

# Language Gene Polymorphism Pattern of Pan Paniscus Adds Indirect Evidence that Qinghai-Tibet Plateau Area was Also a Birth Place of Humankind: a Potential Milestone for Both Science and Education

Zhizhou Zhang<sup>1\*</sup> and Wei Xia<sup>2</sup>

<sup>1</sup>BIOX Center for Marine Biotechnology, Harbin Institute of Technology, Weihai, China.

<sup>2</sup>School of Languages and Literature, Harbin Institute of Technology, Weihai, China 264209.

**Corresponding Author:** Zhizhou Zhang, BIOX Center for Marine Biotechnology, Harbin Institute of Technology, Weihai, China.

**Received:** 📅 2025 Aug 11

**Accepted:** 📅 2025 Aug 30

**Published:** 📅 2025 Sep 10

## Abstract

More and more evidences support that anthropoids in Africa originated from Asia, and why Homo genus emerged earlier in Africa than other continents remain enigmatic. One of the anthropoids closest to Homo is Pan paniscus (Bonobo), now inhabiting only south of Congo river. Language ability of Bonobo seems close to that of Homo sapiens, suggesting that it may be evolutionarily only one or several steps away from humankind. In this study, 148 single nucleotide polymorphism (SNP) sites of 13 language genes were employed to roughly search for language gene polymorphism pattern (LGPP) of 113 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). It was found that the LGPP of Bonobo is highly close to several samples: a 7900y old ancient sub-Sahara African forager and three modern human people from Pakistan, Kenya and Congo. This result could be an indirect evidence for the following possibility: uplifting of Qinghai-Tibet plateau induced huge ecological challenge, thus promoting migration of ancient anthropoids around Pakistan area to Africa and evolution into Bonobo and a certain Homo genus that was highly similar to Bonobo. Later on, the Homo genus likely migrated out of Africa and got a chance to reach Pakistan. The preliminary finding in this study though highly speculative may have a potential to be a milestone in human evolution for both scientific research and cultural education.

**Keywords:** Pan Paniscus, Language Gene, Polymorphism Pattern, Human Evolution, Qinghai-Tibet Plateau

## 1. Introduction

### 1.1. Outlines

Simians in different stages of evolution can be mainly seen in Asia and Africa, but there is a report supporting that the origin of simians is more likely in Asia [1]. Actually, more and more evidences have been collected on this issue [2-10]. Especially, Jacques Jaeger et al [8], pointed out that recent discoveries of older and phylogenetically more primitive basal anthropoids in China and Myanmar, the eosimii forms, sustain the hypothesis that Asia was the birth place of anthropoids, rather than Africa, and African colonization of anthropoids was the result of several dispersal events. If anthropoids originated from Asia, they may also have a chance to evolve into Homo sapiens as some reports suggested in China. In 2013 Chinese scientists claimed that they, collaborated with some American researchers, discovered in Yunnan province an ape skull fossil with about 6 Ma, the age of which is similar to the Sahelanthropus tchadensis (a shah Ape at 7 Ma BP) that can walk upright [11]. But this study seems not well

documented in English and it is pretty hard to get its original data including the full-text article [12]. Meanwhile, reported uplift of the Hengduan mountains on the southeastern margin of the Tibetan plateau in the late Miocene and its paleoenvironmental impact on hominoid diversity. They provided magneto strati graph dated the extinction of hominoids in southeastern Asia at 6 Ma and found that the late Miocene extinction of hominoids was coincident with uplift and climate change of southeastern Tibet margin. This study likely provided extremely important information on why and when the hominoid community in Qinghai-Tibet plateau area dispersed into other places in the world. The study indicated that hominoids occurred in Southeast Asia from the latest Middle Miocene to the terminal Miocene (~13-6 Ma). But the Late Miocene uplift (after 6 Ma) of these N-S oriented mountains on the southeastern margin of the Tibetan plateau led to a cooler and drier climate, bringing paleoenvironmental impact on the diversity of the local hominoid community and their potential dispersal.

