten from Nepal, twelve from other SouthAsia countries, twenty from Africa, twenty-eight from Europe, nine from SouthAm, two from NorthAm, two from SEAsia, one from MAsia and one from WAsia.

Table 3 The 112 whole genomes employed in this study

	abbr	Country	Region	Age (BP)	Supplemental info	Genome file size (G)	References
1	c26	China (a)	EastAsia	5500	China PLT-M312	32	PRJEB36297
2	c25	China (a)	EastAsia	5500	China WGM70	51	PRJEB36297
3	c24	China (a)	EastAsia	3700	China (WD-WT5M2)	41	PRJEB36297
4	c23	China (a)	EastAsia	5300	China (BLSM27S)	67	PRJEB36297
5	c22	China (a)	EastAsia	4000	China (SM-SGDLM27)	222	PRJEB36297
6	c21	China (a)	EastAsia	4079-3913	China (LJM3)	140	PRJEB36297
7	c19	China (a)	EastAsia	5304-5056	China (WGM35)	83	PRJEB36297
8	c18	China (a)	EastAsia	4225-3995	China (PLTM310)	73	PRJEB36297
9	c17	China (a)	EastAsia	4143-3985	China (SM-SGDLM6)	81	PRJEB36297
10	c16	China (a)	EastAsia	3800-4000	China (LJM14)	61	PRJEB36297
11	c15	China (a)	EastAsia	3181-3073	China (JXNTM23)	77	PRJEB36297
12	c14	China (a)	EastAsia	2338-2180	China (LGM79)	101	PRJEB36297
13	c13	China (a)	EastAsia	2200-2000	China (LGM41)	82	PRJEB36297
14	c12	China (a)	EastAsia	4151-3974	China (PLTM311)	87	PRJEB36297
15	c11	China (a)	EastAsia	4089-3983	China (WD-WT1H16)	84	PRJEB36297
16	c10	China	EastAsia		90 Han-3	93	PRJEB11005
17	dc2	China	EastAsia		DaiChina-1 HG00766	87	SRX5983023
18	c9	China (a)	EastAsia	6175-5937	XW-M1R18 in ancient China	117	PRJEB36297
19	c8	China (a)	EastAsia	7000	ZLNR-2 in ancient China	17	PRJEB36297
20	c7	China (a)	EastAsia	7000	WQM4 in ancient China	27	PRJEB36297
21	yc6	China	EastAsia		Yi	9	PRJEB36297
22	c5	China	EastAsia		OROQEN	32	PRJEB36297
23	c4	China	EastAsia		Hezhen	81	PRJEB36297
24	dvi	China/Russia (a)	EastAsia	8000	DevilsGate	41	PRJEB14817
25	c3	China	EastAsia		90 Han NA18547-mix1	156	PRJEB11005
26	c2	China	EastAsia		90 Han NA18561	78	PRJEB11005
27	dc1	China	EastAsia		DaiChina-2	42	SRX5983023

