

Population History and Genome Wide Association Studies of Birth Weight in a Native High Altitude Ladakhi Population

High Altitude Genetics Explored Through Birth Weight Protection

Sushil Bhandari¹

Padma Dolma²

Mitali Mukerji³

Bhavana Prasher³

Hugh Montgomery⁴

Dalvir Kular⁵

Vandana Jain⁶

Vatsla Dadhwal⁶

David J Williams⁷

Aniket Bhattacharyya³

Edmund Gilbert¹

Gianpiero L Cavalleri¹

Sara L Hillman⁷

¹Royal College of Surgeons in Ireland, Dublin, Ireland

²Sonam Norboo Memorial Hospital, Leh, Ladakh, India

³Council Scientific Industrial Research-Institute for Genomics and Integrative Biology, New Delhi, India

⁴Centre for Human Health and Performance, University College London, London, United Kingdom

⁵King's College Hospital, London, United Kingdom

⁶All India Institute for Medical Sciences, New Delhi, 110029 India

⁷University College London Institute for Women's Health, London, United Kingdom

Corresponding author: sara.hillman@ucl.ac.uk (SH)

Abstract

Pathological low birth weight due to fetal growth restriction (FGR) is an important predictor of adverse obstetric and neonatal outcomes. It is more common amongst native lowlanders when gestating in the hypoxic environment of high altitude, whilst populations who have resided at high altitude for many generations are relatively protected. Genetic study of pregnant populations at high altitude permits exploration of the role of hypoxia in FGR pathogenesis, and perhaps of FGR pathogenesis more broadly.

We studied the umbilical cord blood DNA of 316 neonates born to pregnant women managed at the Sonam Norboo Memorial Hospital, Ladakh (altitude 3540m) between February 2017- January 2019. Principal component, admixture and genome wide association studies (GWAS) were applied to dense single nucleotide polymorphism (SNP) genetic data, to explore ancestry and genetic predictors of low birth weight.

Our findings support Tibetan ancestry in the Ladakhi population, with subsequent admixture with neighboring Indo-Aryan populations. Fetal growth protection was evident in Ladakhi neonates. Although no variants achieved genome wide significance, we observed nominal association of seven variants across genes (*ZBTB38*, *ZFP36L2*, *HMG2*, *CDK1*, *PLCG1*) previously associated with birthweight.

Introduction

Birth weight is a complex trait driven by metabolic, vascular and immune interactions between mother and fetus. Successful placental and fetal growth leads to appropriate birth weight, reduced neonatal morbidity and ultimately driving success of a species [1-3]. Fetal growth restriction (FGR), defined as an inability for a fetus to reach its growth potential, leads to lower birth weight, and compromised neonatal survival [1,4-6].

Studies support a genetic influence on birth weight [7]. Genome wide association studies have identified approximately 70 single nucleotide polymorphisms (SNPs) of robust influence [8], although these studies have primarily related to European ancestry cohorts of normal birth weight. Exploring the genetic contribution to *in-utero* fetal growth, particularly in populations where evolutionary selection may have resulted in protected birth weight from negative environmental influences, has the potential to enhance understanding of FGR pathogenesis, and thus to suggest new protective strategies.

Although FGR is more prevalent in those residing above 1500m (high altitude, where oxygen availability is reduced) [4,9,10], populations who have resided at high altitude for many generations appear to be protected [11,12]. For example, babies born to (native ancestral high altitude) Tibetan women at altitudes above 3000m are more than 500g heavier than those born to native lowland Han Chinese [13]. Similarly, in La Paz, Bolivia (3,600m), native Andean babies are born heavier than their European counterparts [14].

To date, genetic studies of birth weight at high altitude have primarily focused on maternal genotypes. The maternal *PRKAA1* gene locus (coding for *AMPK*, a central regulator of cellular

energy metabolism) has been associated with birth weight and maternal uterine artery diameter in high altitude Andean residents [15]. Less is known about fetal genotype and its association with birth weight at high altitude in Tibetan ancestry populations.

Ladakh, in the Jammu and Kashmir region of India, lies between the Karakoram and Himalayan mountain ranges. The term “La-dvags” in Tibetan means “land of high passes” and “Ladakh” is the Persian spelling [16]. The highland Ladakh region connects South Asia with the Tibetan plateau via the ancient trade corridor- “the Silk Road”. Genetic adaptation to the hypoxia of high altitude is well described[17]. Although sharing similar linguistic, cultural and religious practices with Tibetans [16,18], the population structure and genetic selection of the Ladakhi population is relatively poorly studied. The Sonam Norboo Memorial Hospital in Leh [3,540m], which provides maternity care for Ladakh, has an institutional birth rate of >90% [19] making it a unique site to study a pregnant population at high altitude.

We hypothesized that the fetal genotype of native Ladakhis would be enriched for gene variants which protect *in-utero* growth despite the challenging environment of high altitude gestation. We first performed a population genetic survey before seeking to identify fetal genetic elements associated with birth weight in the Ladakhi population.

Materials and Methods

Ethical permissions were granted for the study from the Indian Health Ministry's Screening committee (HMSC, 7thSeptember 2016), the Office of the Chief Medical Officer Leh (3rdAugust 2016), the All India Institute for Medical Sciences, and the University College London Research Ethics Committee (3634/002). All methods were performed in accordance with the relevant guidelines and regulations and informed consent was obtained from all participants and/or their legal guardians. The research was performed in accordance with the Declaration of Helsinki. Pregnant women and their partners were recruited from Sonam Norboo Memorial (SNM) Hospital, Leh [altitude 3,540m] from antenatal clinics and the ultrasound department, over a 2-year period (February 2017-January 2019).