Primates in the Qinghai-Tibet plateau area were supposed to start the departure for Africa or Southeast Asia between 38 and 29 million years ago [12-14]. And the dispersion continued till 6 million years ago. Around 6 million years ago, the uplifting of Qinghai-Tibet plateau was already too high for almost all apes. Only the Hengduan Mountains on the southeastern margin of the Tibetan plateau still maintains small amounts of several types of primates till now. So, it will be valuable to systematically search for different types of primate fossils dated 6~38 million years ago. One example was *Sivapithecus* fossils (5.3~23.7Ma BP) discovered in Pakistan and India [15]. This ancient ape had apparent characters of both chimpanzee and orangutan, suggesting that *Sivapithecus* moving west to Africa evolved into Chimpanzee, while *Sivapithecus* moving to southeast Asia evolved into orangutan. Those *Sivapithecus* hidden in the Hengduan Mountains on the southeastern margin of the Tibetan plateau may be the ancestor of *Sivapithecus lufengensis* (8 Ma BP) and related *Homo erectus* in China [16].

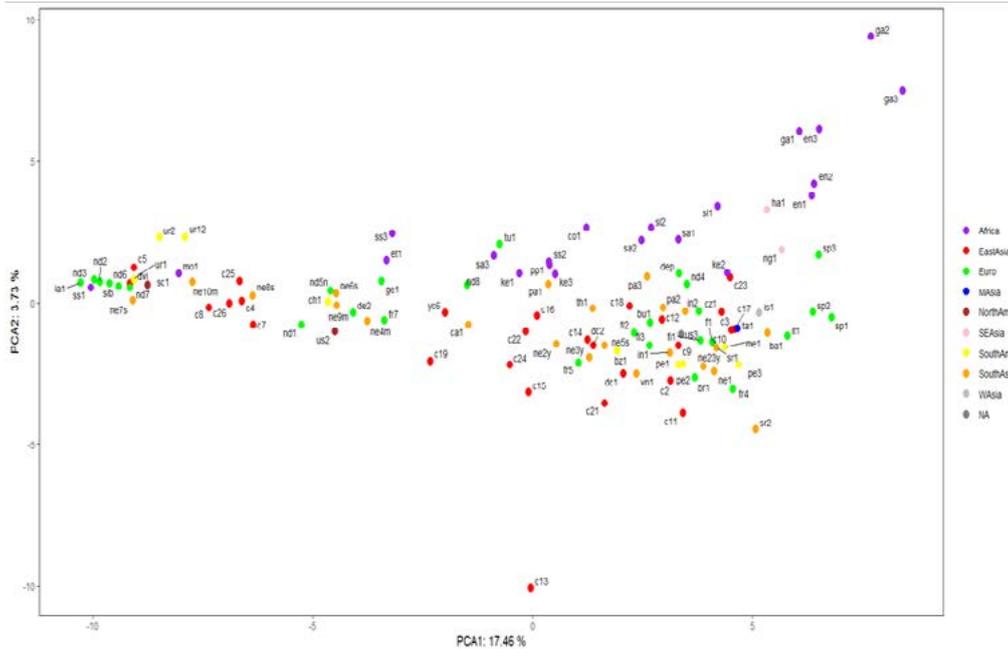
East African Rift Valley is widely regarded as a driving force for human evolution. In fact, the time for the formation of East African Rift Valley was largely overlapped with that for the eminence of Qinghai-Tibet plateau [12,13,17]. It is evidenced that the central Tibetan valley was at a relatively low elevation of 1,700 meters around 40 million years ago, supporting a subtropical "Shangri-La" diverse ecosystem. It rose rapidly at an elevation over 4,000 meters between 38 and 29 million years ago, and this elevation shall be high enough to expulse most animals. Apes in this area shall have experienced the whole uplifting process, especially huge climate and ecological challenges. It is hard to deny that the eminence of Qinghai-Tibet plateau might bring an evolutionary force to apes no less than East African Rift Valley. This postulation has been approved by the fact that fossils of *Homo erectus wushanensis* (2 Ma BP) and many other *Homo erectus* fossils were found Southwest China (east side of Qinghai-Tibet plateau). Of course, there was a big difference between Africa and Qinghai-Tibet plateau. In the Qinghai-Tibet plateau area, geomorphological environment contains mountain, forest, meadow, river and cavern. Especially in the southeast of Qinghai-Tibet plateau, Hengduan mountains may provide apes with good shelters from glacial frigid. But in Africa, grassland was the main place for apes to evolve and the living conditions there were relatively unitary and

harsh. This may be the key element leading to a much faster evolution rate for Africa apes.

