密级:...................

中国科学院大学
University of Chinese Academy of Sciences

博士学位论文

喜马拉雅地区夏尔巴人群的群体历史和高原适应的遗传学研究

作者姓名: **Sushil Bhandari**

指导教师: Professor Su Bing, State Key Lab of Genetic

Resources and Evolution, Kunming Institute of Zoology,

Chinese Academy of Sciences, China

学位类别: Doctor of Natural Sciences

学科专业: Genetics

研究所: Kunming Institute of Zoology,

Chinese Academy of Sciences, China

October, 2015

Genetic analysis of population history and high-altitude adaptation of Sherpas living in the Himalayan region

By

Sushil Bhandari

A dissertation submitted to

Graduate University of Chinese Academy of Sciences

In partial fulfillment of the requirement

For the degree of

Doctor of Natural Science

Kunming Institute of Zoology, Chinese Academy of Sciences,

Kunming, Yunnan, China

October, 2015

摘要

夏尔巴人群生活在喜马拉雅山地区，被誉为是高空登山者，但夏尔巴人群在何时何地人起源并且他们如何能够耐受高海拔的低氧环境仍存在争议。在本研究中，我们收集了居住在尼泊尔和中国西藏自治区的 582 个夏尔巴人的 DNA 样本。

首先，我们研究了夏尔巴人群母系（线粒体 DNA）与父系（Y 染色体）的遗传多样性，发现他们与藏族原住民共享大多数父系与母系的世系。线粒体 DNA 上现存两个夏尔巴人群特有的亚单倍群（C4a3b1 和 A15c1）是其祖先在距今约 1500 年前从青藏高原穿越喜马拉雅山迁徙至尼泊尔后人群扩张的结果。我们的发现并不支持此前夏尔巴与汉族人群是藏族人群的祖先群体的理论，反而支持藏族人群是夏尔巴人群的祖先这一论点。

其次，我们对夏尔巴人群进行了部分基因（*EPAS1*、*EGLN1* 与 TED 区域）的基因型检测，这些基因此前曾报道与藏族人群的高原适应相关。我们对夏尔巴人群进行基因型检测的区域包括藏族人群富集的缺失（TED）以及 30 个单核苷酸多态位点（其中 28 个位于 *EPAS1* 基因区域，另有 2 个位于 *EGLN1* 基因的错义突变）。我们选择这 30 个单核苷酸多态位点与缺失多态位点是基于此前对藏族人群的研究，并且这些多态位点在汉族与藏族人群之间等位基因频率差异明显（F_{ST} 大于 0.45）。我们发现夏尔巴与藏族人群在这些位点有着相似的衍生型等位基因频率（大于 51%），但在世界其他人群中其衍生型频率却非常低（小于 10%）。同样，夏尔巴与藏族人群在 *EPAS1* 与 *EGLN1* 基因区域均共享明显区别于其他人群的特异单倍型。此外，基因型检测结果显示，在 582 份夏尔巴样本中携带藏族人群富集的缺失的频率为 94%，其中 73% 为缺失型纯合子（拥有 0 个拷贝），而 21% 为杂合子（拥有 1 个拷贝）。有趣的是，非缺失型的携带者（拥有 2 个拷贝）在世界其他居住于平原地区的人群中的频率非常高

（2792 个世界范围内的样本中约占 97%），但在藏族人群（大约 10%）与夏尔巴人群（大约 6%）中频率非常低。

最后，我们进行了夏尔巴人群的表型（血红蛋白水平与血氧饱和度水平）与 *EPAS1*, *EGLN1* 和 TED 区域基因型数据的关联研究。有趣的是，我们发现夏尔巴人群 *EPAS1* 基因区域内的几个单核苷酸多态位点（rs113305133、rs116611511 和 rs12467821）以及藏族人群富集的缺失与血红蛋白水平显著相关（p 值均小于 0.03）。我们观察到，夏尔巴人群中 *EPAS1* 与 *EGLN1* 基因区域内这些单核苷酸多态位点的衍生型携带者的血红蛋白浓度低于祖先型的携带者。同样，夏尔巴人群中携带 TED 的个体比非缺失型的携带者的血红蛋白浓度更低。

综上所述，我们的研究表明 *EPAS1*、*EGLN1* 和 TED 可能在人类适应青藏高原的过程中经历了正选择。假设夏尔巴人群是近期从藏族人群衍生而来，他们从定居西藏近三万年的藏族祖先那里继承了这些适应特性。因此，这两个距今约一千五百年在夏尔巴人群中形成的线粒体 DNA 单倍型（C4a3b1 和 A15c1）并没有任何适应高原的特性。恰恰相反，夏尔巴人群和藏族人群形成了相似的适应高海拔的遗传与表型特征。

关键词: 夏尔巴人, 喜马拉雅山脉, 线粒体 DNA, Y 染色体 mtDNA, EPAS1, EGLN1, TED, 遗传多样性, 人类迁移, 高海拔适

Abstract

Sherpas living around the Himalayas are renowned as high-altitude mountain climbers but when and where Sherpa people originated and how they are capable of staying at less oxygen environment of high altitude remains contentious. In this study, we collected DNA samples from 582 Sherpas living in Nepal and Tibet Autonomous Region of China.

Firstly, we analyzed the genetic diversity of both maternal (mitochondrial DNA) and paternal (Y chromosome) lineages in Sherpa and found they share most of their paternal and maternal lineages with indigenous Tibetans. There exist two Sherpa-specific mtDNA sub-haplogroups (C4a3b1 and A15c1) as a result of population expansion after ancestors of Sherpa migrated from Tibetan plateau to Nepal crossing the Himalayas around 1,500 years ago. In addition, our finding rejects the previous theory that Sherpa and Han Chinese served as dual ancestral populations of Tibetans, and conversely suggest that Tibetans are the ancestral populations of the Sherpas.

Secondly, after knowing Sherpa as a recently derived population from Tibetans, we genotyped the key genes (EPAS1, EGLN1 and TED) in Sherpa that were reported previously in Tibetans for high altitude adaptation, including the Tibetan enriched deletion (TED) region and 30 single nucleotide polymorphisms (SNPs)(28 EPAS1 SNPs and two EGLN1 missense mutations) in Sherpas. These 30 SNPs and deletion polymorphism were selected based on previous studies of Tibetans which showed deep allelic divergence ($F_{ST}>0.45$) between Tibetans and Han Chinese. We found similar derived allele frequencies (>51%) among Sherpa and Tibetans, but much less derived allele frequencies (<10%) in other world populations. Similarly, there is a distinct haplotype pattern shared among Sherpa and Tibetans for EPAS1 and EGLN1, and these haplotypes are different from those of other populations. In addition, genotyping results of TED in 582 Sherpa samples showed 94%TED in Sherpa samples, including 73% homozygous deletion (Zero copy) and 21% heterozygous

deletion (one copy) in Sherpa. In contrast, the non-deletion carriers (two copy samples) were seen at high frequency (around 97% of total 2,792 worldwide samples) in other lowlander populations which are rare in Tibetans(<10%) and Sherpa (<6%).

Lastly, we conducted association studies between phenotypic (hemoglobin level and degree of blood oxygen saturation level) and genotyping data of EPAS1, EGLN1 and TED in Sherpa. Interestingly, we found several EPAS1 SNPs (rs113305133G, rs116611511G and rs12467821C) and TED having significant association (p value<0.03) with hemoglobin in Sherpa. We observed that Sherpa having derived allele of EPAS1 and EGLN1 SNPs had lower hemoglobin concentration than others carrying the ancestral allele. Likewise, Sherpa containing TED showed lower hemoglobin concentration in comparison with other non-deletion carriers.

In summary, our studies suggest that EPAS1, EGLN1 and TED likely underwent positive selection for human adaptation in the Tibetan plateau. In case of Sherpa, who were recently derived from Tibetans, inherited these adaptive traits from their Tibetan ancestors who have been living in Tibet since about 30,000 years ago. Hence, the two population-specific mtDNA haplotypes (C4a3b1 and A15c1) in Sherpa may not have any pronounced effect for high altitude adaptation. Instead, Sherpa and Tibetans share similar genetic and phenotypic traits for high altitude adaptation.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES:

1.   INTRODUCTION

The Himalayan mountain range is located in between the north of Tibetan plateau and the south of the Indian subcontinent. This region served as a corridor for human migrations through the ancient Silk Road in different regions of Asia, the Middle East and Europe. Nepal lying in the southern slopes of Himalayas is an inland country bordered by India and China. In the northeast of Nepal nearby Himalayas located Khumbu region which is the homeland of Sherpa people. Sherpa became world-renowned Himalayan population after Tenzing Norgay's Sherpa and Sir Edmund Hillary made their first ascent to Mount Everest, the highest peak of the world in 1953. Most of the Sherpa people are involved in mountaineering field as climbers, porters and trekking guides and have displayed extraordinary adaptive behavior at high altitude. The oxygen level in Sherpa homeland is 40% lower than that at sea level (Beall, 2007). This harsh environmental condition is overcome by several distinctive traits seen in native Sherpa which are lower ventilatory response (Lahiri, 1967), larger spirometric values (Havryk, 2002), lower hemoglobin concentrations(Adams & Shresta, 1974; C. Beall & Reichsman, 1984), higher arterial oxygen saturation(Hackett, Reeves, Reeves, Grover, & Rennie, 1980; Keyl et al., 2000), higher affinity of blood for oxygen (Morpurgo et al., 1976),higher heart rate(Pugh, 1962), less psycho-neurological symptoms (Garrido et al., 1996) and higher work economy (Bastien, 2005). These features show Sherpa seems well adapted at Himalayas and ideal population for studying high-altitude adaptation. But there is a limited genetic study in Sherpa, which is insufficient to give a systematic evaluation of the genetic basis of Sherpa's superior climbing ability. Here, we study the high altitude adaptation   mechanism in Sherpa by analyzing 582 DNA samples. Besides it is still contentious   when   and   where   the   Sherpa   people   originated

and settled at the high altitude regions of Nepal. Thus, our present studies give new insights about origin of Sherpa population and their possible genetic adaptation mechanism in Himalayan region clarifying the previous arguments.

The current dissertation manifests upon the possible genetic mechanism for Sherpa high altitude adaptation and their origin in Nepal. The literature review provides an overview of peopling in Nepal，high altitude adaptations in Tibetans, Andeans, Ethiopians and several physiological studies in Sherpa conducted before. It begins with fundamental knowledge about common haploid DNA markers widely used in phylogeographic studies i.e. mtDNA and Y-chromosome. In the subsequent chapter's linguistics, archeology, paleoanthropological evidence, population history, positive selection, high altitude adaptation in different region of world and physiological behavior in Sherpa is reviewed briefly. The existing controversial opinions about origin of Sherpa and their high altitude adaptation are discussed in the next chapter. Finally, the results of present study are summarized clarifying the previous contentious studies in Sherpa.

1. LITERATURE SYNOPSIS

2.1 The cause of using mtDNA and Y-chromosome markers for phylogenetic studies

2.1.1. Structure and Importance of mtDNA

The nuclear genome and mitochondrial genome are the two independent components of the human genome. Nuclear genome consists of 99.9995% of total genetic constituents and it comprises 24 chromosomes (22 autosomes and two sex chromosomes, X and Y). Mitochondrial genome is a circular DNA molecule of 16,569 nucleotides and includes the remaining minor genetic composition. Every human cell has mitochondria except for mature RBC which lack nucleus. Multiple copies of mitochondrial genome are located in the energy-generating organelles called mitochondria. Mitochondria produce ATP through oxidative

phosphorylation and are called as power house of cell. The mitochondrial DNA (mtDNA) was first recorded by electron microscopy in 1963 by Nass and Nass (Nass & Nass, 1963). In 1981 Anderson et al. first published the sequence and organization of the human mitochondrial genome. The human mtDNA is a circular double stranded molecule of 16,569 bases (Anderson et al., 1981; Andrews et al., 1999). The mitochondrial genes include 13 protein subunits of the enzymes involved in oxidative phosphorylation, two rRNAs of the mitochondrial ribosome, and 22 tRNAs necessary for the translation of the proteins encoded (Boore, 1999). MtDNA has two strands, a guanine rich heavy (H) strand and a cytosine rich light (L) strand. The heavy strands contains 12 of the 13 poly-peptide encoding genes, 14 of the 22 tRNA encoding genes and both rRNA encoding genes(Anderson et al., 1981; Andrews et al., 1999).The substitution rate of mtDNA is almost 10 times higher than nuclear DNA(Brown, George, & Wilson, 1979; Haag-Liautard et al., 2008) and comparatively higher in non-coding control region. In mitochondrial genome there is no proof reading mechanism so any mutation happened in mtDNA will pass to next generations without any repair or recombination. Thus mtDNA has an ability to accumulate large number of mutations. There are many copies of mtDNA in each cell. All copies of mtDNA are genetically identical in each cell and this genetic condition is known as homoplasmy. In homoplasmy, the hundreds to thousands of mtDNA copies within a cell or an individual have the same nucleotide-sequence (Birky, Demko, Perlman, & Strausberg, 1978) whereas in heteroplasmy, there can occur variations in mtDNA among different cells of same individuals.