Pregnancies were included if 1) both parents were aged over 18 years and unrelated (defined as first cousin or closer), 2) they were singleton, 3) the mother planned to have their baby at the SNM hospital and 4) the estimated due date could be calculated from last menstrual period or dating ultrasound. Those pregnancies where the fetus had a clinically obvious fetal structural or chromosomal abnormality were excluded. Women completed a questionnaire documenting their social (including nutritional), family, obstetric and medical histories (including smoking, alcohol, chronic medical problems and medications). These data have already been reported by our group [20]. Only subjects who reported that their parents and grandparents (on both sides) were born in or close to Leh at altitudes >2,500m were considered as Ladakhi and thus recruited to the study.

Following birth and delivery of the placenta and umbilical cord, at least three 1ml aliquots of residual whole umbilical cord blood, with a separate serum sample, were frozen immediately at

-80°C. Information concerning the birth process (mode of delivery) and the neonatal characteristics routinely collected by the hospital (sex, weight, head circumference, crown-heel length and APGAR score) were recorded. The cord blood (and whole blood from their parents for future testing) was then shipped to Delhi and genomic DNA extracted at the Institute for Genomics and Integrative Biology, Delhi using commercial kits (Qiagen, Maryland, USA) and validated, standardized protocols. Common bi-allelic SNP genotyping was performed using Illumina Global Screening Array SNP-microarray technology (Illumina, San Diego, California, USA). Genotype assignment from the microarray fluorescence data was performed using Illumina's Genome Studio software.

We combined the genotypes of 316 Ladakhi individuals with reference individuals from surrounding populations living at high and low altitude. Human Genome Diversity project [21] datasets were included for the following different populations; Burusho, Han, Hazara, Indo Aryan, Japanese, Yakut Siberian. Similarly other population datasets included in this study were Munda [22], Sherpa [23], Tibetan [24] and Tibeto-Burman [25]. These individual cohorts were merged using autosomal SNPs found in common between all separate datasets. This merged dataset was then processed through a number of quality control steps using the software PLINK 1.9[26,27]. We included only individuals or SNPs that had < 5% missing genotypes, SNPs with a minor allele frequency (MAF) >1%, and SNPs with a HWE at significance of >1e⁻⁶. Furthermore, we calculated closely related individuals using King [28], pruning one individual from a pair related by 2nd degree or closer. This left a final dataset of 1,413 individuals. Due to the variety of genotyping platforms between the different population references, the final set of common SNPs contained 60,280 markers.

We calculated principal components using PLINK [26], first pruning SNPs in linkage disequilibrium using PLINK's `--indep-pairwise` command using a window size of 1000, moving 50 SNPs, and using an r^2 threshold of 0.2 – leaving 41,718 SNPs in approximate linkage.

We conducted a genetic clustering analysis of Ladakhi populations with its neighboring populations using ADMIXTURE v1.2[29], estimating individual ancestry proportions in an unsupervised analysis. The same cohort assembled in PCA was used for ADMIXTURE analysis, i.e., non-missing autosomal SNPs, individuals filtered for relatedness, and SNPs pruned for linkage disequilibrium. Unsupervised ADMIXTURE analysis was carried out over $k=2-4$ populations.

We further investigated evidence of admixture within the Ladakh population's history by applying f -statistics [30] testing the strength of evidence that the Ladakh population is admixed between pairs of neighboring populations using the f_3 implementation within *ADMIXTOOLS 2* (<https://uqrmaie1.github.io/admixtools/index.html> - a manuscript describing *ADMIXTOOL 2* is currently under preparation by its authors). The same cohort of individuals and SNPs that were used for PCA and ADMIXTURE analysis was used for f_3 -statistic work.

We detected Runs of Homozygosity (ROH) [31] in Ladakhi individuals and compared population-averages to individuals from neighboring lowland and highland populations using PLINK's `--homozyg` command, with the following specific parameters; `--homozyg --homozyg-window-snp 50 --homozyg-snp 50 --homozyg-kb 1500 --homozyg-gap 1000 --homozyg-density 50 --homozyg-window-missing 5 --homozyg-window-het 1`. This was carried out on a subset of the individuals described in principal component, ADMIXTURE, and f_3 analyses, selecting individuals who were genotyped on SNP-microarray chips with an overlap of common SNPs >100K in number to faithfully detect ROH. This subset included individuals from the

Indian Indo-Aryan, Ladakhi, Sherpa, Tibeto-Burman Bhutanese, Tibeto-Burman Nepalese, or Tibetan population labels. Common SNPs between these individuals were filtered according to the same parameters as in the PCA, leaving 117,044 common SNPs. Genome-wide association tests were conducted using 601,887 genotyped SNPs. Linear regression was applied, implemented in PLINK v1.9 with the--linear command. We applied an additive genetic model adjusting sex and the first four principal components obtained from genome-wide SNP data. Gene variants associated with birthweight in the Ladakhi were compared to variants in the Global Biobank Engine [32] to confirm their association with birthweight in other populations. Sample size calculations using algorithms incorporated in GWA power [33] were based on mean birth weights obtained from audit data collected in Leh prior to the study commencing. Calculations identified the study had an 80% power to detect a variant that would explain 11.4% (or greater) of the variability observed in birth weight given a discovery sample size of 300 neonates and alpha level of 5×10^{-8} .

Results

In total 316 eligible women were recruited and studied to delivery. Maternal and fetal baseline characteristics are summarized in Table 1.