The *Pan paniscus* (Bonobo) only lives in the geographic area south to the Congo River, whilst the north area is occupied by the *Pan troglodytes* (chimpanzee). Even chimpanzees live in only several places in Africa. It is likely that *Pan paniscus* is native animal though its ancestor may come from Asia. *Pan paniscus* can learn hundreds even thousands of human language terms, but they cannot understand grammar, demonstrating that their language levels are never better than a 3-5 years old human baby. It also suggests that *Pan paniscus* may be only one step away from *Homo habilis* and *Homo erectus*, but several steps away from *Homo sapiens* in human evolution. It has been evidenced that the time point when the early anthropoids started to possess modern human-like brain structure was at 1.5~1.7 Ma BP [18,19], represented by significant brain size increase; and this time point was 4.3~4.5 Ma far away from the time when apes disappeared in Qinghai-Tibet plateau (6 Ma BP [12]). So, it can be postulated that ancestors of *Pan paniscus*, after leaving the Tibet plateau for Africa, spent at least 4.3~4.5 Ma to transform into *Homo* genus (*Homo habilis* or *Homo erectus*). Consistent with this calculation, fossils of *Homo habilis* and *Homo erectus* was found gradually in Africa and other continents and given a age record of about 1.5~2.4 Ma BP and 1.9 Ma BP, respectively. Actually, there was another phase of significant brain size increase around 0.7Ma to 10 Ky, which should promote the transformation from *Homo erectus* to *Homo sapiens*, and the language ability of *Homo* genus shall be structurally fulfilled after the long run of natural evolution [18].

## 1.2. Language Gene Polymorphism Pattern of Pan Paniscus

Study on language gene polymorphism patterns (LGPP) across different populations could provide important information on human evolution. In figure 1, we adopted 148 single nucleic polymorphism (SNP) sites from 13 language genes, each with 4-13 SNPs. These 148 SNPs were screened across 113 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). Roughly four distinct LGPPs were preliminarily recognized across human evolution history (figure 1).



**2. Methods**

**2.1. Language Genes and Their SNPs**

Language gene SNP data were all randomly selected for each gene in the dbSNP database: <https://www.ncbi.nlm.nih.gov/>

snp/; Table 1 listed 13 language genes (as a preliminary observation, only 13 language genes were employed at the time the manuscript was written), and a total 148 SNPs (Table 1s) from these 13 genes were selected for this study.

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

**Table 1: Language Genes Employed in this Study**

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 rs535454043	FOXP2-8	FOXP2 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

Table 1s: Selected 148 SNPs of Thirteen Language Genes

## 2.2. Sample Genome Sequences

All genome sequences (Table 2) were downloaded from ENA database (<https://www.ebi.ac.uk/ena/browser/>). Total 113 whole genomes (including 59 ancient genomes) from 5 continents (Africa, Asia, Europe, North America, and South

America) were collected, among which, there are 27 from East Asia (China), 10 from Nepal, 12 from other South Asia countries, 21 from Africa, 28 from Europe, 9 from South America, 2 from North America, 2 from Southeast Asia, 1 from Middle Asia and 1 from West Asia.

	abbr	Country	Region	Age (BP)	Detail	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	117	PRJEB36297
19	c8	China (a)	EastAsia	7000	ZLNR-2 ANCIENT CHINA	17	PRJEB36297
20	c7	China (a)	EastAsia	7000	WQM4 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	Bangladesh	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		SindhiPakistan	101	PRJEB9586
44	sr2	Sri Lanka	SouthAsia		SriLankan No.2	101	
45	pa1	Pakistan	SouthAsia		BalochiPakistan	67	PRJEB9586