28	th1	Thailand	SouthAsia		Thailand Thai	18	PRJEB9586
29	ne1	Nepal	SouthAsia		Nepal Kusunda	30	PRJEB9586
30	ne2y	Nepal (a)	SouthAsia	~2000	Kyang-KS25	34	PRJEB41752
31	ne3y	Nepal (a)	SouthAsia	~2000	Kyang-KS20	31	PRJEB41752
32	ne23y	Nepal (a)	SouthAsia	~2000	Kyang-KS20+ks25	31+34	PRJEB41752
33	ne4m	Nepal (a)	SouthAsia	~2000	Mebrak-mix1	40	PRJEB41752
34	ne5s	Nepal (a)	SouthAsia	~2000	Samdzong S143	45	PRJEB41752
35	ne6s	Nepal (a)	SouthAsia	~2000	Samdzong S183+S13+S173+S153	74	PRJEB41752
36	ne7s	Nepal (a)	SouthAsia	~2000	Samdzong S22+S29+S30+S36+S20+ S16+S163	48	PRJEB41752
37	ne8s	Nepal (a)	SouthAsia	~2000	Samdzong S8+S18+S21	37	PRJEB41752
38	ne9m	Nepal (a)	SouthAsia	~2000	Mebrak-mix2	58	PRJEB41752
39	ne10m	Nepal (a)	SouthAsia	~2000	Mebrak-mix3	54	PRJEB41752
40	ba1	Banglade sh	SouthAsia		Bangladesh Bengali	129	PRJEB9586
41	ca1	Cambodia	SouthAsia		Cambodia Cambodian	80	PRJEB9586
42	pa3	Pakistan	SouthAsia		Pakistan Brahui	128	PRJEB9586
43	pa2	Pakistan	SouthAsia		Pakistan Sindhi	101	PRJEB9586
44	sr2	Sri Lanka	SouthAsia		SriLankan No.2	101	PRJNA552609
45	pa1	Pakistan	SouthAsia		Pakistan Balochi	67	PRJEB9586
46	vn1	Vietnam	SouthAsia		Vietnam-HG2080	123	PRJEB9586
47	sr1	Sri Lanka	SouthAsia		SriLankan-1	102	PRJNA552609
48	in1	India	SouthAsia		60-Indian MIX1	82	PRJEB16019
49	in2	India	SouthAsia		GujaIndian MIX1	109	PRJNA530214
50	ga3	Gambia	Africa		Gambian3	19	PRJEB31736
51	ga2	Gambia	Africa		Gambian2	124	PRJEB31736
52	ga1	Gambia	Africa		Gambian1	125	PRJEB31736
53	en1	Nigeria	Africa		ENigeria-1	120	PRJEB31736
54	en2	Nigeria	Africa		ENigeria-2	101	PRJEB31736
55	en3	Nigeria	Africa		ENigeria-3	122	PRJEB31736
56	ke1	Kenya	Africa		LuhyaKenya-mix1	104	PRJEB31736
57	ke2	Kenya	Africa		Kenya Luhya-2	84	PRJEB9586
58	ke3	Kenya	Africa		Kenya Bantu	21	PRJEB9586
59	sl1	Sierra Leone	Africa		Sierra Leone mix1	160	PRJEB31736
60	sl2	Sierra Leone	Africa		SierraLeone Mende	17	PRJEB9586
61	co1	Congo	Africa		Congo SAMEA3302769/2716/25 69	17	PRJEB9586
62	sa1	Southern Africa	Africa		Southern Africa KB1	26	PRJNA46161
63	sa2	Southern Africa	Africa		Southern Africa NB1	16	PRJNA46161
65	sa3	Southhern Africa	Africa		Southern Africa combined three individuals	10	PRJNA46161
65	et1	Ethiopia(a)	Africa	4500	Ancient Ethiopian 'Mota' genome	31	PRJNA295861
66	ss3	sub-Sahar a (a）	Africa	3160	African foragers I10873_new+I10874_new	25	PRJEB49291
67	ss2	sub-Sahar a (a）	Africa	7900	African foragers I10871_new	42	PRJEB49291
68	ss1	sub-Sahar a (a）	Africa	4500	African foragers I5950_new	51	PRJEB49291
69	mo1	Morocco(a)	Africa	15000	Ancient Morocco genome	18	PRJNA422662
70	bu1	Bulgaria	Europe		Bulgarian	110	PRJEB9586