Table 1 Maternal and fetal characteristics

Characteristic	Term>37 <42+1 weeks n=301	Preterm<37 weeks n=15	Mean n=316
Maternal age (years)	29.1 [18/40]	26 [20/35]	29.0[18/40]
Weight (kg)	57.0 [37/92]	56.9 [40/82]	57.0[37/92]
Hb (g/dl)	12.7 [10.4/14.8]	12.3 [7.5/17]	12.3 (7.5/17)
Primiparous n (%)	111 (36.9)	6 (40)	117 (37)
Birth weight (kg)	3.18[1.98-4.40]	2.45[1.5-3.2]	3.15[1.5-4.4]
Birth weight centile (%)	44.5 [0.13/98.7]	49 [1.5/99]	44.9 [0.13/99.9]
<10 th centile (%)	42 (14)	2 (13.3)	44 (13.9)
Tibetan by history kg (centile) [n=10]			3.6 (70)
Ladakhi by history kg (centile) [n=306]			3.17 (43.9)

Characteristics are summarized as mean [min/max] with birth weight of infant (kg) divided into those born at term gestations of >37 weeks and those born preterm at <37 weeks, birth weight centile (%) derived from Intergrowth, and those born at <10th centile recorded (defined as small for gestational age) and ancestry by history (Ladakhi and Tibetan).

Mean birth weight was 3.18kg (mean birth weight centile 44.5th) in term individuals (>37 weeks' gestation). According to intergrowth charts(<https://intergrowth21.tghn.org>), 14% of infants were born at less than the 10th centile (which defines a small for gestational age [SGA] newborn) and only 7.8% of all infants were born with a low birth weight (LBW, defined as less than 2.5kg). The average age for pregnant women was 28.9 years and 37.9% (120/316) were primigravida. All

women reported 2, and most (96.8%) >3, generations of Ladakhi ancestry (as defined above). Ten women reported Tibetan ancestry: all gave birth at term, with a mean birth weight of 3.6 kg. Preterm birth was infrequent: 15/316 [4.7%] infants were born before 37 weeks, whose birth weight was, as expected, lower than those born at term (mean 2.4 vs 3.2kg, $P=0.0001$) although birth weight centile was similar (52nd vs 45th respectively, $p=0.324$) suggesting appropriate growth for gestational age. Recognizing that being preterm would confound birth weight if used in isolation, for further analysis we instead used birth weight centiles as a more useful measure of growth potential, given that it adjusts for gestational age.

Reconstruction of Ladakhi population history

We first sought the history and structure of the Ladakhi population from the perspective of genetics. We utilized our dataset of 316 Ladakhi neonates and compared it with other surrounding populations based on major language groups including Tibeto-Burman (Tibetans, Sherpa, Ladakhi), Indo-European (Indo-Aryan, Hazara), Austro-Asiatic (Munda) and isolate languages (Bursushaski) (Fig. 1) using Principal Component Analysis (PCA). To understand the nature and extent of Himalayan ancestry in the Ladakhi population, we compared them with high altitude populations (Sherpa, Tibetans) and other Tibeto-Burman populations adjacent to the foothills of the Himalayas in Nepal and Bhutan.

Fig 1. Principal Component Analysis (PCA) of Ladakhi population

PCA of Ladakhi and its neighboring populations are indicated in the figure legend. First and second components of PCA analysis with individual clusters labeled by different colours.

In the PCA, the Ladakhi individuals project together, forming a genetic cluster that is distinct from the neighboring Indo-Aryan Indian population. This Ladakhi cluster forms a genetic cline between the two termini of principal component 1, which separates Indo-Aryan Indian individuals from individuals with Himalayan or East Asian ancestry. The second principal component appears to separate out Han (East Asian) individuals from Sherpa (high altitude), where Ladakhi individuals are placed towards the high-altitude-associated Sherpa population alongside individuals from Himalayan regions such as Tibet, Nepal, and Bhutan. The intermediate position of Ladakhi individuals on principal component 1 is suggestive of admixture between Indo-Aryan and Himalayan populations. The position of Ladakhi individuals on principal component 2 suggests the Himalayan genetic group that the Ladakhi share affinity to is Tibetan or Sherpa with no evidence in PCA of admixture with Han-related East Asian ancestries.

We further contextualized our population structure analysis using the maximum-likelihood estimation of individual ancestries using ADMIXTURE (Fig. 2). At $k=2$ the two ancestral components are maximized in either Himalayan or the East Asian Han populations (in red), and Indo-Aryan Indians (green). We observed that Ladakhi individuals are modelled with a slightly higher proportion of the East Asian ancestry than South Asian ancestral component, in agreement with their average location on principal component 1. At $k=3$, one ancestry component, represented by blue, is maximized in the East Asian Han, the second (green) maximized in Sherpa, and the third (red) in Indo-Aryan Indians. The majority average component in Ladakhi individuals is the Sherpa-maximized ancestry (52.4%), then the Indo-Aryan-maximized component (32.3%), suggesting the majority of Ladakhi are closer to Tibeto-Burmans than Indo-Aryans. At $k=4$ an ancestry component separates ancestry which is maximized in Tibetan-Bhutanese individuals from other East Asian sources. This component (in purple) is also found in appreciable proportions in

other high-altitude populations. Ladakhi individuals also show appreciable proportions of this ancestry component common in both Nepalese Tibeto-Burman and Tibetans. However, the blue ancestry component at $k=4$ in Ladakhi individuals continues to support admixture with other South Asian population such as Hazara, Burushoor Indo-Aryan Indians. Thus, Ladakhi appear genetically closer to Tibeto-Burman speakers of Himalayan region than Indo-Aryan populations of South Asia. Long-term isolation in remote highland, practices of endogamy culture and a founder effect might have added greatly to the formation of distinct genetic cluster in the Ladakhi agreeing with PCA result (Fig. 1). In summary, the data supported Ladakhi as a genetically distinct Himalayan population closer to Tibeto-Burman population and Tibetan populations than to Indo-Aryan.

Fig 2. ADMIXTURE analysis of the Ladakhi population

ADMIXTURE analysis for $K=2$ (top), $K=3$ (middle) and $K=4$ (bottom). ADMIXTURE analysis shows Ladakhi contain higher number of Tibeto-Burman ancestries and are relatively closer to Tibetans.