During fertilization, a sperm cell contributes its nuclear genome but not mitochondrial genome. Since sperm does not have mitochondria. So, the zygote contains the mitochondrial genome originally found in oocyte. Thus it shows mitochondrial genome is maternally inherited (Giles, Blanc, Cann, & Wallace, 1980; Stoneking & Soodyall, 1996) and doesn't undergo any recombination (Breton, Beaupre, Stewart, Hoeh, & Blier, 2007; Macaulay,

Richards, & Sykes, 1999; Merriwether et al., 1991; Neiman & Taylor, 2009; Olivo, Van de Walle, Laipis, & Hauswirth, 1983; White, Wolff, Pierson, & Gemmell, 2008). Such specific mode of inheritance makes it a unique tool for studying human origin and migration (Cann & Wilson, 1983; Cann, 1994, 2001; Kivisild, Bamshad, et al., 1999; Kivisild, Kaldma, et al., 1999; Kivisild et al., 2002; Walker, Smith, & Smith, 1987). Thus, in every generation, we only have one mitochondrial ancestor whereas in nuclear DNA the number of ancestors increases by a factor of 2n, where n is the generation number. This peculiar character of mtDNA compel us to think that all modern human today had a single common mitochondrial ancestor, probably female at some point in the past which is also regarded as "Mitochondrial Eve".

In addition, there are two small regions in human mtDNA known as the hypervariable regions I and II (HVI and HVII). Mutation rates in HVI and HVII are especially high on average and there is evidence that the rates vary within the regions as well (Jazin et al., 1998). The high number of sequence variations in the non-coding control region can make differentiation among populations (Greenberg, Newbold, & Sugino, 1983; Lutz, Weisser, Heizmann, & Pollak, 1998; Wilson, Stoneking, Holland, DiZinno, & Budowle, 1993). There is displacement loop (D-loop) region in mtDNA because the H-strand replication often pauses a few hundred base pairs after its initiation, resulting in a structure consisting of the nascent H-Strand associated with its template and displaces third single strand (Hans-Jürgen Bandelt, Richards, & Macaulay, 2006).The D-loop region is rapidly evolving in human mtDNA with accumulation of base substitutions, insertions or deletions which is faster than single copy nuclear DNA. The rate of substitutions in human D-loop region is estimated in range between 2.8 (Cann, Brown, & Wilson, 1984) to 5 (Aquadro & Greenberg, 1983) times higher than the rate of other parts of the mtDNA. The variations in different mtDNA sequences occurred mostly due to mutational events rather than recombinational

rearrangements. The different mutations accumulate sequentially in mtDNA as time passes and form independent lineages or haplotypes.

2.1.2. Dating using mtDNA sequence

The process of molecular dating of the most recent common ancestor (TMRCA) based on set of DNA sequences, either taken from within or between species, is very important interpretations made from genetic data for estimating divergence. Previously mtDNA mutations rate were calibrated just based on short fragments of mtDNA, i.e. HVS-I and HVS-II. The HVS-I mutation rate of $1.79*10^{-7}$ is commonly used for calculating different lineage age (Forster, Harding, Torroni, & Bandelt, 1996). But the age can also be estimated from complete mtDNA sequence data. The mutation rate is not the same throughout the whole mtDNA genome and the control region of mtDNA has more than five times higher sequence variation than the coding region (Endicott, Ho, Metspalu, & Stringer, 2009; Ingman, Kaessmann, PaÈaÈbo, & Gyllensten, 2000).The advances in sequencing technology, has made available large number of mtDNA complete genome sequences. These complete sequence data sets have significantly improved the molecular resolution of phylogenetic studies (Palanichamy et al., 2004; Kong et al., 2011; Olivieri et al., 2006; Van Oven & Kayser, 2009).

The commonly used mutation rates for human mtDNA complete sequences are based on interspecies calibrations assuming certain split times of the humans-chimpanzees clade. Several studies have assumed 6.5 million years old coalescent time of human and chimp mtDNA lineages (Kivisild et al., 2006; Mishmar et al., 2003). But Soares et al. 2009 used a more ancient human-chimp split time in the calibration of mtDNA mutation rate and as a consequence the inferred mutation rate was relatively slower (Soares et al., 2009). Different software and mathematical formula can now easily calculate the age of any mtDNA lineages.

2.1.3. Structure and importance of Y-Chromosome

The Y- chromosome is 60 megabases (Mb) in size and it is male specific having a small number of genes than other chromosome (Jobling & Tyler-Smith, 1995). The Y-chromosome have a short (Yp) and a long (Yq) arm. It contains 27 genes distributed 9 on short arm and 18 on long arm (Skaletsky et al., 2003). In majority of (about 95%) the Y-chromosome there is no X-Y crossing over in male meiosis and it is called the non-recombinant region of Y (NRY), or the male specific region (MSY). NRY of the Y chromosome splits into two large parts: heterochromatic and euchromatic portions. Mostly the gene transposed to the NRY region of the Y-chromosome are at higher risk of degeneration in their successive generations. However, recent studies shows human Y chromosome has not lost any genes since the divergence of humans and chimpanzee between 6-7 million years before present (YBP) (Hughes et al., 2010; Rozen et al., 2003). The sex chromosomes are thought to evolve from a pair of autosomes within the last 300 million YBP (Hughes et al., 2010; Jobling & Tyler-Smith, 2003; Ohno, 2013; Skaletsky et al., 2003).

The Y chromosome determines sex and is inherited clonally from father to son. It is the only chromosome having haploid characteristics and no material is exchanged with a homologue through recombination making all sites linked to each other. The 95% length of the human Y-chromosome is MSY whereas the 5% is genetically similar to X chromosome called as pseudo autosomal region (PAR). PAR is located in the telomere ends of the Y chromosome and recombines with the X chromosome during meiosis. But mainly three types of polymorphisms: indels, SNPs and microsatellites/STR are used for studying the Y chromosomal phylogeography of world (Hammer et al., 1998; Jobling & Tyler-Smith, 2003; Underhill et al., 2001; Underhill et al., 2000). Indels are insertions or deletions at particular locations on the chromosome, e.g. YAP (Y-chromosome Alu Polymorphism) (Hammer et al., 1998). SNPs are single nucleotide polymorphisms in which a particular locus is altered. STRs

are the short sequences of nucleotides (mainly tri or tetra nucleotide), which are repeated over several times in tandem. There is absence of recombination in indels, SNPs and microsatellites of Y chromosome so they are linked to each other forming a haplotype. Such haplotypes of the Y-Chromosome are the major attractions to study genetic differentiations of human populations. The extant distribution of Y-chromosomal haplotype diversity is used as a tool for reconstructing the peopling across the world by modern humans, from the paternal perspective (Jobling & Tyler-Smith, 2003; Su, Jin, et al., 2000; Su et al., 1999; Underhill & Kivisild, 2007; Underhill et al., 2001). Since Y chromosome SNPs have relatively low mutation rates, non- recombinant nature and paternal transmission, they have great importance in evolutionary studies. An evolutionary force changes the pattern of Y chromosome by changing repeat content, mutation rate, gene content and haplotype structure. The non-recombinant region (NRY) of the human Y-chromosome conserves compound haplotype information over time scale spanning prehistory of modern humans (Underhill & Kivisild, 2007).The Y-chromosome SNPs have low mutation rates and are also called as unique event polymorphisms (UEP).These markers usually show regional distributions, tracing back to the origin and thus are the best markers to measure male gene flow. Thus NRY of Y-chromosome contains record of mutational events that have occurred along the paternal lineages throughout the evolution.

There is different mechanism of transmission of autosomes and mtDNA/Y chromosome to next generation. The autosomal DNA gets reshuffled by recombination and other random mixtures. However, mtDNA and Y-chromosome remain intact while passing from one generation to next. The mtDNA is transmitted from mother to child and help for tracing maternal lineages in both men and women. The Y- chromosome (i.e. MSY) is passed from father to son without recombination with other DNA. Thus, Y chromosome marker is widely used to trace paternal lineages. X-chromosome and autosomes both have multiple

ancestors whereas mtDNA and MSY region of Y chromosome have small fraction of ancestors coming from maternal and paternal line. So, Y-chromosome and mtDNA are widely used markers for analyzing population history (Underhill & Kivisild, 2007).

2.2   Peopling of Nepal

Nepal is bordered by China on the North and India on the East, West and South. Nepal can be divided into three major parts geographically: Mountains (High Land), Hills (Middle land) and Terai (Lowland). The mountain region is located along the southern slopes of the Himalayas. Terai is the lowland region of Nepal which comprises the northern edge of the Gangetic plain that boarder North India. In between these two (high land and lowland) region lie the intermediate hills and valleys where majority of the Nepalese population resides. The peopling of Nepal can be studied by knowing its languages, archeological studies and population genetic structure.

2.2.1 Major language groups

In tracing the origin of any population, studying of language, culture and genes are very important. The spread of language or any other cultural element can occur together with total replacement of genes (demic diffusion) or without any exchange of genetic material between the human groups that interact with each other (cultural diffusion). Nepal is linguistically diverse country having around 126 languages (Ethnologue (2005).The different languages of Nepal can be classified into two major language family- Indo-European and Tibeto-Burman. Nepal is the meeting place for these two language family. The Indo-European branch of linguistic family are dominant in east and west and the Tibeto-Burman language family is prevalent in north and east of Nepal. The Austroasiatic and Dravidian language speakers are also seen in Nepal but in minor number. In addition, there is also Kusunda, an isolate language speaking population in Nepal.

In Nepalese context, Indo-European family of languages mainly comprise Indo-Aryan group of languages, which forms the largest group of languages in terms of speakers, viz. nearly 70.81% .Nepali is an official language of Nepal and which belongs to Indo-European language family. Indo- European languages range geographically from Srilanka (Colombo - Sinhalese) in the southeast to Iceland (Reykjavik) in the northwest. In Nepal 14 languages belong to the Indo-European language family (Kansakar, 1996). There are around 70.81% speakers of this family. Nepali language being national language is more dominant and has 44.64% speakers spreading all different region of the country. Indo-European languages are supposed to evolve over the past 8,000 years. The Indo-European languages are a group of more than 400 languages that contains different languages from Polish and French to Icelandic and Hindi, and scientists have claimed that they all originally came from one single language spoken in the region of Anatolia, in present-day Turkey (Bouckaert et al., 2012). The Indo-European languages spread from Anatolia through farming to various parts of Europe and Asia, changing in different places as they migrate and give rise to different language with diverse dialects.

Tibeto-Burman was first recognized as a language family in 1823 in Paris by the well-travelled and well-read German scholar Julius von Klaproth. He showed that Tibetan, Burmese and Chinese belonged to a single family of languages, whereas languages such as Thai, Mon and Vietnamese each belonged to other separate families. The ancestors of Tibeto-Burman population originated from the basin of Yellow river (Su et al., 2000; Su et al., 1999) and dispersed in other parts of Tibet, Nepal, Northeast India, Bhutan and Burma. In Nepal, 57 languages belong to the Tibeto-Burman subfamily. The Tibeto-Burman speakers are seen mostly in hills and mountains of Nepal.

Austroasiatic languages are also spoken by some tribal groups in the eastern Nepal. There is only one language named Santhali which belongs to Austric branch of the Austro-

Asiatic family (Kansakar, 1996). Their introduction to Nepal, i.e., how and when they happen to be in Nepal is unknown. However, presence has been consistent and reported in all censuses.

Dravidian languages were first recognized as separate language family in year 1816 by Francis Ellis. The speakers of the Dravidian languages are settlers in the eastern Nepal. Dhangar is the language spoken by Dravidian speakers in Nepal. The Dhangar language is also spoken in India but there is different dialect among Dhangar speakers of Nepal and India. It is said to be a regional variant of Kurux spoken in Jharkhand State of India though it shows divergence from its vocabulary and grammar (Gordon, 1976; Yadava, 2002).The proto Dravidian speakers are suggested to migrate from Near East to Indus Valley and Indian subcontinent with the farming dispersal (Cavalli-Sforza, 1995;Cavalli-Sforza, 1996; Renfrew, 1996).The Eastward migration of Dravidians from South India towards East of India and Nepal are proposed previously (Sengupta et al., 2006).

Beside the above four language, Nepal also has speakers of isolates language-Kusunda. The term isolates language means the language which is unrelated to other surroundings languages. A linguistics study suggests their link with Ocenian as well as Andaman Islanders (Whitehouse, Usher, Ruhlen, & Wang, 2004). However, currently there is not enough genetic information available to support this putative link. Some linguistic similarity between Kusunda and Greater Yenisseian language was also pointed by linguist, Van Driem. He further predicts Kusunda might be the remnant of the ancient Greater Yenisseian migration into the Himalayas. The paternal lineage Q (M242) of Greater Yenisseian in Kusunda need yet to examined if their language and gene both are connected or they lost genetic connection (van Driem, 2013).