46	vn1	Vietnam	SouthAsia		Vietnam-HG2080	123	PRJEB9586
47	sr1	Sri Lanka	SouthAsia		SriLankan-1	102	
48	in1	India	SouthAsia		60-Indian MIX1	82	PRJEB16019
49	in2	India	SouthAsia		GujaIndian MIX1	109	
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 BantuKenya	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 2569	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 individuala	10	PRJNA46161
65	et1	Ethiopia(a)	Africa	4500	Ancient Ethiopian 'Mota' genome	31	
66	ss3	sub-Sahara (a)	Africa	3160	African foragers I10873_new+I10874_new	25	PRJEB49291
67	ss2	sub-Sahara (a)	Africa	7900	African foragers I10871_new	42	PRJEB49291
68	ss1	sub-Sahara (a)	Africa	4500	African foragers I5950_new	51	PRJEB49291
69	mo1	Morocco(a)	Africa	15000	Ancient Morocco genome	18	
70	bu1	Bulgaria	Europe		Bulgarian SAMEA3302842 SAMEA3302718	110	PRJEB9586
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 nd5n	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 ancimnt	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	
97	la1	Latvia (a)	Europe	6179-5750	ZVEJ31 13G Latvia	13	
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		MexicanLosAngeL-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 SAMEA3302650	77	PRJEB9586
111	ta1	Tajikistan	MAsia		Tajikistan Tajik	47	PRJEB9586
112	is1	Israel	WAsia		Israel	53	PRJEB9586
113	pp1	Congo	Africa		Salonga Pan Paniscus	63	SRR741770 SRR741768 SRR741785

Table 2: The 113 Whole Genomes Employed in this Study

### 2.3. Sample SNP Information Abstraction and PCA Analysis

The authors used 010Edit software to extract all 148 SNP information from each genome (Data can be extracted from the author). In all 113 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 and ggplot2. The main R codes are listed as follow.

#### • The main R codes

```
>library(FactoMineR)
>library(factoextra)
>library(ggplot2)
>country <- read.delim('C:/RBook/20220516fastqSNPdata.txt', row.names = 1, sep = '\t')
>country <- t(country)
>country.pca <- PCA(country, ncp = 2, scale.unit = TRUE, graph = FALSE)
>plot(country.pca)
```

```
>pca_sample <- data.frame(country.pca$ind$coord[,1:2])
>head(pca_sample)
>pca_eig1 <- round(country.pca$eig[1,2], 2)
>pca_eig2 <- round(country.pca$eig[2,2], 2)
>pca_eig1
>pca_eig2
>group <- read.delim('C:/RBook/group3.txt', row.names = 1, sep = '\t', check.names = FALSE)
>group <- group[rownames(pca_sample),]
>pca_sample <- cbind(pca_sample, group)
>pca_sample$samples <- rownames(pca_sample)
>head(pca_sample)
>library(ggplot2)
>ggplot(data = pca_sample, aes(x = Dim.1, y = Dim.2))
+geom_point(aes(color = group), size = 3) + scale_color_manual(values = c('purple', 'red', 'green', 'blue', 'brown', 'pink', 'yellow', 'orange', 'grey')) + theme(panel.grid = element_blank(), panel.background = element_rect(color = 'black', fill = 'transparent'), legend.key = element_rect(fill
```

```
= 'transparent')) + labs(x = paste('PCA1:', pca_eig1, '%'), y = paste('PCA2:', pca_eig2, '%'), color = '') + geom_text_repel(aes (label = samples), size = 3, show.legend = FALSE, box.padding = unit(0.25, 'lines'))
```