We next performed two additional analyses to further confirm that the Ladakhi are admixed between lowland Indo-Aryan and highland Tibeto-Burman sources. Firstly, we leveraged f -statistics [30] in the form of $f_3(X; A, B)$ where X is tested for evidence of admixture between sources A and B , where a negative f_3 -statistic is indicative of admixture. We placed Ladakhi as X , and tested combinations of lowland/Indo-Aryan populations (Burusho, Indo_Aryan, Hazara, Munda) as X and highland/East or North Asian populations (Han, Japanese, Sherpa, Tibetan,

TB_Bhutanese, TB_Nepalese, Yakut_Siberian) as Y, reporting those f_3 results with an absolute Z score >3 (Fig. 3).

Figure 3. Evidence of admixture in Ladakhi between lowland or Indo-Aryan populations and highland or north/east Asian populations measured using f_3 outgroup statistics.

Along the x axis is the f_3 statistic score. The more negative that statistic is, the greater is evidence of admixture. The error bars correspond to three standard errors of the estimated f_3 statistic. Along the y axis are different pairs of putative source populations for an admixture event which creates the modern Ladakhi population.

The strongest evidence of admixture (most negative f_3 -statistic) is between Tibetan or Sherpa sources and Indo-Aryan or Burusho sources – supportive of results from PCA. Additionally, we performed Runs of Homozygosity detection using PLINK, comparing Ladakhi to a subset of neighbouring lowland and highland references. We detect elevated ROH (S1 Fig.) in the high altitude Tibeto-Burman and Sherpa populations, agreeing with previous estimates [23], but only modest levels of ROH in Ladakhi more comparable to the general Tibetan or Indo-Aryan labelled individuals. These modest levels of ROH would be consistent with admixture between different ancestries.

Birth weight analysis on Ladakhi subjects

In our analysis in reconstructing the genetic history of the Ladakhi population, we determined that our dataset is a homogenous sample of genetic variation within the Ladakhi population. We

therefore leveraged this homogenous sample of 316 Ladakhi neonates to determine if there were any genome-wide significant predictors of birth weight in the Ladakhi population, performing a GWAS of birth weight. We did not find any signals that were genome-wide significant after correction for multiple testing (i.e. $p < 5 \times 10^{-8}$) (S2 Fig.). However, when we looked in the tail of the association statistics from our birth weight GWAS (i.e. $p = 1 \times 10^{-4}$ to 1×10^{-7} uncorrected), we noted the presence of multiple variants previously associated with haematological/cardiovascular traits of potential relevance to high altitude adaptation: rs16893892 in *RP5-874C20.3* associated with mean corpuscular hemoglobin, rs2298839 in *AFP* associated with mean platelet volume, rs9261425 in *TRIM31* associated with WBC count and rs362043 in *TUBA8* associated with atrial fibrillation [32].

Next, in order to determine whether the genetic architecture of birth weight in the Ladakhi population overlaps with that observed in lowland populations, we sought the association with birth weight in Ladakhis of the 70 genetic signals previously associated with birth weight in lowland populations [8]. Overall, 32 of these 70 signals were either genotyped or captured through linkage disequilibrium ($r^2 > 0.8$) in our Ladakhi dataset. Seven of these 32 signals were significantly associated ($P < 0.05$, uncorrected) with birth weight in the Ladakhi GWAS study (Table 2). These variants mapped to five different birth weight associated genes (*ZBTB38*, *ZFP36L2*, *HMGA2*, *CDKALI* and *PLCG1*) with the direction of effect being consistent with the original report [8] in all seven cases.

Table 2 Single nucleotide polymorphism (SNP) of Ladakhi neonates compared with UK biobank

SNP	Mapped Gene	SNP_Annotation	P-values			Beta values	
			Ladakhi	UK biobank	GWAS Catalog	GWAS_Catalog	Ladakhi
rs6440006	<i>ZBTB38</i>	Upstream	0.03515	1.099e ⁻⁵	4 x 10 ⁻¹²	-0.021224	-0.0818
rs8756	<i>HMGA2</i>	3_prime_UTR	0.000078	1.97e ⁻¹³⁸	1 x 10 ⁻¹⁹	-0.027582	-0.176
rs1351394	<i>HMGA2</i>	intronic	0.000078	3.28e ⁻¹³⁷	2 x 10 ⁻³³	-0.043	-0.176
rs7968682	<i>HMGA2</i>	intergenic	0.000078	4.97e ⁻¹³⁰	4 x 10 ⁻⁶⁰	-0.041831	-0.176
rs4952673	<i>ZFP36L2</i>	intergenic	0.04	NA-not available	2 x 10 ⁻¹¹	-0.020476	-0.09315
rs35261542	<i>CDKAL1</i>	intronic	0.01864	NA	3 x 10 ⁻⁴⁵	-0.040621	-0.08755
rs753381	<i>PLCG1</i>	missense variant	0.02988	NA	3 x 10 ⁻⁹	-0.01512	-0.0792

Single Nucleotide Polymorphism rs (rs) number and gene location along with P and Beta values

Discussion

This study is the first to undertake a comprehensive fetal genotypic exploration of a high altitude population in relation to a birth weight phenotype. We showed that the Ladakhi are more closely related to Tibeto-Burman speaking populations than to the Indo-Aryan groups of South Asia, providing clear evidence that Tibeto-Burman expansion occurred in North East India crossing the Himalayan range [34]. The majority of Ladakhi form a common clustering PCA, indicating the

presence of a Ladakhi specific genetic ancestry component that is intermediate between Tibeto-Burman and Indo-Aryan ancestries. Analysis utilising f -statistics and ROH data further supports a demographic history of admixture between these sources and suggests a distinct demographic history in the Ladakh region, plausible due to its close proximity with the Tibetan plateau. Lack of large overlap between Ladakhi genotype data and existing references limited haplotype analyses of Identity-by-Descent segments or “ChromoPainter” based methods [35], which limited the power to analyse the demographic history of the population in more detail. Future work overcoming these technical limitations may provide further insights.