Figure 1. Linguistic map of Nepal showing locations of major languages
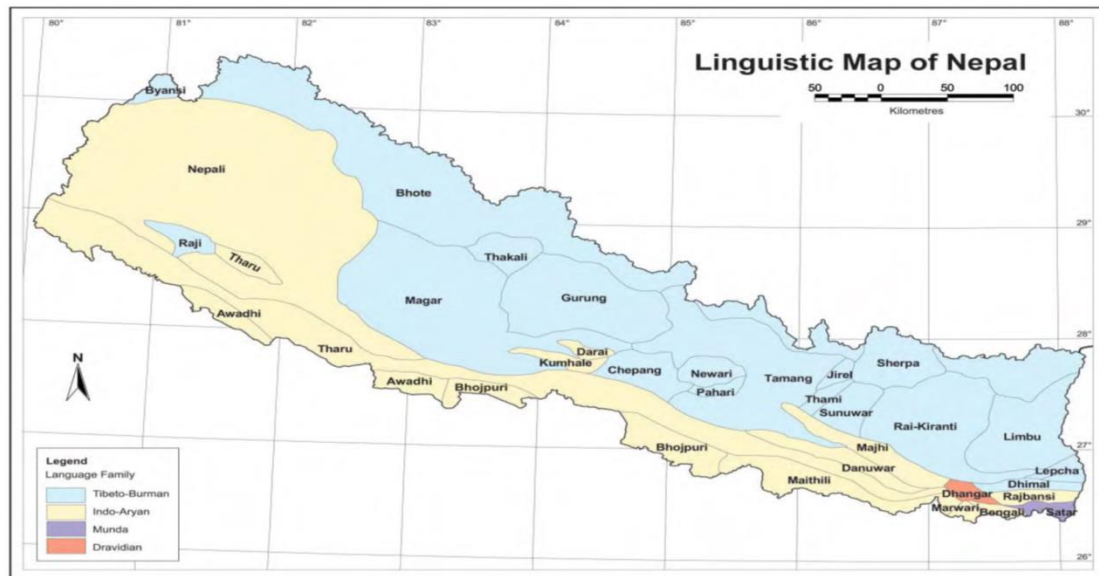
Figure 6: Linguistic map of Nepal showing the distribution of spoken languages
Source: Harka Gurung (2006). *Nepal Atlas and Statistics*. Himal Books, printed by permission

2.2.2 Archaeological and paleoanthropological evidence

In 1980, in the western Terai of Nepal, Nepal-USA Scientific Expedition team found tooth of Ramapithecus. They claim it might be upper left molar of Ramapithecus and was dated approximately 11 million years old on the basis of palaeomagnetic dating method. Ramapithecus are considered as ancestors of modern orangutans. Some other scientific explorations by Department of Archaeology of Nepal found some palaeolithic tools and stone tools in Nawalparasi district. They also found Neolithic tools from the Dang valley.

The ancient historical evidences are also identified in Lumbini Zone relating to birthplace of Gautam Buddha in Nepal. These shows they are associated with Pre-Buddha and Buddha era.

Professor Mark S. Aldenderfer, anthropologist and archaeologist from University of California, has done anthropological studies in caves of Upper Mustang region of Nepal for knowing its settlement history. They search for ancient human samples in the caves for understanding the people migration pattern and to know from where they came from. The

11

Mustang caves are located about 13,800 feet above sea level. Dr. Mark and his team discovered Buddhist cave art, possible meditation chambers for monks, and skeletons dating back to 200 to 700 AD in cave. They have collected bones of 34 individuals from the caves located in cliff of mountain. Their preliminary report highlighted only the cultural practices for handling dead body. The team compared their findings in Kathmandu with bones and remains that were found earlier in the southern Mustang region by a Nepali-German team. Jacqueline T. Eng analyzed all the bone and found nearly 67 percent of all the bones had been de-fleshed, most likely with a metal knife, which the team believe is evidence that these people observed a funerary tradition that is similar to one known as a Tibetan "sky burial," where the deceased's body is cut up and left on a mountaintop, exposing the remains to be eaten by predatory animals and birds. They further propose that different ancient trade might happen passing by Mustang. They collected lots of beads belonging to different areas such as Pakistan, India, and Iran, and they also found iron daggers.

Another prehistoric study was done in Nepal by Prof. Dr. Freund and Dr. Gudrun Corvinus Institute for prehistory, Erlangen University, Germany. Their team focuses in search of ancient tools and settlement time in Nepal. They predict the Neolithic settlement might have occurred in Kathmandu and Dang Valley. They found affinities to microlithic culture by finding variety of small flakes of chert, quartz and quartzite. They didn't define the possible time when these stone tools were used. The different stone tools and fossils of rare mammals as well as sea-animals found in different regions of Nepal indicate there might be ancient human presence in Nepal after the African exodus.

2.2.3 Population genetic studies in Nepalese populations

Tharu-Indo European Speakers

Tharus are staying in Terai -the lowland region of southern Nepal. The Terai region was highly malarial area so less population was staying until a few decades ago. But Tharu become resistant to malaria and are capable of staying in the malarial Terai region.

Different hypotheses on Tharu are postulated telling Tharu shared ancestry with Austroasiatic, or Tibeto-Burman-speaking populations as well as aboriginal roots in the Terai. Tharu speak Indo-Aryan languages and their genetic ancestry is composed of mainly two origin (East Asian and South Asian) (Chaubey et al., 2014). Chaubey et al studied Tharu from Nepal and India, and found mitochondrial DNA haplogroup M43 (Thangaraj et al., 2008) and Y-chromosomal haplogroup: O3a2c1-M134 is shared among different Tharu group. Thus, Tharu demographic history shows there is gene flow from Nepal to India and vice versa. Tharu genetic composition can be classified into mainly three different components East Asian, West Eurasian and Indian components (Fornarino et al., 2009).

Tharus may have admixed with Tibeto-Burman speakers due to which East Asian signature seen on them. The East Asian mtDNAs haplogroups B5a, D4, G2a are present in Tharus. Similarly, the East Asian Y chromosome haplogroups seen in Tharus are C(xC5), D, N, O3, Q, and K*. Among these Hg O3(O3-M117) is highly prevalent in Tharus (39.8%). Tharus also contain Tibetan markers, haplogroups D (4.5%) and Q (0.7%). The presence of haplogroups B5a and F1 in the Tharu of Eastern Uttar Pradesh (near the Indo-Nepal border) points to past human movements from East Asia to India, perhaps through Nepal (Thangaraj et al., 2008).

West Eurasian markers are absent in the mtDNA of Tharus, whereas they are present in their Y chromosomes lineages as J2-M410* and J2-M241*, with a frequency of 30%. These could be due to recent gene flow from the Middle East or, remnant of out of Africa

migration. J-M410, which was associated with the first farmer dispersal in Europe and therefore are compatible with a dispersal of this lineage from somewhere in the Middle East/Asia Minor.

A great majority of the Tharu mtDNA and Y-chromosome gene pools is represented by lineages shared or derived from Indian subcontinent haplogroups. In particular, Tharus share with Indians ancient mtDNA haplogroups (for example, the M clades M31, M33, M35, M38, the new M52 and also the R30, almost all dated ~30 kilo years ago) and Y-chromosome haplogroups (such as H-M69, O2-P31Tdel, R1-M17* and R2-M124) There exist deep shared ancestry between Tharus and Indians. Some Y-chromosome lineages in Tharu are shared even with some tribal of South Indian. Therefore, Tharu population represents their genetic affinity with both North and South Indian.

Tibeto-Burman speaking populations of Nepal

Another study led by Peter de Knijff of Leiden University on different Tibeto Burman(TB) speaking populations of Nepal and Bhutan try to find the uniqueness and antiquity of the Himalayan populations (Kraaijenbrink et al., 2009).They try to link between language and genes by comparing 15 autosomal STRs loci among the speakers of Tibeto Burman language family and observe genetic similarity between speakers of same language family.

The Y-chromosome haplogroup and Y−short tandem repeat (STR) haplotypes were investigated for knowing the genetic affinities of three populations from Nepal-including Newars, Tamang, and population of Kathmandu. These populations were studied for knowing the role of Himalayas in peopling of Nepal. They found East Asian haplogroup in Nepal indicating Himalayas have been permeable for recent gene flow from Tibetan plateau to Nepal. Newars and Kathmandu population contain both East Asian and Indian component. In Tamang they found higher amount of East Asian component but much less Indian genetic

14

component (Gayden et al., 2011). Tibetans contain very less Indian component suggesting Himalaya act as unidirectional barrier for gene flow from Indian subcontinent to Tibetan plateau but permeable for other directions (Gayden et al., 2007). The genetic similarities between Tibeto-Burman (TB) speaking population of Nepal and Tibetans shows gene flow from Tibetan plateau to Nepal and not from Northeast India (Gayden et al., 2009; Wang et al., 2012). Since Nepal lies in Himalayan region, peopling in Nepal at mountainous region occurred through gene flow by crossing Himalaya's barrier recently.

Other populations- Brahmain, Magar, Chhetri, Newar and Madhesi

The Nepalese populations are highly diverse, with some populations showing a closer relationship to East Asian populations, while others are closer to South Asian populations. An examination of the ethnicity of the Nepalese individuals reveals that individuals from the ethnic groups derived from the caste system, including Madhesi, Brahman and Chhetri, show a closer relationship to South Asian populations (especially Indian Brahmins). Individuals from the two indigenous Nepal ethnic groups (Newar and Magar) are closer to Central/East Asian populations (Xing, 2010). Majority of Nepalese population is highly admixed. The Nepalese also share common ancestral component found in Indians and Pakistanis. However, Nepalese are highly heterogeneous and substantial ancestry components occurred from Central Asia, East Asia and Southeast Asia (Xing, 2010).

Kusundas

Kusunda language is a member of the Indo-Pacific family (Whitehouse et al., 2004). The Indo-Pacific family is located on New Guinea and surrounding islands. Kusunda is an isolate language in Nepal. There is limited genetic data on Kusunda to compare their genetic relatedness with other populations. However, some studies found Kusunda cluster together with Southeast Asian (Rasmussen et al., 2011). Previous linguistic studies proposed Kusunda can be the remnant of the migration that led to the initial peopling of New Guinea and

Australia. Kusunda can be ancient relic populations of South Asia but further more genetic studies is needed to confirm it.

1.3   High Altitude Adaptation

The original concept of high altitude adaptation is guided by Charles Darwin, below is the quote of Charles Darwin written in his book "On the Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life," 1859, p. 162.. "*Variation is a feature of natural populations and every population produces more progeny than its environment can manage. The consequences of this overproduction are that those individuals with the best genetic fitness for the environment will produce offspring that can more successfully compete in that environment. Thus the subsequent generation will have a higher representation of these offspring and the population will have evolved*." Thus, the Darwin theory of natural selection also known as "survival of the fittest," is the process where individuals with favorable traits survive in changed environmental conditions because of those traits whereas other who doesn't have that trait fails to compete. These favorable traits are inherited and seen in next generation leading to change in populations over successive generations by process defined as "descent with modification" by Darwin.

After modern human originated in Africa nearly 200 thousand years ago (kya) their rapid migration occurs to other part of the world and encounter diverse environmental conditions (Henn, Cavalli-Sforza, & Feldman, 2012; Henn et al., 2011; Oppenheimer, 2012; Laura B Scheinfeldt, Soi, & Tishkoff, 2010). The tolerance of these different environmental stresses (Ultraviolet rays, temperature, Oxygen, diets) over long time creates local adaptations producing different phenotype. For example, there is variation in human skin color depending on amount of UV available in that region (Jablonski & Chaplin, 2000; Jablonski & Chaplin, 2010). Similarly, after modern human arrive to Himalayan plateau around 30,000 years before, their long term settlement produces adaptive features like lower

level of hemoglobin concentration in high altitude natives of Tibetans and Sherpa for adaptation(Wu, Liu, Cui, Qi, & Su, 2013). The recent genome wide studies have identified several candidate genes for high altitude adaptation in Tibetans. But majority of studies found EPAS1 and EGLN1 as the top two candidate genes having strong signal of selection.