					SAMEA3302842 SAMEA3302718		
71	sib	Russia (a)	Europe	45000	45000yr Siberia	13	PRJEB6622
72	nd8	Russia (a)	Europe	60000	Neandertal Mezmaiskaya-2	20	PRJEB21881
73	nd7	Germany(a)	Europe	120000	Neandertals Hohlenstein-Stadel Cave	27	PRJEB29475
74	nd6	Belgium(a)	Europe	120000	Neandertals Scladina Cave in Belgium	36	PRJEB29475
75	nd5n	Russia (a)	Europe	60000	Neandertal Mezmaiskaya	5.4	PRJEB1757
76	nd4	Russia (a)	Europe	50300	Neanderthal Altai	158	PRJEB1265
77	nd3	Spain (a)	Europe	60,000-120,000	Neanderthal Devils Tower	6	PRJEB31410
78	nd2	Spain (a)	Europe	60,000-120,000	Neanderthal ForbesQuarry	143	PRJEB31410
79	nd1	Russia (a)	Europe	50000	Neandertal-MIX1	64	PRJEB29475
80	sp1	Spain	Europe		SPAIN1	200	PRJNA42557
81	sp2	Spain	Europe		SPAIN2	32	PRJNA42557
82	sp3	Spain	Europe		Spain3	98	PRJNA42557
83	fr4	France (a)	Europe	4000	France4000	167	PRJEB9586
84	fr5	France (a)	Europe	5000	France5000	122	PRJEB9586
85	fr7	France (a)	Europe	7000	France7000	34	PRJEB9586
86	fi1	Finnish	Europe		Finnish mix-1	119	PRJNA38449
87	fi2	Finnish	Europe		Finnish mix-2	61	PRJNA38449
88	fi3	Finnish	Europe		Finnish mix-3	62	PRJNA38449
89	cz1	Czech (a)	Europe	45000	Czechia ancient	112	PRJEB39040
90	de2	Russia (a)	Europe	100000	Denisova2	109	PRJEB20653
91	dep	Russia (a)	Europe	74000-82000	DenisovaPha	95	PRJEB3092
92	it1	ITALY	Europe		Italian	46	PRJEB9586
93	f1	France	Europe		France	53	PRJEB9586
94	tu1	Turkey	Europe		Turkey Turkish	11	PRJEB9586
95	br1	England	Europe		British in England and Scotland	42	PRJEB31249
96	ge1	Georgia (a)	Europe	9529 – 9895	Georgia kk1	14	PRJNA670050
97	la1	Latvia (a)	Europe	6179-5750	ZVEJ31 Latvia	13	PRJNA670050
98	pe3	Peru	SouthAm		peru--ERR042533_FINAL	28	PRJEB31736
99	pe1	Peru	SouthAm		PERU ERR042535-MIX1	67	PRJEB31736
100	pe2	Peru	SouthAm		PERU-ERR042532_FINAL	27	PRJEB31736
101	me1	Mexica	SouthAm		Mexican LosAngeL-1	103	PRJEB31736
102	ur1	Uruguay(a)	SouthAm	668	Uruguay (CH13)	19	PRJEB48360
103	ur2	Uruguay(a)	SouthAm	1400	Uruguay (CH198)	12	PRJEB48360
104	ur12	Uruguay(a)	SouthAm	~1000	Uruguay (CH13+CH198)	31	PRJEB48360
105	ch1	Chile（a）	SouthAm	4700	Ayayema	35	PRJEB29074
106	bz1	Brazil (a)	SouthAm	8000	Sumidouro Cave, Lagoa Santa Brazil	41	PRJEB29074
107	sc1	US(a)	NorthAm	10000	US Spirit Cave	10	PRJEB29074
108	us2	US(a)	NorthAm	12500	US ancient Anzick	25	PRJEB29074
109	ha1	Hawaiian	SEAsia		USA Hawaiian SAMEA3302908	106	PRJEB9586
110	ng1	PapuaNewGuinea	SEAsia		PapuaNewGuinea SAMEA3302871	77	PRJEB9586

					SAMEA3302650		
111	ta1	Tajikistan	MAsia		Tajikistan Tajik	47	PRJEB9586
112	is1	Israel	WAsia		Israel	53	PRJEB9586

Note: (a): ancient sample; SouthAm: South America; NorthAm: North America; SEAsia: Southeast Asia; MAsia: Middle Asia; WAsia: West Asia. References are ENA (European Nucleotide Archive) accession numbers.

2.3 Sample SNP information abstraction and PCA analysis.

The 010Edit software was employed to extract all one hundred and forty-eight SNP information from each genome (supplementary file 1). In all one hundred and twelve genomes, the sizes mainly range from 10G to 200G. Genomes less than 10G were neglected or only used as a reference. Principal Component Analysis（PCA）was performed using R packages FactoMineR, factoextra, ggrepel and ggplot2.The main R codes can be requested from the corresponding author.

3. Result and Discussion

3.1 LGPP profile for Asian

Asian have all five LGPPs, especially one Chinese sample, c5, is actually a modern OROQEN person. This suggests that LGPP-1, the supposed oldest version of human language gene polymorphism pattern, still exist in some modern human populations. Actually, the authors speculate that all oldest version LGPPs may only exist remote areas in the world, where food resources are enough for small groups of people but human wars almost never happen for them during the extremely long history.