Individuals in Leh are born at a significantly higher birth weight than expected, based on existing literature for average birth weights in India at 2.8-3kg [36,37] and seen from previous [20] studies. One such mechanism of protection of birth weight is through genetic selection as evidenced in other phenotypic traits of high altitude populations [17].

Birthweight is a complex trait with potentially conflicting parental inheritance patterns. As such, it is perhaps unsurprising that no one clear signal was identified at GWAS level, especially given the low numbers in the final study. Study of parental DNA and genetic signals in genes responsive to the hypoxia-responsive transcription factors *hypoxia inducible factors (HIFs)* in relation to offspring birth weight is an obvious follow up study to further explore HIF function and high altitude adaptation.

In relation to SNPs previously associated with birth weight by GWAS, that of rs1351394 of the *HMGA2* gene (encoding the high mobility group-A2 protein) was replicated in Ladakhi offspring. High-mobility group (HMG) proteins are ubiquitous nuclear proteins that bind to DNA and induce structural change in chromatin, thereby regulating gene expression [15]. HMGA2 (the gene

encoding the high mobility group-A2 protein) has been associated with human height and birth weight in lowland populations [38,39] and with adipose mass variation in pigs [40], making it a biological plausible functional candidate.

ZBTB38 encodes a zinc finger transcription factor that binds methylated DNA to enhance or repressing transcription in a way which is complex and likely cell-dependent [41]. Its expression appears to play a role in skeletal development [39]. Recent GWAS have reported SNPs in this gene to be associated with human height, perhaps through upregulation of insulin growth factor 2, a potent fetal growth factor [42].

One of the strengths of this study relates to accurate dating of gestation from last menstrual periods and early dating ultrasounds. This allowed for calculation of birth weight centiles which adjusts neonatal weight for gestational age and sex at birth, allowing for better correlation with pathological poor growth. This study focused on fetal genotype, but it is recognized that maternal [43] and paternal genotype also has an effect. We have parental DNA and it would be beneficial in the future to analyze inheritance patterns of these fetal gene variants of interest.

Conclusion

We found Tibetan ancestry in the Ladakhi population, and evidence of recent limited admixture with neighboring Indo-Aryan populations. The effects of a subset of variants known to predict birthweight in European-descent populations was replicated in the Ladakhi population. Replicated variants were focused on body anthropomorphic traits and hematological parameters. Further genomic study of high altitude infants is indicated to validate these findings in conjunction with

mechanistic study to gain insight in potential mechanisms of action that may be amenable to therapeutic intervention.

References

1. Ludvigsson JF, Lu D, Hammarstrom L, Cnattingius S, Fang F (2018) Small for gestational age and risk of childhood mortality: A Swedish population study. *PLoS Med* 15: e1002717.
2. Hales CN, Barker DJ, Clark PM, Cox LJ, Fall C, et al. (1991) Fetal and infant growth and impaired glucose tolerance at age 64. *BMJ* 303: 1019-1022.
3. Gilbert WM, Danielsen B (2003) Pregnancy outcomes associated with intrauterine growth restriction. *Am J Obstet Gynecol* 188: 1596-1599; discussion 1599-1601.
4. Moore LG (2003) Fetal growth restriction and maternal oxygen transport during high altitude pregnancy. *High Altitude Medicine & Biology* 4: 141-156.
5. Barker DJ, Osmond C, Winter P, Margetts B, Simmonds SJ (1989) Weight in infancy and death from ischaemic heart disease. *The Lancet* 334: 577-580.
6. McCowan LM, Figueras F, Anderson NH (2018) Evidence-based national guidelines for the management of suspected fetal growth restriction: comparison, consensus, and controversy. *Am J Obstet Gynecol* 218: S855-S868.
7. Warrington NM, Beaumont RN, Horikoshi M, Day FR, Helgeland Ø, et al. (2019) Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nature Genetics* 51: 804-814.
8. Horikoshi M, Beaumont RN, Day FR, Warrington NM, Kooijman MN, et al. (2016) Genome-wide associations for birth weight and correlations with adult disease. *Nature* 538: 248-252.
9. Julian CG, Wilson MJ, Moore LG (2009) Evolutionary adaptation to high altitude: a view from in utero. *American Journal Human Biology* 21: 614-622.
10. Moore LG, Charles SM, Julian CG (2011) Humans at high altitude: hypoxia and fetal growth. *Respiratory physiology & neurobiology* 178: 181-190.
11. Smith C (1997) The effect of maternal nutritional variables on birthweight outcomes of infants born to Sherpa women at low and high altitudes in Nepal. *Am J of Hum Biol* 9: 751-763.
12. Moore LG, Zamudio S, Zhuang J, Sun S, Droma T (2001) Oxygen transport in Tibetan women during pregnancy at 3,658 m. *Am J of Physic Anthro* 114: 42-53.
13. Moore LG, Young D, McCullough RE, Droma T, Zamudio S (2001) Tibetan protection from intrauterine growth restriction (IUGR) and reproductive loss at high altitude. *Am J Hum Biol* 13: 635-644.
14. Haas JD FE, Stepick C, Beard J, Hurtado L. (1980) Altitude, ethnic and sex differences in birth weight and length in Bolivia. *Hum Biol* 52: 459-477.
15. Hock R, Furusawa T, Ueda T, Bustin M (2007) HMG chromosomal proteins in development and disease. *Trends Cell Biol* 17: 72-79.
16. R JTaT (2018) Cultural Relationship between the People of Ladakh and Tibet. *The Tibet Journal* 43: 65-71.