2.3.1. Genes involved in Tibetans Adaptation

The different genes involved for Tibetan high altitude adaptations are listed below:

1) EPAS1-(Beall et al., 2010; Bigham et al., 2010; R.-L. Ge et al., 2012; Huerta-Sánchez et al., 2014; Jeong et al., 2014; Peng et al., 2011; Simonson, 2010; Xu et al., 2011; Yi et al., 2010)

2) EGLN1-(Bigham et al., 2010; Bigham et al., 2009; Jeong et al., 2014; Lorenzo et al., 2014; Peng et al., 2011; Simonson, 2010; Xiang et al., 2013a; Xu et al., 2011; Yi et al., 2010)

3) PPARA-(R.-L. Ge et al., 2012; Scheinfeldt et al., 2012; Simonson, 2010)

4) HMOX2-(Peng et al., 2011; Simonson, 2010; Wuren et al., 2014)

5) CYP17A1-(Simonson, 2010; Wuren et al., 2014)

6) PKLR-(Simonson, 2010; Wuren et al., 2014)

7) HBB and HBG2-(Simonson, 2010; Wuren et al., 2014)

8) HFE-(Wuren et al., 2014; Yi et al., 2010)

9) Hypoxia up regulated protein 1 (HYOU1)-(Chene, 2006; Jeong et al., 2014; Sanson et al., 2008)

10) Hydroxy methylbilane synthase (HMBS)-(Gubin & Miller, 2001; Jeong et al., 2014; Schadt et al., 2008; Zeller et al., 2010)

11) Others (gene in HIF pathway) (ARNT ,HIF3A,TCEB2,UBE2D1,VHL,CA9, EGLN2, EPO,RBX1, UBA52,UBE2D2, CREBBP, EGLN3, HIF1A, RPS27A, UBB, UBE2D3, CUL2, EP300, HIF1AN, TCEB1, UBC, VEGFA, CUL2, RPS27A, UBB,

UBE2D3, (EGLN2, HIF1A, TCEB1, UBC, VHL, EGLN3, HIF3A, TCEB2, UBE2D1, RBX1, UBA52, UBE2D2) (Jeong et al., 2014), Clorf124,DISC1, ATP6V1E2, SPP1, PKLR, C4orf7, PSME2, OR10X1, FAM9C, LRRC3B, KRTAP21-2, HIST1H2BE, TTLL3, HISTIH4B, ACVR1B, FXYD6, NAGLU, MDH1B, OR6Y1, OTX1, MBNL1, IFI27L1, C18orf55, RFX3, HBG2, FANCA, HISTIH3C, TMEM206)(Yi et al., 2010), (CYP2E1, EDNRA, ANGPTL4, CAMK2D, PTEN)(Simonson, 2010), TGFBR3, GCH1, PIK3R1) (Peng et al., 2011)

High altitude hypoxic adaptation seems the result of interaction among multiple genes. However, recent several studies highlighted mainly EPAS1, EGLN1, and TED as the three key genes (or regions) for Tibetans' adaptation. The endothelial Per-Arnt-Sim (PAS) domain protein (EPAS1) is a transcription factor induced under hypoxic conditions. Several studies have shown individuals carrying the derived alleles of EPAS1 intronic genetic variants have lower haemoglobin levels than individuals homozygous for the ancestral allele. EPAS1 is also known as hypoxia-inducible factor 2α (HIF-2α). There is not any functional variant in the coding region of EPAS1 gene. *Lou et al* (2015) proposed the copy number variations (CNVs) might also have possibilities for HAA directly or indirectly acting together with EPAS1 (Lou et al., 2015). Since several previous studies (Schlattl,2011; Weischenfeldt, Symmons, Spitz, & Korbel, 2013) found copy number variations also as having functionally significant in gene expression. *Lou et al 2015* found a 3.4-kb copy number deletion at 80 kb downstream of EPAS1 gene in 90% of Tibetans. This Tibetan-enriched deletion (TED) is seen in three different forms: homozygous deletion (0 copy), heterozygous deletion (1 copy) and non-deletion (2 copies). Tibetans carried 50% of homozygous deletion (0 copies) whereas non Tibetans didn't have homozygous deletion. There is only heterozygous deletion found at low frequency (3%) in other non-Tibetans population (Lou et al., 2015). The TED

was tightly linked with EPAS1 SNPs (Beall et al., 2010) having association with lowering hemoglobin concentrations in blood. It also have strong LD with the 5-SNP Denisovan introgressed motif (Huerta-Sánchez et al., 2014), but the TED itself was absent in Denisovan sequence.

EGLN1 gene encodes prolyl hydroxylase 2(PHD2) (Lorenzo et al., 2014). PHD2 activates the degradation of HIF 1 transcription complex. The adaptive variant of EGLN1 shows lower Km value for oxygen, suggesting that it promotes increased HIF degradation under hypoxic conditions (Lorenzo et al., 2014). In addition, variation at the EGLN1 locus is associated with protection against polycythemia in Tibetan highlanders and SNPs at the EPAS1 locus are associated with hemoglobin levels in Tibetans.

Genes involved in hypoxia inducible factor 1 pathway

HIFs are transcription factors that function as master regulators of oxygen homeostasis. They are composed of a hypoxia-inducible α subunit and a constitutively expressed β subunit. The α subunit has three different isoforms; HIF-1α, HIF-2α (encoded by EPAS1) and HIF-3α. The HIF-1α subunit is ubiquitously expressed, whereas HIF-2α and HIF-3α have tissue-specific patterns of expression. Mainly HIF-2α (HIF-2α and HIF-1 α both) are suggested to involve in stimulating erythroid progenitors and regulates the production of erythrocytes.

In the presence of oxygen, HIF α subunits are hydroxylated by PHD2, which generates a binding site for the von Hippel–Lindau (VHL) protein and results in their polyubiquitination and proteasomal degradation. Thus, PHD2 and VHL are the major oxygen-dependent negative regulators of HIFs. The oxygen atom of the hydroxyl group (OH) is derived from molecular oxygen. When oxygen level decreases, prolyl hydroxylation of HIF is reduced and finally suppressed PHD2 activity. Thus decrease in von Hippel–Lindau (VHL)

protein binding results in decrease of degradation of HIF. So there occurs transcription of hypoxia responsive gene.

The HIF 1 transcription complex is composed of two subunits: HIF-2 α (also known as EPAS1) and HIF-1β (also known as ARNT2).EGLN1 is involved in degradation of HIF 1 transcription complex. The HIF 1 transcription complex inhibits peroxisome proliferator-activated receptor- α (PPARA) expression. The PPARA encodes the nuclear peroxisome proliferator activated receptor alpha (PPARα) that regulates fatty acid metabolism and is in turn regulated by HIF (Simonson, 2010). There are several gene involved in down regulation of fatty acid oxidation, including PPARA, which has been observed in rats exposed to hypoxia (Kennedy, 2001).

Gene involved in metabolic changes of Tibetans for their high-altitude adaptation

The energy production from oxidative metabolism might be effected in case of low oxygen environment at high altitude. In this situation, if oxidative metabolism occurred reactive oxidative intermediates will accumulate in mitochondria, resulting in cell death. A previous study on the whole organism level has revealed that HIF plays important role in regulating metabolism, showing a relationship between HIF and metabolic demands in humans (Formenti et al., 2010). The HIF signaling will change the metabolism for decreasing tissue oxygen demand (Denko, 2008; Semenza, 1999). In limited oxygen level, the ATP production will be efficient from carbohydrate metabolism than fatty acid. Because fatty acid oxidation consumes more oxygen than glycolysis for energy production. There is limited use of fat metabolism in people living at high altitude are shown in previous studies (Roberts et al., 1996). Additionally, there occurs glycolysis then oxidative glucose metabolism for maintaining energy production (Brooks et al., 1991). It means there will be up-regulation of glucose uptake and glycolysis and down-regulation of mitochondrial glucose oxidation

(Papandreou, 2006). It shows natural selection on HIF regulated genes has maintained these metabolic changes for hypoxic adaptation at high altitude. Some studies shown that the putatively advantageous EPAS1 haplotype previously associated with hemoglobin (Beall et al., 2010; Yi et al., 2010) was also highly associated with changes in serum lactate concentration (Ge et al., 2012). In addition, the putatively adaptive PPARA haplotype also exhibit an association with serum free fatty acid concentration in Tibetan population (Ge et al., 2012).

The mice lacking EPAS1 have a syndrome of multiple-organ pathology, biochemical abnormalities and altered gene expression patterns. EPAS1 knockout mice have lactic acidosis, altered Krebs cycle function, hypoglycemia and dysregulated fatty acid oxidation (Scortegagna et al., 2003).

2.3.2. Positive selection for adaptation at the Tibetan plateau

The positive selection on genes can occur by three different models: classic selective sweep/hard sweep, selection on standing variation and selection on complex traits/polygenic adaptation (Pritchard & Coop, 2010; Scheinfeldt & Tishkoff, 2013). In classic selective sweep model (Smith & Haigh, 1974), a new advantageous mutation arises in a population and fixed rapidly where as in selection from standing variation (Daub et al., 2013), a variant that is already present in the population becomes advantageous in a new environment and increases frequency as time passes. In case of selection on a complex trait, there will be a set of adaptive variants lying in different chromosomes and these set of genetic variants will be more common in successive generations (Coop et al., 2009).Several previous studies have defined the positive selection on Tibetans might have occurred by hard sweep model. Whereas other claim it might have occurred through selection from standing variations (Jeong et al., 2014).They proposed for admixture mediated Tibetan adaptation.

Admixture is the process of exchanging genes due to interbreeding among different populations. *Jeong et al*., 2014 proposed admixture as the main process involved in Tibetan adaptation. According to them, Tibetans are considered to have dual ancestors, i.e. Sherpa and Han Chinese and the interbreeding between Sherpa and lowlander Han Chinese give rise to adaptive features in Tibetans. Since they speculate the origin of high altitude adaptive mutations might have initially developed 30,000 years ago in Sherpa. Thus, the transfers of advantageous mutation from Sherpa to Tibetans were considered as the process for Tibetan high altitude adaptation. After admixture, natural selection favors for maximizing these Sherpa genes in Tibetans. They reported the high altitude adaptive genetic variants are more in Sherpa than in Tibetans. But they didn't mention the specific genetic differences observed between Sherpa and Tibetans. This study gives controversial idea either selection works through new advantageous mutations or on existing variants becoming beneficial in a new environment.

However, recent studies in Tibetans proposed these simple models of selection from de novo mutation or from standing variation can't explain the several genetic variants observed in EPAS1 gene (Huerta-Sánchez et al., 2014). Thus, they suggested Denisovans introgression might be the possible explanation for unusual EPAS1 genetic variants seen in Tibetans. The availability of whole genome sequences of two archaic human groups; Neanderthals (Green et al., 2010; Prüfer et al., 2014; Reich et al., 2010) and Denisovans (Meyer et al., 2012) made possible to study the level of gene flow between archaic humans and modern humans staying in different region of globe. The Denisovan ancestry (at different level) is reported widely in different modern human populations' from island Southeast Asia, Oceania, Eastern Eurasian and Native American populations. Among these, higher level of Denisovan ancestry is found in Oceania populations (Qin & Stoneking, 2015).The Neanderthal ancestry is also found more in East Asians than West Eurasians and Native

Americans (Qin & Stoneking, 2015). These evidence shows there was admixture between archaic and modern human groups in East Asian populations. The archaic introgressions have provided various advantages in non-African populations (Racimo, Sankararaman, Nielsen, & Huerta-Sánchez, 2015).

EPAS1 encodes a transcription factor which play important role in response to low oxygen environment at high altitudes (Beall et al., 2010). A highly differentiated 5 SNP haplotype motif (AGGAA) within a 2.5kb window in intronic region of EPAS1 in Tibetans are suggested to be introgressed from Denisovan. This Denisovan introgressed segment of EPAS1 is present at higher frequency (~80%) in Tibetans and absent in other individual from 1000 Genomes (1000 Genomes Consortium, 2012) except in 2 Han Chinese. The haplotype network of modern human and Denisovan haplotypes revealed that Tibetans contained the closest-matching haplotype with Denisovan. The highly significant D statistics, S* statistics and haplotype length provide the statistical evidence for introgression (Huerta-Sánchez et al., 2014). Additionally, they reported introgressed haplotype of EPAS1 are significantly associated with lowering haemoglobin concentrations in Tibetans. However, there is not any functional data to support that the introgressed region is responsible for high altitude adaptation in Tibetans.

2.3.3 High altitude adaptation in Sherpa

The term Sherpa ("Shyar" (east) and "Pa" (people), means "people who came from the east," in Tibetan language. The Khumbu region of Solukhumbu district in Nepal is the homeland of Sherpa people. From their ancestral place, Sherpa have migrated to different areas of Nepal like Dolakha, Ramechap, Olkadunga, Sankhuwasabha, Sindhupalchok, Taplejung and Khotang. Currently some Sherpa are staying in India and China and they are considered as the descendants of Solu-Khumbu emigrants (Oppitz, 1974). In India, some Sherpa are living in Darjeeling district of West Bengal (Oppitz, 1974; Stevens, 1996). In China, Sherpa are residing at Dingjie County and Zhangmu Town in Tibet (Kang et al., 2013). However, majority of Sherpa are staying in Nepal and their population size is around 112,946 according to records of Nepal National Population Census, 2011.