The LGPP of Bonobo is very close to ss2, pa1, ke1, ke3 and co1. The ss2 was 7900y old ancient sub-Saharan African forager sample (PRJEB49291, genome file size: 42G); pa1 was a Balochi Pakistan sample (PRJEB9586, genome file size: 67G); ke1 and ke3 represented Luhya-Kenya (PRJEB31736, genome file size: 104G) and Bantu-Kenya (PRJEB9586, genome file size: 21G) samples, respectively. co1 was a modern Congo sample (PRJEB9586, genome file size: 17G). This result is apparently an indirect evidence that the ancestor of Bonobo originated from Pakistan (within southeastern margin of the Tibetan Plateau) moved to Congo and Kenya areas [19]. Some native human populations in Pakistan, Kenya and Congo still keep the ancient LGPP with minor changes till now. This result is also a confirmation that Bonobo is very close to *Homo sapiens* in the context of language ability evolution, and probably the direct ancestor of *Homo habilis* or *Homo erectus*.

### 3. Results and Discussion

It is well known that human originates from Africa. This statement is general and crude. When detailed information is needed in both scientific research and cultural education, inaccuracy in this statement becomes obvious. The better statement may be "Anthropoids originate from Asia and *Homo sapiens* originate from Africa". High-level language ability is unique for *Homo sapiens*. There must be several key points for human language evolution within the process line of apes--anthropoids--*Homo habilis*—*Homo erectus*—*Homo sapiens*. *Pan paniscus* is not capable of grasping high-level language ability such as grammar; Its invalid vocalization for human language speaking is also a problem though its learning ability and thinking behavior seem all decent [20,21].

It is relatively easy to understand the LGPP of *Pan paniscus* is highly similar to ancient or modern human samples from the central Africa areas, because the habitat of *Pan paniscus* itself is within this region. However, it is not easy to decipher why the LGPP of *Pan paniscus* is highly similar to some modern people in Pakistan. It is well known that Pakistan was the birthplace of *Sivapithecus*, an important anthropoid that was also found in several Africa sites including Kenya [15]. One possibility is that ancient *Sivapithecus* moved to Africa from the Qinghai-Tibet plateau area and evolved into *Pan paniscus*, while some *Sivapithecus* had a chance to stay in Pakistan and evolve into human populations, the LGPP of which was relatively stable till now for at least 6 million years. Another subtle possibility is, those *Homo* genus that moved out of Africa into Asia (including Pakistan area) had very close relationship with *Pan paniscus*, though the latter never had a chance to become real *Homo*. By the way, it is not strange that LGPP could be preserved for very long time during human history, as exemplified by the c5 sample (figure 1), which represents the modern orogen population in northeast Asia but possesses LGPP-1 exactly the same

as nd2, nd3, nd6 and nd7 (Neanderthals) that were almost 120,000 years old.

The results and postulations in this study are preliminary. The randomly selected 148 SNP sites from 13 language genes may be not enough to accurately distinguish the potential LGPPs. Future investigations would employ all known SNPs or SNVs of all known language genes [21,22].

**Figure Legends:** Figure 1 PCA results for 113 human genomes in the context of 148 language gene SNPs. From left to right, there are roughly four aggregations of samples representing LGPP-1, LGPP-2, LGPP-3 and LGPP-4, respectively (also see ref. 21). *Pan paniscus* (pp1) is at the central position of the figure.

### Acknowledgments

This study was supported by State Language Commission Research Grant (YB135-117) and National Research Center for Foreign Language Education Grant (ZGWYJYJJ10A042).

### Declaration

The authors declared that there are no potential conflicts of interest for this study.