17. Bhandari S, Cavalleri GL (2019) Population History and Altitude-Related Adaptation in the Sherpa. *Frontiers in Physiol.* Aug 28;10: 1116
18. Rowold DJ, Perez Benedico D, Garcia-Bertrand R, Chennakrishnaiah S, Alfonso-Sanchez MA, et al. (2016) Ladakh, India: the land of high passes and genetic heterogeneity reveals a confluence of migrations. *Eur J Hum Genetics* 24: 442-449.
19. Gutschow K DP (2012) Global policies and local implementation: Maternal mortality in Rural India Kappas M, Gross U, Kelleher D (eds) *Global Health: A Challenge for Interdisciplinary Research* p215-238.
20. Dolma P, Angchuk P, Jain V, Dadhwal V, Kular D, et al. (2022) High-altitude population neonatal and maternal phenotypes associated with birthweight protection. *Pediatric Research: Jan*; 91(1): 137-142
21. Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, et al. (2002) A human genome diversity cell line panel. *Science* 296: 261-262.
22. Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, et al. (2011) Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Mol Biol and Evol* 28: 1013-1024.
23. Cole AM, Cox S, Jeong C, Petousi N, Aryal DR, et al. (2017) Genetic structure in the Sherpa and neighboring Nepalese populations. *BMC Genomics* 18: 102.
24. Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, et al. (2010) Natural selection on EPAS1 (HIF2 α) associated with low hemoglobin concentration in Tibetan highlanders. *Proceedings of the National Academy of Sciences* 107: 11459-11464.
25. Arciero E, Kraaijenbrink T, Asan, Haber M, Mezzavilla M, et al. (2018) Demographic History and Genetic Adaptation in the Himalayan Region Inferred from Genome-Wide SNP Genotypes of 49 Populations. *Mol Biol Evol* 35: 1916-1933.
26. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genetics* 81: 559-575.
27. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, et al. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4: 7.
28. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, et al. (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics* 26: 2867-2873.
29. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 19: 1655-1664.
30. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, et al. (2012) Ancient admixture in human history. *Genetics* 192: 1065-1093.
31. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, et al. (2008) Runs of homozygosity in European populations. *Am J Hum Genetics* 83: 359-372.
32. McInnes G, Tanigawa Y, DeBoever C, Lavertu A, Olivieri JE, et al. (2018) Global Biobank Engine: enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics* 35: 2495-2497.
33. Feng S, Wang S, Chen CC, Lan L (2011) GWAPower: a statistical power calculation software for genome-wide association studies with quantitative traits. *BMC Genetics* 12: 12.
34. Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, et al. (2007) The Himalayas as a directional barrier to gene flow. *Am J Hum Genetics* 80: 884-894.
35. Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genetics* 8: e1002453.

36. Kumar VS, Jeyaseelan L, Sebastian T, Regi A, Mathew J, et al. (2013) New birth weight reference standards customised to birth order and sex of babies from South India. *BMC Pregnancy and Childbirth* 13: 38.
37. Marete I, Ekhaguere O, Bann CM, Bucher SL, Nyongesa P, et al. (2020) Regional trends in birth weight in low- and middle-income countries 2013-2018. *Repro Health* 17: 176.
38. Weedon MN, Lettre G, Freathy RM, Lindgren CM, Voight BF, et al. (2007) A common variant of HMGA2 is associated with adult and childhood height in the general population. *Nature Genetics* 39: 1245-1250.
39. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, et al. (2008) Many sequence variants affecting diversity of adult human height. *Nature Genetics* 40: 609-615.
40. Kim KS, Thomsen H, Bastiaansen J, Nguyen NT, Dekkers JC, et al. (2004) Investigation of obesity candidate genes on porcine fat deposition quantitative trait loci regions. *Obesity Research* 12: 1981-1994.
41. de Dieuleveult M, Miotto B (2018) DNA Methylation and Chromatin: Role(s) of Methyl-CpG-Binding Protein ZBTB38. *Epigenetics insights* 11: 2516865718811117-2516865718811117.
42. Lin YJ, Liao WL, Wang CH, Tsai LP, Tang CH, et al. (2017) Association of human height-related genetic variants with familial short stature in Han Chinese in Taiwan. *Scientific Reports*. Jul 25;7 (10): 6372.
43. Hughes AE, Nodzenski M, Beaumont RN, Talbot O, Shields BM, et al. (2018) Fetal Genotype and Maternal Glucose Have Independent and Additive Effects on Birth Weight. *Diabetes* 67: 1024-1029.

Acknowledgments

This work would not have been possible without the tireless commitment of the staff and patients at Sonam Norboo Memorial Hospital, Leh Ladakh. The work has been supported by several hospital Medical Superintendents, in particular Dr Phunchok Wangchuk Kokluk and by the Chief Medical Officer, Dr Motup Dorje.

481 **Supporting Information**

482 **S1 Fig. Runs of homozygosity(ROH)in different populations**

483 The plot shows the average total length of ROH in each of the populations. The coloured bar shows
484 the arithmetic mean, and the box plots are normal box plots.