Sherpa speaks Tibeto-Burman languages and follow Buddhism. They stay at altitudes above 2,000 meters. Sherpa also engaged in agriculture and grows mostly barleys and potatoes. They involved in pastoralism of yak and sheep. Sherpas cultures are closely related with Tibetans. According to historical literature, Sherpas ancestors were thought to migrate from eastern Tibet to Nepal which is located 1,200 kilometers away from their present homeland of Khumbu region (Oppitz, 1974). Sherpa ancestors migrate to Khumbu region in Nepal crossing Himalayas via Nangpa La pass. Although Nangpa La is located at high altitude (5200m) it suggest ancestors of Sherpa were already adapted at high altitude so they were capable to migrate defeating harsh environmental conditions of Himalayas. During migration they bring yak, Himalayan goat and sheep from Tibetan plateau. Different hybrids of yaks (jyokyo) are still raised in Sherpa villages in Nepal for carrying goods in the Himalayan trail.

Sherpa extraordinary performances in expeditions to Mount Everest are widespread in mountaineering field since 1920s. Sherpas are the legendary climbers and they are crowned with different titles like the Tigers of climbers, King of mountain and so on. Sherpa became

world-renowned Himalayan population of Nepal after Tenzing Norgay's Sherpa and Sir Edmund Hillary make their first ascent to Mount Everest, the tallest peak of the world in 1953. Sherpas have been the major attractions for scientists, climbers, journalists and for other general public because of their heroic performances in Himalayas. Several previous studies have defined Sherpa in different physiological and genetics aspects. The summary of previous studies in Sherpa is described below.

Hypoxic ventilatory response-respiration physiology

The hypoxic ventilatory response (HVR) among different highlander populations is different. It was speculated that highland populations ventilate less than other acclimatized persons at high altitude. Different studies suggested that Sherpas showed less ventilation at high altitude (Lahiri & Milledge, 1967; Lahiri et al., 1967; Milledge & Lahiri, 1967; Samaja, & Cerretelli, 1997). It was considered as the sign of hypoxic adaptation since hyperventilation wasted energy (Santolaya & Schoene, 1989). In contrast, some studies showed that the HVR in Sherpa is comparatively similar with acclimatized lowlanders at high altitude (Hackett et al., 1980; Bengt Kayser et al., 1994).

The hyperventilation has both advantages and disadvantages for performances at high altitude. On one hand, hyperventilation can increase oxygenation which may help in high altitude (Masuyama et al., 1986; Schoene et al., 1984). But on the other hand, hyperventilation may cause brain damage due to decrease in cerebral blood flow caused by hypocapnia (less than the normal level of carbon dioxide in the blood) (Hornbein, Townes, Schoene, Sutton, & Houston, 1989). Thus, this implicates that having optimum HVR is a better strategy.

Spirometric values

The comparison of spirometric values of Sherpa with other lowlanders shows Sherpa has significantly larger spirometric values (Havryk et al., 2002). Sherpa shows greater forced expiratory volume (FEV1) and forced vital capacity (FVC). Pugh et al (Pugh et al, 1962) suggested Sherpas' have greater cardiac output and higher diffusing capacity of lungs for adaptation at high altitude.

Hemoglobin

Hemoglobin is involved in the regulation of O_2 transport in two ways: a long-term adjustment in red cell mass is mediated by erythropoietin (EPO), a response to renal oxygenation. Short-term, rapid-response adjustments are mediated by ventilation, cardiac output, hemoglobin oxygen affinity (P50), barriers to O_2 diffusion, and the control of local micro vascular tissue perfusion(Winslow, 2007).

Several previous reports on hemoglobin concentrations (Hb) showed that Sherpa have lower Hb concentration than other acclimatized lowlanders at high altitude (Adams & Shresta, 1974; Adams & Strang, 1975; Beall & Reichsman, 1984; Samaja, Veicsteinas, & Cerretelli, 1979; Winslow et al., 1989; Wu et al., 2013). The unalleviated Hb concentration of Sherpa was considered as the symbol of adaptation at high altitude. The EPAS1, EGLN1 and PPARA are reportedly involved in regulating hemoglobin levels in Tibetans (Beall, 2007; Simonson, 2010).

Oxygen saturation

Some studies reported that Sherpa had higher arterial oxygen saturation (SaO_2) than Westerners (Hackett et al., 1980; Bengt Kayser et al., 1994). Others mention there is a not significant difference in SaO_2 between Sherpa and other lowlander (Keyl et al., 2000; Michele Samaja et al., 1997). This inconsistency in arterial oxygen saturation may be due to different

factors like sample size, participants health history, instrument used, saturation probe location and measurement protocols (Weitz & Garruto, 2007). Beall *et al 2004* reported higher infant survival of Tibetan women with high oxygen saturation genotypes (Beall & Goldstein, 2004).

Sherpas staying permanently at 4000 m do not have increased hematological parameters (i.e., red cell number, hematocrit, hemoglobin content, and 2,3-diphosphoglycerate/hemoglobin ratio) and have a higher affinity of blood for oxygen as compared with acclimatized Caucasians. In contrary, Sherpas permanently living at low altitude have lower affinity of blood for oxygen than do Caucasians living at comparable altitude (Morpurgo et al., 1976). The P50 value in Sherpa was similar with acclimatized lowlanders. P50 means hemoglobin oxygen affinity. The rise in the ratio of 2, 3-DPG concentration to Hb at high altitude in both natives and in newcomer is balanced by hyperventilation and alkalosis induced by hypoxia (Winslow, 2007).

Heart and cardiac metabolism

Sherpa have higher heart rate in comparison to lowlander climbers at high altitude (L. Pugh et al., 1964; L. G. C. E. Pugh, 1962). Even Sherpa from lowlands have similar heart rate as seen at highlands suggested it might be a genetic feature (Marconi et al., 2004).

Cardiac Metabolism

The glucose metabolism in mammalian heart is 25-50% more O_2 efficient than the metabolism of free fatty acids. Sherpa had higher myocardial glucose uptake rates after an overnight fast compared with lowlanders was determined by positron emission tomography (Holden et al., 1995).

Given the significant energy demand of the heart, the preferred respiratory fuel of the resting cardiac muscle in the fasted state is normally fatty acids. This is due to the greater degree of reduction and thus superior adenosine triphosphate (ATP) yield per carbon compared with glucose. But under hypoxic conditions, there is a switch in substrate

preference toward glucose instead of fatty acid. Since ATP production per oxygen molecule is higher with glucose than with free fatty acids. Even in such conditions skeletal muscles also prefer carbohydrate substrates instead of fatty acids (Murray, 2009). Thus, shifting towards glucose substrate for energy production makes oxygen demand less and ATP production is also less which reduces energy saving. Six Sherpa heart were examined using 31P-magnetic resonance (MR) spectroscopy under normoxic and hypoxic conditions. The concentration ratio of phosphocreatine (PCr) to ATP in the Sherpa heart at low altitude was about half in comparison to lowlanders. (Hochachka et al., 1996) The PCr/ATP was not further decreased in the Sherpa heart following sustained inhalation of a hypoxic gas mixture, perhaps suggesting a metabolic optimization for hypoxic conditions.

In Sherpas under acute hypoxia, the heart rate increased by 20 beats per min from resting values of about 70 beats per min, and the percent saturation of hemoglobin decreased to about 75%. However, these changes did not effect on the PCr/ATP concentration ratios, which remained at about 50% of the values expected in healthy lowlanders (Hochachka et al., 1996). Interestingly, in lowlanders returning from high altitude, myocardial PCr/ATP was also found to be lowered initially but later on after de-acclimatization cardiac PCr/ATP recovered on them (Holloway et al., 2011).Whereas, in case of Sherpa heart, 27 days of residence at low altitude also show no rise in cardiac energetic reserves (Hochachka et al., 1996). Thus, this reduction of PCr/ATP in Sherpa might be response to hypoxic adaptation. The decrease in PCr/ATP after hypoxic exposure has also been shown in animal studies (Portman , 1996).

At high-altitude there will be stress in cardiovascular system to meet the metabolic demand for O2. The cardiac structure and function may have adapted to facilitate O2 delivery. The previous studies in Sherpa provide limited information and they mention Sherpa having higher maximal heart rate, lower pulmonary vascular resistance and no differences in resting

cardiac output in comparison to lowlanders. Ventricular form and function are intrinsically linked through the left ventricular (LV) mechanics that facilitate efficient ejection, minimize myofibre stress during contraction and aid diastolic recoil. The recent data in adult Sherpa shows smaller absolute and relative LV size (Stembridge et al., 2014). The cardiac mechanics also differ in Sherpa when compared with lowlanders at high altitude. These differences are characterized by a reduction in resting systolic deformation and slower diastolic untwisting, a surrogate of relaxation. These changes may reflect a functional cardiac adaptation that affords Sherpa the same mechanical reserve seen in lowlanders at sea level, which is absent when they ascend to high altitude (Stembridge, 2014).

Some studies reported elderly people living at high altitude have a higher risk of cardiovascular disease than low altitude peers due to arterial wall stiffening (Otsuka et al., 2005).Thus they suggest lifelong adaption to high altitude does not appear to provide any functional cardio protective benefits.

Artery structure and function

Some researcher focuses on vascular adaptions at high altitude. They compared vascular function and structure in: healthy lowlanders (n=12) during acute hypoxia and prolonged (~2 weeks) exposure to high altitude, and Sherpa (n=12) at 5050 m. They measured brachial endothelium-dependent flow-mediated dilatation (FMD), endothelium-independent dilatation (via glyceryl trinitrate GTN), common carotid intima-media thickness (CIMT) and diameter and arterial stiffness via pulse wave velocity (PWV). Cephalic venous biomarkers of free radical-mediated lipid peroxidation (lipid hydro peroxides, LOOH), nitrite (NO_2) and lipid soluble antioxidants were also measured. Finally, they found compared to sea level, high altitude reduced lowlanders' FMD and GTN-induced dilatation and raised central PWV. But these values were similar with Sherpa at high altitude suggesting these changes not dependent on time spent at high altitude. However, compared to lowlanders at sea level

and high altitude, highlanders had a lower carotid wall: lumen ratio, narrower CIMT and wider lumen. It is unclear if these differences are reflective of a beneficial adaptation in high altitude or specific mechanism commonly seen in all for hypoxic response. They further suggested high altitude exposure can increase carotid diameter in lowlanders showing as an adaptive response to high altitude (Lewis et al., 2014).

Cerebral function

There is a risk of brain damage when low-landers attempt to climb the highest summits. Some studies did comparative studies among Sherpas and lowlander climbers, and found Sherpa having less psycho-neurological symptoms during sojourns to extreme altitude (>8,000 m). Similarly magnetic resonance abnormalities was seen in lowlander climbers after their return from high altitude but it was absent in most of the Sherpa (Garrido et al., 1996). It shows Sherpa having better brain protection when exposed to extreme altitude.

Cerebral autoregulation (CA) was studied in Sherpa and lowlander newcomers at high altitude. But CA of both Sherpa and lowlander was not functioning at 4243m altitude (Jansen, & Odoom, 2000). Those people living at 3,440m and lower have functioning CA (Jansen & Ince, 2007). Thus the CA becomes impaired above 4243m. Another study suggested cerebral artery diameter is changing during alterations of inspired oxygen partial pressure. They measured cerebral blood flow (CBF) at 7,950m of lowlander climber and demonstrate that above 5,300m, middle cerebral artery diameter increases (n=24 at 5,300 m, 14 at 6,400 m, and 5 at 7,950 m) (Wilson et al., 2011). It is not known in case of Sherpa if their cerebral diameter also increases or not at 7,950m.

The brain of hypoxia-tolerant vertebrates is assumed to be adapted by reducing rate of ATP utilization and ATP production. But the results of positron emission tomography (PET)

of cerebral glucose metabolism in Sherpa maintain normal values similar with lowland controls (Hochachka et al., 1996).

Pregnancy and fetal birth weight

At high altitude there is decrease in birth weights, averaging a 100-g fall per 1000 m elevation gain, as the result of restriction of third trimester fetal growth (Moore, 2003). But Sherpa do not show the reduction in birth weight at altitude. The mean birth weights were similar in Sherpa women (n=17) living at low (1,330 m) and high (3,930 m) (n=21) altitudes (C. Smith, 1997). It is unknown about the cause of fetal growth protection. It may be due to several factors like: greater delivery or metabolism of oxygen, glucose or other substrates or to other considerations such as mechanical factors protecting fragile fetal villi, the creation of a reserve protecting against ischemia/reperfusion injury, or improved placental O_2 transfer as the result of narrowing the A-V O_2 difference and raising uterine PvO_2 (Moore, Charles, & Julian, 2011).

Skeletal muscle structure and its mitochondrial volume

Sherpas (n=5) possess significantly greater number of capillaries per cross-sectional area of muscle compared with sedentary unacclimatized lowlanders (Kayser, Hoppeler, Claassen, & Cerretelli, 1991). They have reduction in muscle fiber cross-sectional area and increased capillary density-to-muscle fiber ratio (Kayser et al., 1991) and could serve for proper oxygen flow to the working muscles. Sherpa have 25% lesser mitochondrial volume density than lowlanders but still have higher maximal oxygen consumption-to-mitochondrial volume ratio (Kayser et al., 1991). However, it was the same in both Sherpa and acclimatized climbers. Based on this fact, we can't say Sherpa have endurance capacity.