3.2 LGPP profile for European

Europeans have only four LGPPs in this study, and LGPP-2 seems absent for them. LGPP-2 is mainly represented by Chinese and Nepal people, plus some ancient American and African samples. This may suggest that south margin of the Tibetan plateau was not a place for ancient Europeans and Asian to interact and admix. The interaction between ancient Europeans and Asian shall be mainly mediated by the middle Asia, the whole Russia area and the northeast China. Some LGPP-2 possessing ancient human migrated to America and Africa, but seemingly not Europe.

3.3 LGPP profile for African

African people have all five LGPPs. From LGPP-1 to LGPP-4, African samples had a relatively larger evolutionary distance from animal samples than all other human samples; but in LGPP-5, some African sample such as sa2, had a relatively smaller evolutionary distance from animal samples (classical primates) than other modern human samples. The results may suggest that, though Africa may be not the origin place for human being, it was the origin place for most modern human. This is consistent with the conventional out-of-Africa hypothesis.

3.4 LGPP profile for American

American samples have LGPP-1, LGPP-2, LGPP-3 and LGPP-5, but not LGPP-4 in this study. Ancient Americans had LGPP-1 and LGPP-2, indicating interactions of ancient human populations among several main continents happened much earlier than expected, and these interactions continued till for LGPP-5, but LGPP-4 possessing populations may never get chance to migrate to America.

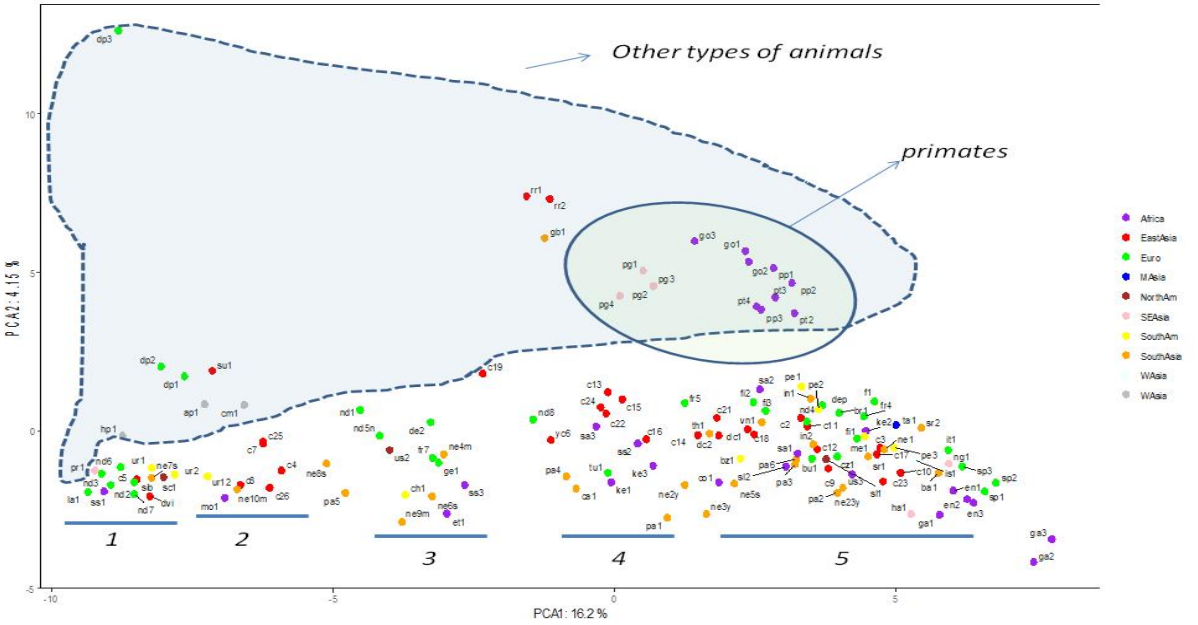

Fig. 1 PCA results suggest that human populations have mainly five distinct LGPPs.