485

486 **S2 Fig. Manhattan plot showing GWAS results of birth weight genes across chromosomes**

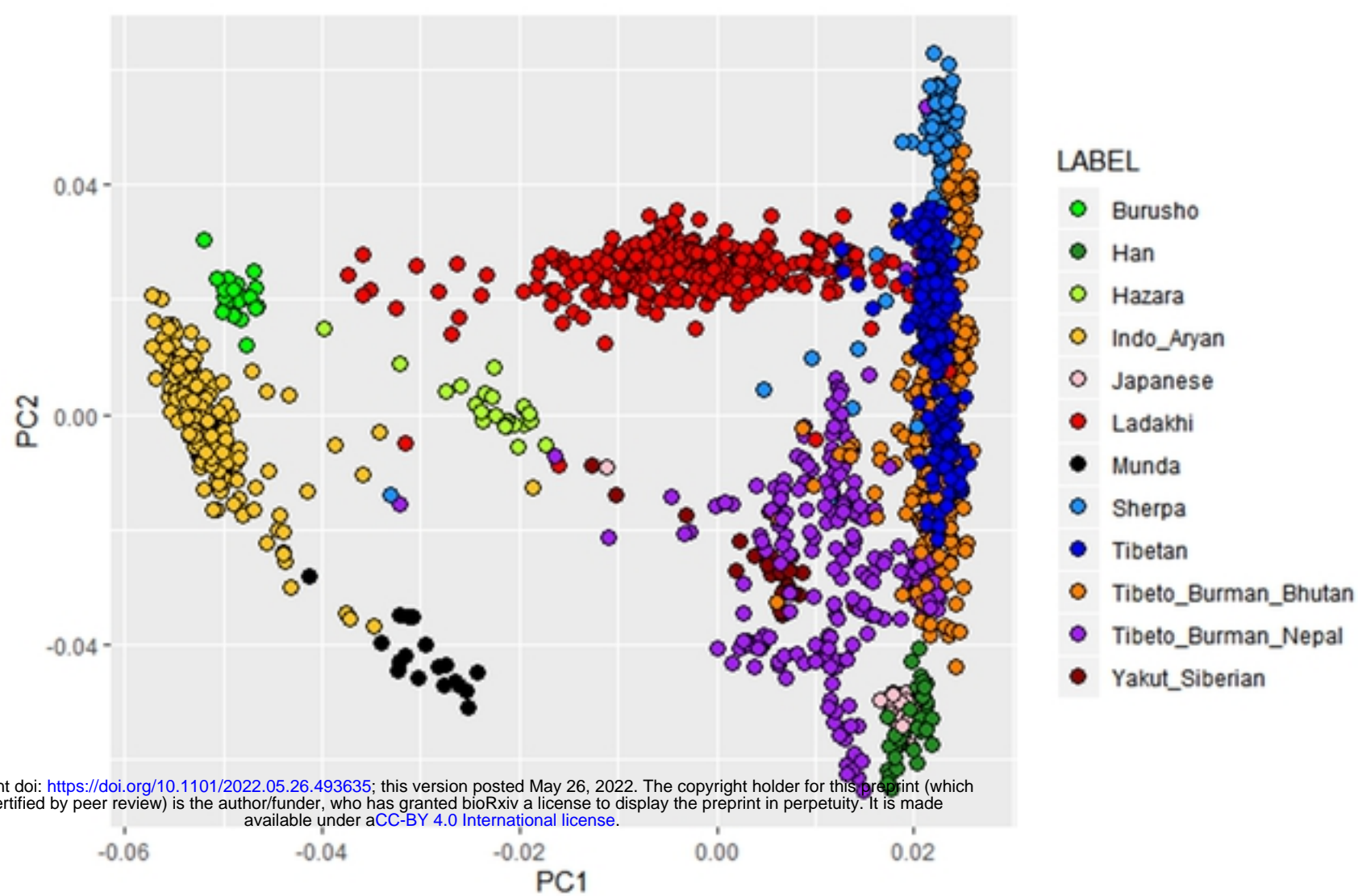


Fig 1