The role of low mitochondrial volume might help to improve better coupling between ATP demand and supply pathways as well as better metabolite homeostasis. It will also help for choosing metabolism of carbohydrates as a fuel and reducing intramyocellular lipid

substrate stores (Hoppeler, Vogt, Weibel, & Flück, 2003). Overall it will increase in amount of ATP produced per volume of oxygen consumed and thus favors a high yield of muscle work-to-O2 ratio (Hoppeler & Vogt, 2001). There is controversy yet that if hypoxia causes increase in reactive oxygen species production (ROS) or not. There are some evidences suggesting mitochondria acting as oxygen sensors by variation of ROS production. Some studies pointed Sherpa are protected from ROS-induced tissue damage and possess specific metabolic adaptations (Gelfi et al., 2004). However, there is no common agreement on role of mitochondria in Hypoxic adaptation.

The blood lactate at high altitude in acclimatized lowlander is same as at sea level (West, 1986). There is lower than expected blood lactate in native highlanders compare with lowlanders. At higher altitude also there is lower post exercise blood lactate peak. The lactate production is a function of how metabolic and physiological control contributions are organized in the complex pathways of ATP supply and demand should be generally applicable to muscle during exercise in many differing physiological states. The role of lactate in high altitude performance is paradoxical (Hochachka et al., 2002).The noninvasive 31P magnetic resonance spectroscopy (31P MRS) was used to explore the nature of muscle metabolism during rest-exercise-recovery transitions in Sherpas (Allen et al., 1997). In this 31P MRS studies they found an inverse relationship between lactate and phosphocreatine (PCr) concentrations in Sherpa. They conclude such relationship can be easily seen from exercise than during exercise in other populations too. The lactate paradox in high altitude physiological performance was observed (Allen et al., 1997).

V̇O2max exercise capacity

Sherpa superior performances in carrying heavy loads at high altitude are also discussed by several studies. They used different methods to know about their efficiency on

carrying loads on their backs. They found Nepalese porters (mainly Sherpa) working at Everest region can carry heavier load in comparison with other porters from different parts of the world since they have higher metabolic economy (Bastien et al., 2005; Minetti, Formenti, & Ardigò, 2006).The actual mechanism for this process is not known but some considered mean maximal oxygen consumption (V̇O2max) as the index for determining working capacity and fitness at high altitude.

Either Sherpas have a greater V̇O2max or not compared with lowlanders at the same altitude remains controversial. Most of the studies claimed Sherpa have greater V̇O2max (Garrido et al., 1997; Lahiri et al., 1967; Pugh et al., 1964; Pugh, 1962) but others disagree with them. They found similar or even lower V̇O2max values in Sherpa (staying at low altitude) in comparison with lowlanders (Kayser et al., 1994). The differences in study design might have make variation in results. Another recent study concluded that high-altitude natives achieved a higher mean V̇O2max at hypoxia and smaller V̇O2max decrement with increasing hypoxia (Brutsaert, 2008).

Work efficiency and work economy

Work efficiency (WE) may be defined as the ratio of total work done (internal and external) to total energy expenditure (aerobic, anaerobic, and contributions to the phosphagen system).This is the theoretical definition but in practical it is difficult to measure non oxidative energy expenditure as well as the internal work. In reality, WE is the ratio of external work(output) to metabolic cost(input).The VO_2 can only be measured for input.

In contrast, work economy is defined as the oxygen cost for a specific activity, and the external work measures are not necessary. For example as in case of treadmill exercise (Brutsaert, 2008). Regarding Work efficiency also there are different views. Some say work efficiency is similar in both Sherpa/Tibetans and lowlanders (Kayser et al., 1994; Lahiri et al., 1967). While other argue that Tibetans have significantly higher work efficiency (Ge et al.,

1994; Niu, Wu, Li, Chen, & Song, 1995). Tibetans are staying at higher altitude than Sherpa. So, there may be intra-population differences also since Tibetans living at high altitude (4,400 m) demonstrate a greater work performance for given oxygen uptake compared with those at lower altitudes (3,658 m)(Curran, Zhuang, Droma, & Moore, 1998).

One previous study shows higher work economy in Sherpa porters. The Sherpa porters usually carry head supported loads exceeding their body weight. They can carry heavier loads in comparison with other lowlander porters at high altitude (Bastien et al., 2005). Work economy can't only say Sherpa have higher work performance at high altitude. Different other social and economic factors may have compelled them to carry higher loads. They are paid at high altitude based on the weight they carry. There is insufficient scientific data to support that Sherpa have higher work economy although they are well adapted at high altitude.

Growth, body Weight and basal metabolic rate

Growth and body weight of Sherpa children living at low and high altitude were compared. They found both Sherpa children growth was retarded compared to other high altitude populations of Peru and Ethiopia (Gupta & Basu, 1991; Pawson, 1976).The increased chest circumference represented as developmental acclimatization to hypoxia among Peruvian high-altitude natives was also not seen in Sherpa (Pawson, 1977). The growth can be effected by nutritional and genetic factors also. The height in Sherpa affected by hypoxic environment or the result of inherited pattern of development is unknown.

The low lander acclimatized at high altitude by increasing their basal metabolic rate (BMR)(Butterfield et al., 1992; Gill & Pugh, 1964).The BMR is not depended on the environmental temperature(Nair, Malhotra, & Gopinath, 1971) so the lower temperature at higher altitude have no any relation in BMR. There is weight loss in lowlander population when they climb high altitude. In contrast, Sherpa have constant weight at high altitude

(Boyer & Blume, 1984). Sherpa are normally very lean so there is not significant weight loss on them. The weight loss below 5,400m is mostly from fat loss and above that elevation it is from muscle protein catabolism (Boyer & Blume, 1984). Weight loss is the symptom of high-altitude deterioration (Ward, 1954). Sherpa resistance to weight loss and their lean body size may be due to resilience to high altitude deterioration.

NO metabolism

Nitric oxide (NO) is a ubiquitous signaling molecule produced through the metabolism of L-arginine by nitric oxide synthases (NOS). NO controls the activity of cytochrome c oxidase (CcO) which is involved in cellular energy production of mitochondria (Taylor & Moncada, 2010).When there is higher oxygen concentration CcO is in oxidized state and it consumes NO. But when there is lower oxygen concentration CcO is reduced and NO is not consumed and accumulates in the microenvironment (Taylor & Moncada, 2010). NO is involved in control of blood pressure, blood flow and other vital bodily functions (Moncada, Palmer, & Higgs, 1991).Beside this, NO is an antioxidant(Wink et al., 2001), and a regulator of intermediary metabolism(Jorgen, Fried, Fu, Meininger, & Wu, 2006).The vascular system of Sherpa were studied under baseline conditions and during maximal vasodilation after 2 min leg occlusion. It shows superior ability to increase blood flow velocity as a response to muscular ischemia in Sherpa compared to sea-level natives (Schneider et al., 2001).The authors concluded it might be due to the differences in conduit vessel function and speculated this could be due to effect of NO.

Sherpas have increased basal levels of angiogenic (vascular endothelial growth factor A (VEGF-A), interleukin (IL-8)and lymphangiogenic factors [VEGF-C and D] which help for increased microcirculatory flow (Patitucci & Lugrin, 2009).They suggest these may be also the adaptation features in Sherpa. The levels of circulating nitrogen oxides in Sherpa is

lower where as in Tibetans are higher. The cause of differences in nitrogen oxides between Sherpa and Tibetans is unknown (Beall, Laskowski, & Erzurum, 2012).

The genotypes of Glu298Asp and eNOS4b/a polymorphisms of the endothelial nitric oxide synthase (eNOS) gene were identified in Sherpa (Droma et al., 2006). They measured metabolites of nitric oxide (NO x: nitrite and nitrate) in serum. The frequencies of the Glu and eNOS4b alleles were significantly higher in Sherpas (Glu: 87.5%; eNOS4b: 96.7%) than in non-Sherpas (Glu: 77.9%, $p = 0.036$; eNOS4b: 90.5%, $p = 0.009$). In addition, the combination of the wild types of Glu298Glu and eNOS4b/b was significantly greater in Sherpas (66.7%) than non-Sherpas (47.7%, $p = 0.008$) (Droma et al., 2006). However, the serum NO x was significantly lower in Sherpas (53.2 ± 4.6 µmol/L) than in non-Sherpas (107.3 ± 9.0 µmol/L, $p < 0.0001$) (Droma et al., 2006). Finally, they concluded that the nitric oxide metabolites (NO x) in serum vary individually, thus it is not a reliable indicator for endogenous nitric oxide production (Droma et al., 2006).

Genetics

A novel dinucleotide repeat polymorphism in intron 13 of the hypoxia-inducible factor (HIF)-1α gene was found in Sherpas in comparison to lowlanders (Suzuki et al., 2003). These genetic variations were initially speculated as having role in hypoxic adaptation. Another study performed a human genome scan using polymorphic DNA markers in eight unrelated Sherpa porters living in the Solu-Khumbu area. They find three polymorphic DNA markers (D6S1697, D14S274, and D17S1795) likely to involve in adaptation to hypobaric hypoxia (Malacrida et al., 2007). The angiotensin-converting enzyme (ACE) gene was also analyzed in 105 Sherpas (Droma et al., 2008). They found high allelic frequency of insertion allele (I) in ACE gene of Sherpas and conclude that it may be associated with maintaining low serum levels of ACE (Droma et al., 2008). The insertion allele (I) of the ACE gene was also

reported before as having elite endurance performance among British mountaineers (Montgomery et al., 1998).

Several genome-wide allelic differentiation scans (GWADS) have reported high allele frequencies across SNP loci in mainly EPAS1 and EGLN1 gene for Tibetan high-altitude adaptation (HAA). In Sherpa few tag single-nucleotide polymorphisms (SNPs, rs13419896, rs4953354, and rs4953388) of EPAS1 gene were analyzed and they found similar patterns in Sherpa and Tibetans in case of these three tag SNPs (Hanaoka et al., 2012).They also found equal average level of serum erythropoietin between Sherpas at 3440 m to that of non-Sherpas at 1300 m. This resistant response of erythropoietin in Sherpas was marked as genetic features of high-altitude adaptation (Hanaoka et al., 2012).

Figure 2: Key physiological differences between Sherpas/Tibetans and lowlanders (Gilbert-Kawai, Milledge, Grocott, & Martin, 2014)
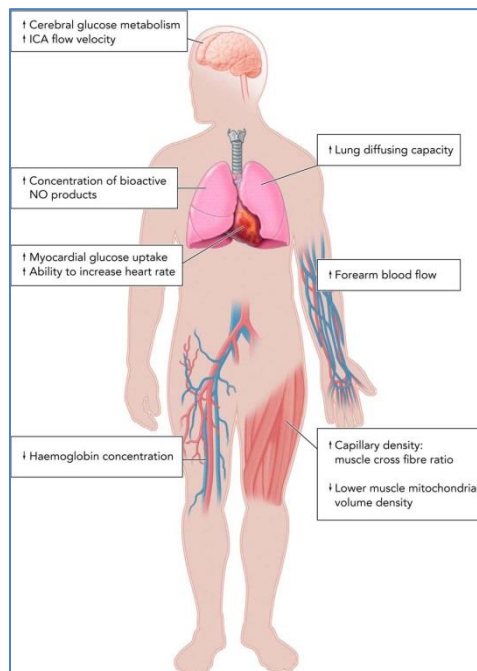
After modern human migrated out of Africa 60,000 years ago, they colonized in different high altitude regions across the planet; the Tibetan Plateau, the Andean Altiplano, and the Ethiopian Highlands. Each of these populations has different genetic mechanism for

adaptation at high altitude. Their physiological changes are also different for behaving in decreased oxygen environment. The archeological data suggest the Tibetan Plateau and the Andean Altiplano were first populated nearly above 25,000 and 11,000 years ago respectively (Aldenderfer, 2003; Moseley, 1997; Zhao et al., 2009). The different time duration for residing at high altitude are thought to involve for their variation in adaptation mechanism. There occurred local adaptation to cope with hypobaric hypoxia in different regions (Foll, Gaggiotti, Daub, Vatsiou, & Excoffier, 2014). The Tibetan adaptation is discussed in previous chapter; here we describe briefly Ethiopian and Andean adaptation.