### References

- Ni, X., Gebo, D. L., Dagosto, M., Meng, J., Tafforeau, P., Flynn, J. J., & Beard, K. C. (2013). The oldest known primate skeleton and early haplorhine evolution. *Nature*, 498(7452), 60-64.
- Beard, C. (2004). *The hunt for the dawn monkey: unearthing the origins of monkeys, apes, and humans*. Univ of California Press.
- Beard, K. C. (2006). Primate Biogeography, eds Lehman SM, Fleagle JG.
- Jaeger, J. J., Thein, T., Benammi, M., Chaimanee, Y., Soe, A. N., Lwin, T., ... & Ducrocq, S. (1999). A new primate from the middle Eocene of Myanmar and the Asian early origin of anthropoids. *Science*, 286(5439), 528-530.
- Ni, X., Wang, Y., Hu, Y., & Li, C. (2004). A euprimate skull from the early Eocene of China. *Nature*, 427(6969), 65-68.
- Ni, X., Gebo, D. L., Dagosto, M., Meng, J., Tafforeau, P., Flynn, J. J., & Beard, K. C. (2013). The oldest known primate skeleton and early haplorhine evolution. *Nature*, 498(7452), 60-64.
- Jaeger, J. J., Chavasseau, O., Lazzari, V., Naing Soe, A., Sein, C., Le Maître, A., ... & Chaimanee, Y. (2019). New Eocene primate from Myanmar shares dental characters with African Eocene crown anthropoids. *Nature Communications*, 10(1), 3531.
- Jaeger, J. J., Beard, K. C., Chaimanee, Y., Salem, M., Benammi, M., Hlal, O., ... & Brunet, M. (2010). Late middle Eocene epoch of Libya yields earliest known radiation of African anthropoids. *Nature*, 467(7319), 1095-1098.
- Wayman, E. (2012). Out of Asia: How monkey and ape ancestors colonized Africa. *Smithsonian, June*, 4.
- Choi, C. Q. (2010). Out of Asia: new origin proposed for humans, monkeys, apes. *Live Science, October*, 27.
- Daver, G., Guy, F., Mackaye, H. T., Likious, A., Boissarie, J.

- R., Moussa, A., ... & Clarisse, N. D. (2022). Postcranial evidence of late Miocene hominin bipedalism in Chad. *Nature*, 609(7925), 94-100.
12. Li, S., Ji, X., Harrison, T., Deng, C., Wang, S., Wang, L., & Zhu, R. (2020). Uplift of the Hengduan Mountains on the southeastern margin of the Tibetan Plateau in the late Miocene and its paleoenvironmental impact on hominoid diversity. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 553, 109794.
13. Xiong, Z., Liu, X., Ding, L., Farnsworth, A., Spicer, R. A., Xu, Q., ... & Yue, Y. (2022). The rise and demise of the Paleogene Central Tibetan Valley. *Science Advances*, 8(6), eabj0944.
14. Zanolli, C., Kullmer, O., Kelley, J., Bacon, A. M., Demeter, F., Dumoncel, J., ... & Macchiarelli, R. (2019). Evidence for increased hominid diversity in the Early to Middle Pleistocene of Indonesia. *Nature ecology & evolution*, 3(5), 755-764.
15. Leakey, R. E. F., & Walker, A. (1985). New higher primates from the early Miocene of Buluk, Kenya. *Nature*, 318(6042), 173-175.
16. Habinger, S. G., Chavasseau, O., Jaeger, J. J., Chaimanee, Y., Soe, A. N., Sein, C., & Bocherens, H. (2022). Evolutionary ecology of Miocene hominoid primates in Southeast Asia. *Scientific Reports*, 12(1), 11841.
17. Macgregor, D. (2015). History of the development of the East African Rift System: A series of interpreted maps through time. *Journal of African Earth Sciences*, 101, 232-252.
18. Gingerich, P. D. (2022). Pattern and rate in the Plio-Pleistocene evolution of modern human brain size. *Scientific Reports*, 12(1), 11216.
19. Ponce de León, M. S., Bienvenu, T., Marom, A., Engel, S., Tafforeau, P., Alatorre Warren, J. L., ... & Zollikofer, C. P. (2021). The primitive brain of early Homo. *Science*, 372(6538), 165-171.
20. Staes, N., Sherwood, C. C., Wright, K., De Manuel, M., Guevara, E. E., Marques-Bonet, T., ... & Bradley, B. J. (2017). FOXP2 variation in great ape populations offers insight into the evolution of communication skills. *Scientific reports*, 7(1), 16866.
21. Pilcher, H. (2005). Apeing our language. *Nature*.
22. Xia, W., & Zhang, Z. (2022). Language gene polymorphism pattern survey provided important information for education context in human evolution. *bioRxiv*, 2022-10