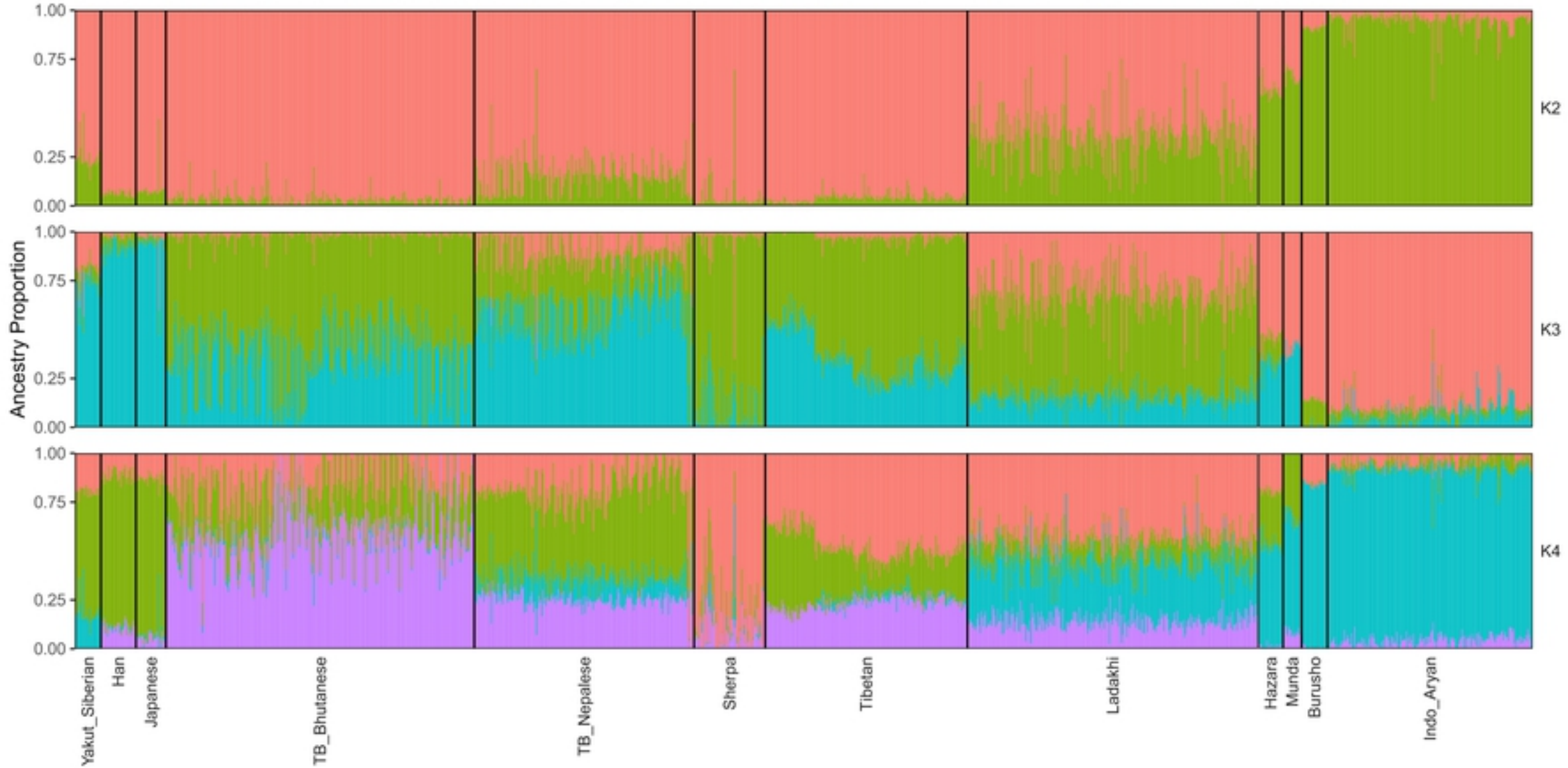


Fig 2

$f(\text{Ladakh}, X, Y)$

X-Y Population Pair

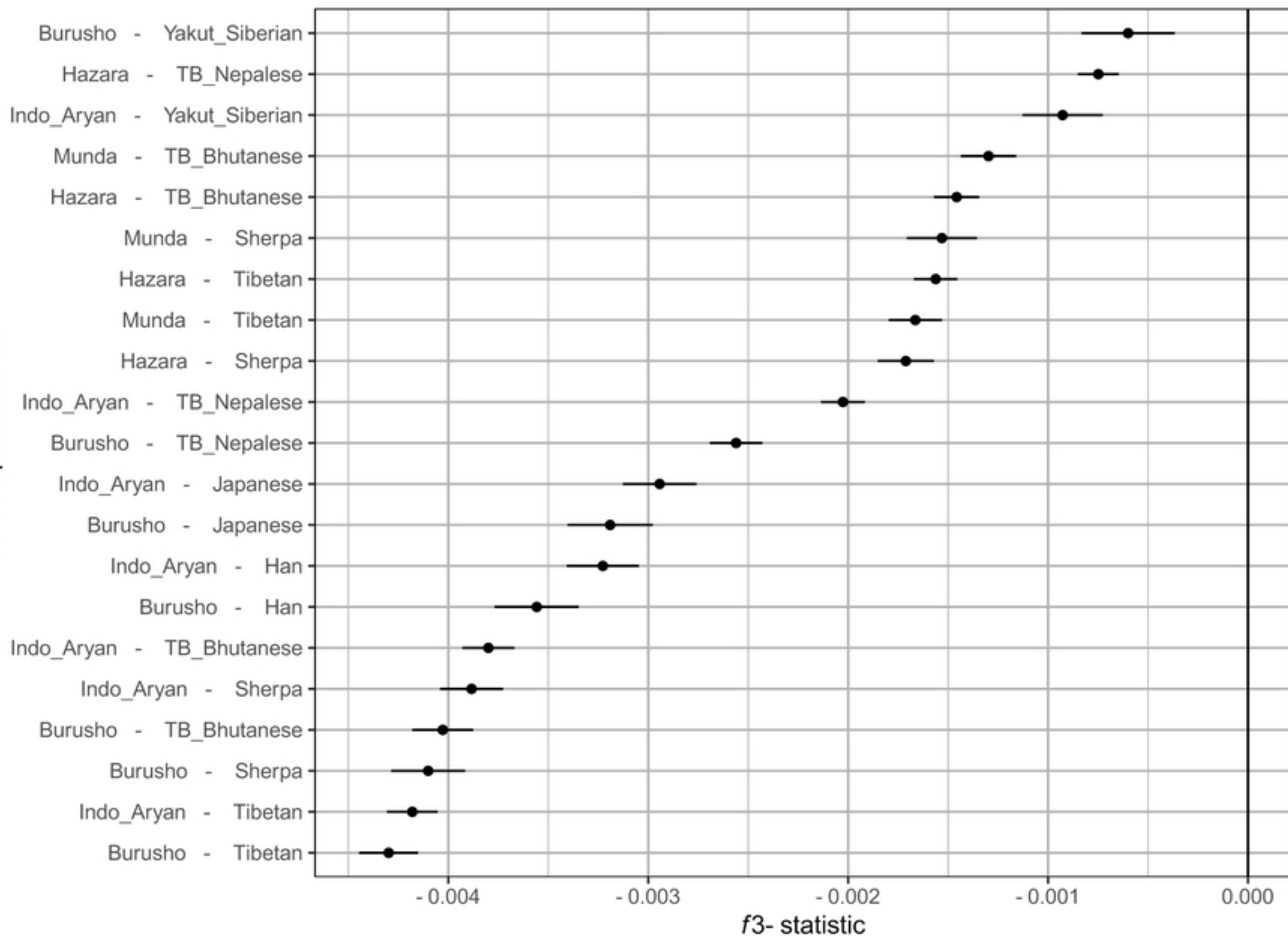


Fig 3