2.3.4 High altitude adaptation in Ethiopians

There is a different set of candidate genes (CBARA1, VAV3, ARNT2 and THRB), for high altitude adaptation in Ethiopian and these genes were not the key candidate genes in Tibetan and Andean populations (Laura B Scheinfeldt et al., 2012). However, ARNT2 and THRB found in Ethiopian were also involved in HIF-1 pathway in case of Tibetan and Andean populations. EPAS1, EGLN1 and PPARA are reported as three major candidate genes in Tibetan studies. EPAS1 and EGLN1 are not having selection signal in Ethiopian highlander but PPARA is having signal of selection in Ethiopian. Like Andeans, Ethiopian highlanders also have higher hemoglobin levels at high altitude.

There is not similar variant for association with hemoglobin among Tibetans and Ethiopians. The SNP, rs10803083 on chromosome 1 was associated with variation in Hb levels in Amhara ethinic group of Ethiopia (Alkorta-Aranburu et al., 2012) There are no known genes within 600 kb of these SNP and the nearest genes is phospholipase D family member 5 (PLD5) and centrosomal protein 170 kDa (CEP170), are not the candidate genes for variation in Hb levels (Alkorta-Aranburu et al., 2012). Next SNP; rs2899662, also shows association signal with Hb in Ethiopians and is located in retinoid-related orphan receptor

alpha (RORA), a hypoxia candidate gene. RORA encodes a protein that induces the transcriptional activation of hypoxia-inducible-factor-1alpha (HIF-1α) (Kim et al., 2008). There was not significant genotype association with oxygen saturation levels in Ethiopians highlanders.

Another study (Huerta-Sánchez et al., 2013) in Ethiopian finds the strongest signal of selection in BHLHE41 gene (also known as DEC2 or SHARP1). This gene is involved in hypoxic-response pathways and has interaction with HIF-1α. The HIF-1 α /ARNT1 protein heterodimer plays an important role in the hypoxia-induced transcription of vascular endothelial growth factor (VEGF)(Forsythe et al., 1996), and BHLHE41 negatively controls VEGF expression by its interaction with HIF-1 α /ARNT1 activation (Sato et al. 2008). In addition, BHLHE41 is also a part of the circadian clock pathway but the relation between hypoxic conditions and circadian cycles is unknown. It shows HAA arose independently due to convergent evolution in Ethiopian.

2.3.5 High altitude adaptation in Andeans

There are different sets of candidate genes for Andean adaptation which includes; ADRA1B, ARNT2, ATPIA1, ATPIA2, CDH1, COPS5, CXCR4, EDN1, EDNRA, EGLN1, EGLN2, ELF2, FRAP1, ILIA1, ILIB, IL6, IGFBP1, IGFBP2, MDM2, MMP2, NOS1, NOS2A, NOTCH1, NRP1, NRP2, POLRA, PIK3CA, PIK3CG, PRKAA1, PRKAA2, SNAI3, SPRY2, TF, TGFA, TNC, TNF, VEGF(Bigham et al., 2009). However EGLN1, one HIF pathway gene is under positive selection in both Andeans and Tibetans (Bigham et al., 2010; Bigham et al., 2009). In addition to EGLN1, the other genes like HBE1, PRKAA1(protein kinase AMP-activated alpha 1 catalytic subunit),VEGFA (vascular endothelial growth factor A) and NOS2A (nitric oxide synthase 2A) are the top candidate genes in Andean adaptation(A. Bigham et al., 2010).EPAS1 gene is not having selection signal in Andeans.

The concentrations of hemoglobin are elevated in high altitude Andean population in contrast Tibetans have lower hemoglobin concentration (Beall, Brittenham, Macuaga, & Barragan, 1990).

PRKAA1 acts as a cellular energy sensor under ATP-deprived conditions, which usually happens in hypoxic condition. The gene product of PRKAA1 is also essential for HIF-1 transcriptional activity (Ivan et al., 2001). NOS2A interacts with nitric oxide synthase isoforms and produces nitric oxide (NO) from arginine and oxygen.NO helps in arterial smooth muscle relaxation, vasodilation and increased blood flow. The NO production is higher in Tibetans in comparison to sea level control (Erzurum et al., 2007).Several studies in Tibetans and Andeans have shown that the increased blood flow in the uteroplacental circulation help in maintaining fetal growth in Tibetans as well as Andeans (Julian et al., 2009; Moore, Zamudio, Zhuang, Sun, & Droma, 2001;Wilson et al., 2007). HBE1, a globin family gene, which is part of beta globin gene cluster also contain significant SNPs for Andean adaptation. Similarly, there is positive signal found in RAS gene in Andeans but signal strength is weak (Bigham et al., 2010).Thus, it shows, the high altitude adaptation in Andean is also governed by different genes of HIF-1 pathway.

The three different highlander populations: Ethiopians, Andeans and Tibetans have different high altitude hypoxic adaptation mechanism although there is an overlap of EGLN1 gene between Tibetan and Andean adaptation. These studies suggest their occurred convergent evolution for high altitude adaptation among these populations, since there are changes in different genes belonging in the same HIF pathway (Foll et al., 2014)

2.　OBJECTIVES OF THE PRESENT STUDY

Sherpa superior adaptation capacity at high altitude is a popular story among scientists, climbers, and to other general public. The extraordinary performance displayed by Sherpa in climbing field might have compelled us to believe in their superior adaptation capacity. Many researchers have defined Sherpa in different physiological and genetics aspects which were discussed briefly in literature review section of this thesis. Most of the physiological behavior studied previously among Sherpa and Tibetan are found similar except Nitric oxide concentration. There are many different traits uniquely seen in Sherpa and which is different from Lowlanders suggesting Sherpa are well adapted in high altitude. There have been extensive genetic studies in Tibetans, and EPAS1 and EGLN1 were reported as the key genes for adaptation in Tibetan plateau (Beall et al., 2010; Bigham et al., 2010; Huerta-Sánchez et al., 2014; Lorenzo et al., 2014; Peng et al., 2011; Simonson, 2010; B. Wang et al., 2011; Xiang et al., 2013; Xu et al., 2011; Yi et al., 2010).

The recent study on Sherpa (Jeong et al., 2014) regarding the evolutionary history of these adaptive alleles of EPAS1 and EGLN1 gene notify these adaptive genetic traits were initially originated in Sherpa. They further conclude that admixture causes transfer of adaptive alleles from Sherpa to Tibetans, claiming Sherpa and Han Chinese as the ancestral populations of Tibetans. They show higher amount of high altitude proxy in Sherpa compared with Tibetans but didn't mention the distinct high altitude proxy in Sherpa which is different from Tibetans. This study creates controversy regarding the origin of Sherpa since historical literature on Sherpa were viewing Sherpa as recent descendent of Tibetans. So, in order to know the origin of Sherpa population was our major objectives of this study. After knowing the ancestors of Sherpa, the further study in high altitude adaptation will be easier.

Another studies on mtDNA of Sherpa proposed two ND1 variants (G3745A and T4216C) in Sherpa specific Lineages (C4a3b1 and A15c1) as having possible role in High

altitude adaptation (Kang et al., 2013). This contradicts either this lineage occur due to positive selection or population expansion.so we study mtDNA and Y-chromosome marker for resolving these various issues.

It is still unclear about the molecular mechanism of Sherpa adaptation which occurred either by genetic variants of EPAS1/EGLN1 gene or variants of mtDNA Sherpa specific lineages. In comparison to Tibetans, there is limited studies on Sherpa so, the real differences between them is difficult to distinguish. The comparative study between Sherpa and Tibetans was necessary to see if Sherpa have superior adaptation capacity or it is similar with Tibetans, so these studies try to fill the gap. Most of the previous studies in Sherpa are done in limited sample size, testing few genetic loci so it can't give the exact genetic mechanism for high altitude adaptations in Sherpa.

The questions undertaken for the present study can be summarized are as follows:

- Who are the ancestors of Sherpa?

- How did Sherpa population originate in Nepal?

- What might be the possible reason for Sherpa high altitude adaptation?

- Do Sherpa specific lineages of mtDNA exist due to positive selection or population expansion?

- Do Sherpa have different genes from Tibetans for adaptation at high altitude?

- Are the adaptation capacities of Sherpa and Tibetans similar or different?

3.   **Materials and Methods**

Sample collection

Blood samples of 2-5ml were collected from 582 Sherpa individuals (350 from four villages of the Khumbu region of Nepal and 232 from Zhangmu Town of Tibet) (Figure 1). We also sampled 90 non-Sherpa Nepalese from Solukhumbu district of Nepal (the lowland region with elevation <2000m) who speak Tibeto-Burman languages. Volunteers were asked to reveal the ethnic affinity of their parents and grand-parents in an oral interview, and only volunteers whose parents and grandparents were reported to be Sherpas were included in this study. Sample collection was done randomly from individuals unrelated for at least three generations. All participants provided written informed consent prior to inclusion in the study. Besides blood samples, we also measured hemoglobin and arterial oxygen saturation level from 297 healthy adult Sherpas (126 man and 171 woman) residing in highland villages of Khumbu region, Nepal. The arterial oxygen saturation (SaO2) was recorded using a hand-held pulse oximeter (Nellcor NPB-40, CA) after the individuals take rest of 5-10 minutes. A HemoCue Hb 201+ analyzer (Angelholm, Sweden) was used to measure hemoglobin of fingertip capillary blood.   The protocols of this study were approved by the Internal Review Boards of the Kunming Institute of Zoology, Chinese Academy of Sciences and the Nepal Health Research Council, Kathmandu, Nepal. The DNA extraction was done using the phenol chloroform method and necessary genotyping/sequencing mention in this study was carried out on working DNA Solution prepared from stock DNA.

Part I- Methodology used for studying population history of Sherpa

Y-chromosomal genotyping

The Y-chromosomal SNP markers were genotyped using methods similar to those described in our previous studies (Shi et al., 2005; Shi et al., 2008; Zhong et al., 2011), and

paternal haplogroups were classified based on the up-to-date high-resolution Y-chromosomal phylogenetic tree (Karafet et al., 2008). The commonly used eight Y-chromosomal STR (short tandem repeat) markers (DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393) were typed using fluorescence-labeled primers in an ABI 3130XL Genetic Analyzer (Applied Bio systems, USA). The nomenclature of Y-chromosomal STRs followed the process proposed by Butler et al. (2002) (Butler et al., 2002).

Mitochondrial DNA sequencing and genotyping

The mtDNA HVS-I (range:16024–16466), HVS-II (range:65-417) and a coding region (range:10,220-10,610) were sequenced following the protocol described previously(Yao, Kong, Bandelt, Kivisild, & Zhang, 2002) and then we sequenced several other coding region sites for further confirmation used to define each individual within their respective mtDNA haplogroups. We then conducted amplicon sequencing of the entire mitochondrial genomes of 89 Sherpa samples using Illumina Miseq, including 79 individuals belonging to the two Sherpa-specific sub-haplogroups (32 A15c1 and 47 C4a3b1) and the other 10 individuals belonging to other haplogroups (5 M9a, 1 M70, 1 C4a1a1, 1 D4j1b and 1 A15c). The complete mtDNA was amplified via PCR of two overlapping fragments (9.3 kb and 9.2 kb)(Dames et al., 2013). Amplicons were analyzed by agarose gel electrophoresis, individually purified via resin purification (Promega), quantified on a Bioanalyzer (Agilent Technologies 2100 Bioanlyzer) and equimolarly pooled. The Nextera DNA Sample Preparation guide (15027987) was followed for pooled amplicons to create indexed, paired-end libraries that were later sequenced on an Illumina Miseq (2×300 base reads) using the manufacturer's protocol (Illumina, San Diego, CA). The resulting high quality sequence data from Miseq had an average coverage of 2000×. Sequence analysis was then done using mtDNA MSR Plug-In and the mtDNA Variant Analyzer provided by Illumina in Human mtDNA Genome Guide (15037958) for the Illumina Sequencing platform.The mtDNA

variants were identified by comparing each sequence with the revised Cambridge Reference Sequence (rCRS)(Andrews et al., 1999), and haplogroups were assigned based on the mtDNA phylogenetic tree (Phylo Treemt Build 16,19 Feb 2014)(Van Oven & Kayser, 2009). Comparing a sample the variants produced by Miseq with those from Sanger sequencing confirmed that our MiSeq data were highly accurate and reliable.

Data analysis

The Sherpa mtDNA phylogenetic tree was constructed based on the available whole genome mtDNA sequences of 165 Sherpa samples, including 76 Sherpa samples from previous studies (Kang et al., 2013) and 89 samples obtained in this study. The Phylo Tree mt Build 16(Van Oven & Kayser, 2009) was followed for constructing the phylogenetic tree with exclusion of several mutations (309.1C (C), 315.1C, AC indels at 515-522, 16182C, 16183C, 16193.1C (C) and 16519). The median-joining network was constructed using NETWORK 4.6.1.0 (Fluxus Engineering) (Bandelt, Forster, & Röhl, 1999) (Bandelt et al. 1999) and was constructed for the major haplogroups of both mtDNA and Y chromosome.