4. Summary

There have been five different types of LGPPs throughout human evolution, with LGPP-1 potentially being the oldest version shared by animals and primates, though data also suggests that LGPP-1 still exists in some modern human populations. Asia and Africa possess all five types of LGPPs, whereas Europe seems to lack LGPP-2, which was only observed in Asia, Africa, and America. Surprisingly, African samples had a significantly larger evolutionary distance from animals than other populations in LGPP1-4, while in LGPP-5 (modern human type), some African samples had a relatively small evolutionary distance from animals than other human populations. All but LGPP-2 contained Asian, African, and European races, suggesting that there has always been vigorous interaction among these three continents throughout human evolution. In this study, ancient American samples were only found in LGPP1-3, suggesting that intercontinental migration may have occurred much earlier than expected or that ancient Americans had little interaction with other populations after migrating to the Americas.

This study disclosed some apparent differences among different populations in language gene polymorphism, suggesting that different populations will demonstrate differential ability for learning a specific language, since language ability itself was first just a muscle movement behavior. Today, personal genome information is penetrating personal medicine and nutrition, and is believed to gradually penetrating personal education. From views of molecular biology and linguistics, most people are common people with various genetic defects, so the learning ability and capacity is different among individuals. Acquiring a higher efficiency of learning by education technology is a normal need for each student, including by taking advantage of personal genome information in the future.

Acknowledgments

This study was supported by State Language Commission Research Grant (YB135-117), Association of Chinese Graduate Education Grant (B-2017Y0505-079) and National Research Center for Foreign Language Education Grant (ZGWYJYJJ10A042).

References

[1] Bacon C & GA Rappold. The distinct and overlapping phenotypic spectra of FOXP1 and FOXP2 in cognitive disorders. Human Genetics,2012,131(11):1687-9168.
[2] Lai CS et al. A fork-head domain gene is mutated in a severe speech and language disorder. Nature, 2001, 413(6855):519-523.
[3] Newbury,D. F.et al.Investigation of dyslexia and SLI risk variants in reading-and language-impaired subjects. Behav. Genet. 41, 90-104; (2011).
[4] Vernes,S. C. et al. A functional genetic link between distinct developmental language disorders.N. Engl. J.Med. 359, 2337-2345;(2008).
[5] Whitehouse,A.J., Bishop, D. V.,Ang, Q. W., Pennell, C. E. & Fisher, S. E. CNTNAP2 variants affect early language development in the general population.Genes Brain Behav. 10, 451-456;(2011).
[6] Villanueva P et al. Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population. European Journal of Human Genetics, 2011, 19(6):687-695.
[7] Fattal I et al.The crucial role of thiamine in the development of syntax and lexical retrieval:a study of infantile thiamine deficiency. Brain, 2011, 134(6):1720-1739.
[8] Scerri,T. S. et al. DCDC2,KIAA0319 and CMIP are associated with reading-related traits. Biol. Psychiatry 70, 237-245;(2011).
[9] Deffenbacher,K.E. et al.Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses. Hum. Genet. 115, 128-138; (2004).
[10] Schumacher,J.et al. Strong genetic evidence of DCDC2 as a susceptibility gene for dyslexia.Am. J. Hum.Genet. 78, 52-62;(2006).
[11] Paracchini,S.et al. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319,a novel gene involved in neuronal migration. Hum.Mol.Genet. 15, 1659-1666;(2006).
[12] Francks,C. et al. A 77-kilo base region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. Am. J. Hum.Genet.75, 1046-1058;(2004).
[13] Wiszniewski W et al. TM4SF20 ancestral deletion and susceptibility to a pediatric disorder of early language delay and cerebral white matter hyperintensities. American Journal of Human Genetics, 2013, 93(2):197-210.
[14] Gialluisi A et al. Genome-wide screening for DNA variants associated with reading and language traits. Genes Brain and Behavior, 2014, 13(7):686-701.
[15] Newbury,D. F.et al. CMIP andATP2C2 modulate phonological short-term memory in language impairment. Am. J.Hum. Genet. 85, 264–272; (2009).
[16] Hannula-Jouppi,K.et al. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. PLoS Genet. 1,e50; (2005).
[17] Bates,T. C. et al. Genetic variance in a component of the language acquisition device: ROBO1 polymorphisms associated with phonological buffer deficits. Behav. Genet.41, 50-57;(2011).
[18] StPourcain, B. et al. Common variation near ROBO2 is associated with expressive vocabulary in infancy. Nat.Commun. 5, 4831; (2014).
[19] Villanueva,P. et al. Exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. PLoS Genet.11,e1004925; (2015).