The coalescence times of the dominant mtDNA haplogroups were estimated using the ρ statistics (Forster et al., 1996) and standard errors were calculated following Saillard et al (Saillard, 2000). The complete mtDNA genome sequences were used for age estimations by excluding 309.1C (C), 315.1C, AC indels at 515-522, 16182C, 16183C, 16193.1C (C) and 16519). The mutation rate of 3,624 years per mutation was used for entire mtDNA genome to estimate the timeframe for the most recent common ancestor (TMRCA) of a haplogroup (Soares et al., 2009).

To observe the genetic relationships of Sherpas with other populations, we constructed PCA maps based on both mtDNA and Y-Chromosome haplogroup frequencies as described   previously using MVSP3.13 (Kovach Computing Services, Anglesey, UK)(Richards, Macaulay, Torroni, & Bandelt, 2002). For the neutrality test, the ratios of non-

synonymous vs. synonymous substitutions in the ND1 gene of Sherpas were analyzed using the Nei-Gojobori model (Nei & Gojobori, 1986) in the MEGA5 package (Tamura et al., 2011). We also conducted Tajima's D test using DnaSP Version 5(P. Librado & J. Rozas, 2009). We reconstructed the demographic history of Sherpas using BSP (Drummond, Rambaut, Shapiro, & Pybus, 2005) in BEAST 1.7.5(Drummond, Suchard, Xie, & Rambaut, 2012) using MCMC algorithms (Drummond, Nicholls, Rodrigo, & Solomon, 2002). The HVS-I (16038-16462) sequences of available 661 Sherpa individuals were used for generating BSP. A clock rate of 1.784×10^{-7} substitutions per site per year (Soares et al., 2009) was applied. MCMC sample was based on run of 5×10^{8} steps with sampling every 5,000 steps, and the initial 5×10^{7} were regarded as burn-in. The effective sample size values were in acceptable range (>100) for all the runs. Tracer 1.5 (http://tree.bio.ed.ac.uk/ software/tracer) in the MCMC Trace analysis packages was used to visualize the BSP.

Part II- Methodology used for studying Sherpa high altitude adaptation

Genotyping of EPAS1, EGLN1 and TED

The selected EPAS1 SNPs defined in this study was genotyped by partial sequencing method covering the respective genomic region of these SNPs. Primers were designed using Primer3 software and genotyping was performed doing Sanger sequencing on an ABI 3730 sequencer (Applied Biosystems, Foster City, CA). The LD map of EPAS1 was constructed using Haploview version 4.1(Barrett, Fry, Maller, & Daly, 2005). Similarly the genotyping of two missense mutations (rs12097901G, rs186996510C) of EGLN1 was done using SNaPshot method on an ABI 3730 sequencer (Applied Biosystems, Foster City, CA).The SNaPshot method was applied as described previously(Xiang et al., 2013b).The 5.5 kb re-sequencing of EGLN1 in 50 Sherpa samples was also done using Sanger sequencing in ABI 3730 sequencer (Applied Biosystems, Foster City, CA).In addition, the genotyping of TED was We

genotyped TED using Long-PCR done following the method described in previous studies(Lou et al., 2015).

Genetic association analyses

We collected hemoglobin and arterial oxygen saturation level data along with blood samples from 297 healthy adult Sherpas (126 men and 171 women) for genetic association studies. The three different copy number variations in TED were marked as like SNP genotypes (for example, AA for zero copy, TA for one copy and TT for two copies) and similar association studies method was used for SNPs and TED. The association analysis of eleven EPAS1 SNPs, two EGLN1 SNPs and TED were done using linear regression with an additive model in PLINK v1.07(Purcell et al., 2007). We also tested Hardy-Weinberg equilibrium (HWE) and found these SNPs were not deviated from HWE.

Haplotype network analysis

The 28 EPAS1 SNPs and 17 EGLN1 SNPs were used to construct haplotype network of Sherpa, Tibetans and other five different populations from the 1000 Genomes Projects. The haplotype reconstruction was done by using PHASE program embedded in DnaSP Version 5(Librado & Rozas, 2009).The median joining network was constructed using NETWORK 4.6.1.0, Fluxus Engineering (Hans-Jurgen Bandelt et al., 1999).

F_{ST} calculation

The unbiased estimates of F_{ST} were calculated using method described previously (Weir & Cockerham, 1984).We measured the genetic divergence of EPAS1 SNPs between Sherpa and other populations (CHB, JPT, CEU, YRI).

4. RESULTS AND DISCUSSION

To address the objectives of our study, we explain our results below by two different headings: Population history of Sherpa and Sherpa High Altitude Adaptation. The result presented in this doctoral thesis is based on our results (obtained by Bhandari *et al* 2015;) presented in two different articles, one of them is published in scientific report and another one is under submission.

5.1 Population history of Sherpa

This part of result is included from our studies (*Bhandari et al 2015),* published in scientific report with titled "*Genetic evidence of a recent Tibetan ancestry to Sherpas in the Himalayan region*" 2015.

Y-chromosomal diversity in Sherpa population

In total, DNA samples of 582 unrelated Sherpa individuals representing two geographic populations from Nepal (350 individuals) and Tibet (232) were collected along with samples of 90 non-Sherpa Nepalese from Solukhumbu district of Nepal (Figure 3).

Figure 3. Sampling locations of Sherpa populations in Nepal and Tibet. The altitudes of the locations (dark spots) range from 2,500m-4,000m. The proposed Tibet-Nepal migratory route of the Sherpa ancestors through Nangpa La Pass is indicated with an arrow. The figure was generated using Microsoft Power point 2011 (Microsoft Corporation, USA).
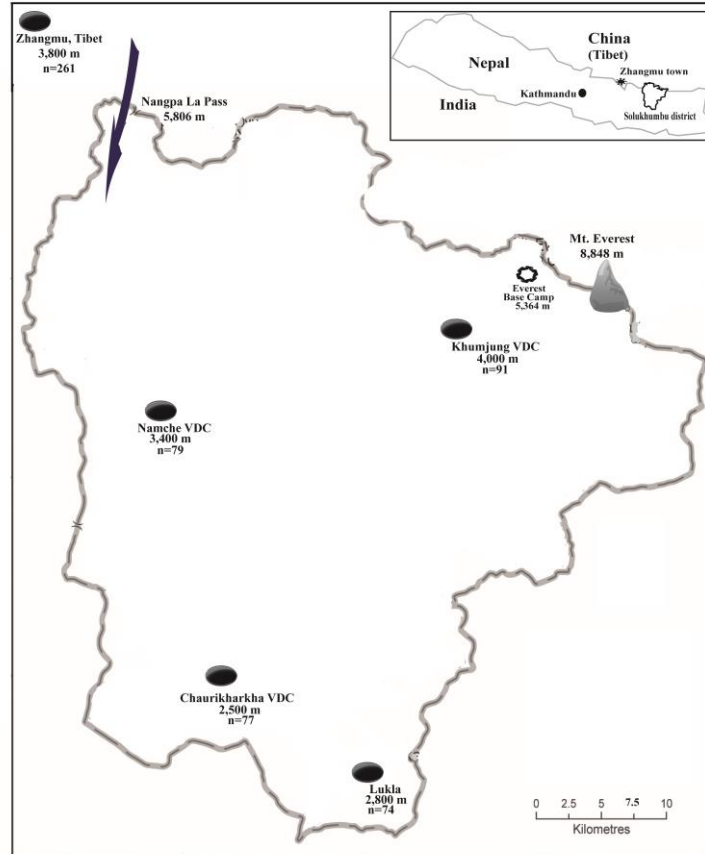
Using the current phylogeny of the human Y-chromosomes, we assigned the samples from 277 Sherpa males into four major Y-chromosomal haplogroups including D-M174 (44.04%), O-M175 (27.08%), F-M89 (9.75%), and K*M9 (7.22%) (Figure 4A). Of these groups, D-M174 and O-M175 are also the two dominant Y-haplogroups in Tibetans (52.84% and 33.13% respectively) (Qi et al., 2013), suggesting a close paternal relation between Sherpas and Tibetans. Likewise, we observed several rare (0.36-3.61%) haplogroups in Sherpas (G-M201, J-M304, M-P256, N-M231, P-M45, Q-M242 and R-M207) that are also seen among Tibetans in similarly low frequencies (Figure 4A). Due to the similarities in the distributions of Y-haplogroups between Sherpas from Nepal and those from Tibet, the data were merged together for further analyses.

Comparing Sherpas, Tibetans, and Han Chinese showed that the D-M174 is the predominant haplogroup in Sherpas (43.38%) and prevalent in Tibetans (52.84%)(Qi et al.,

2013), but rare among both Han Chinese (1.4-6.51%)(Su et al., 2000; Su et al., 1999) and other Asian populations (0.02-0.07%)(Wells et al., 2001), aside from Japanese (34.7%) who possesses a distinct D-M174 lineage highly diverged from those in Tibetans and other Asian populations(Hammer et al., 2006; Shi et al., 2008). Among Tibetans there are five D-M174 sub-haplogroups (D*-M174, 0.08%; D1*-M15, 1.40%; D1a-N1, 13.93%; D3*-P99, 8.62%; D3a-P47, 30.29%), three of which are also present among Sherpas (D1a-N1, 8.41%; D3*-P99, 15.24%; D3a-P47, 19.73%; see Figure 4A). To explore the structure of the D-M174 sub-haplogroups, we constructed the Y-STR network of these sub-lineages and observed that the Y-STR haplotypes among Sherpas represent a subset of those found in Tibetans, implying that the paternal lineages of Sherpas were likely derived from those in Tibetans (Figure 4B-D). Moreover, there were Sherpa-specific Y-STR haplotypes under the D3*-P99 sub-haplogroup closely linked with those in Tibetans (Figure 4C), implying a shallow divergence between Sherpas and Tibetans at this paternal lineage.

A similar pattern of Y-STR network was also seen among the second most common haplogroup O3a3c1*-M117 in Sherpas (23.82 %), which is also present in Tibetans (29.82%), Han Chinese (9.6-16.3%) and many other East Asian and Southeast Asian populations (5.5-16%)(Qi et al., 2013), but Sherpas shared most of their Y-STR haplotypes of O3a3c1-M117 with Tibetans (Figure 4E).

Taken together, D-M174 and O-M175 account for 71.12% of the paternal lineages in Sherpas, bolstering the case for Tibetans as the ancestral population of Sherpas. Likewise, the relative rarity of other Y haplogroups (<10%) among Sherpas (Figure 5) were also rare among Tibetans, and mostly absent in other East Asian populations. Most of these haplogroups are prevalent in India (F*-M89, J2b*-M12 and R1a1-M17) and Island Southeast Asia (K*-M9), suggesting either shared ancient Y-chromosome lineages or limited recent admixture of Sherpas with surrounding populations.

Figure 4. Comparison of Y chromosome diversity among Sherpas, Tibetans and other Asian populations. (A) Phylogenetic tree of the Y chromosome haplogroups and their frequency distributions in Sherpas and Tibetans. (B-E) The Y-STR networks of the major haplogroups showing the distributions of STR haplotypes in Sherpas and other Asian populations. Populations were labeled with different colors based on their belonged language families.
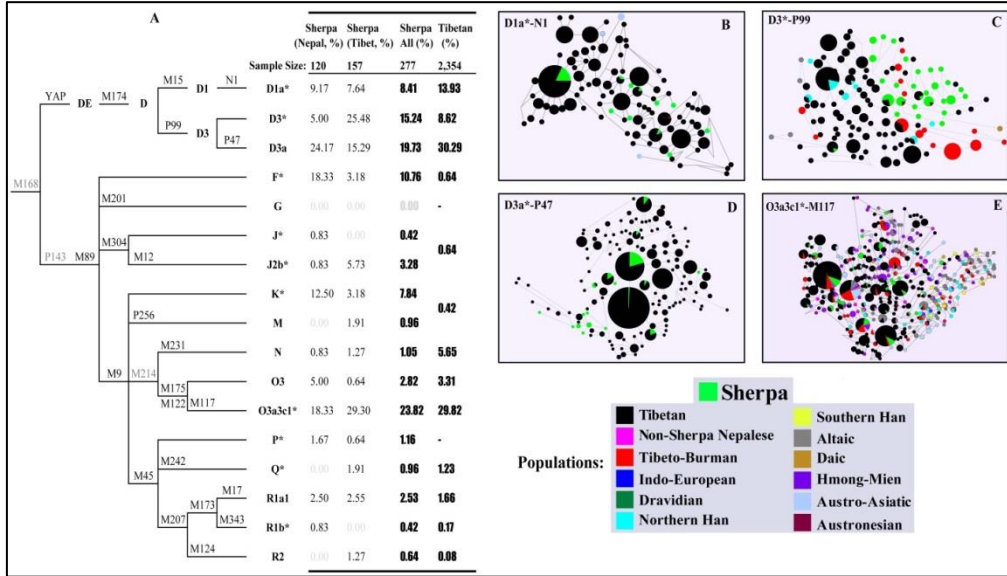
Figure 5. The Y-STR network of four minor haplogroups in Sherpas.
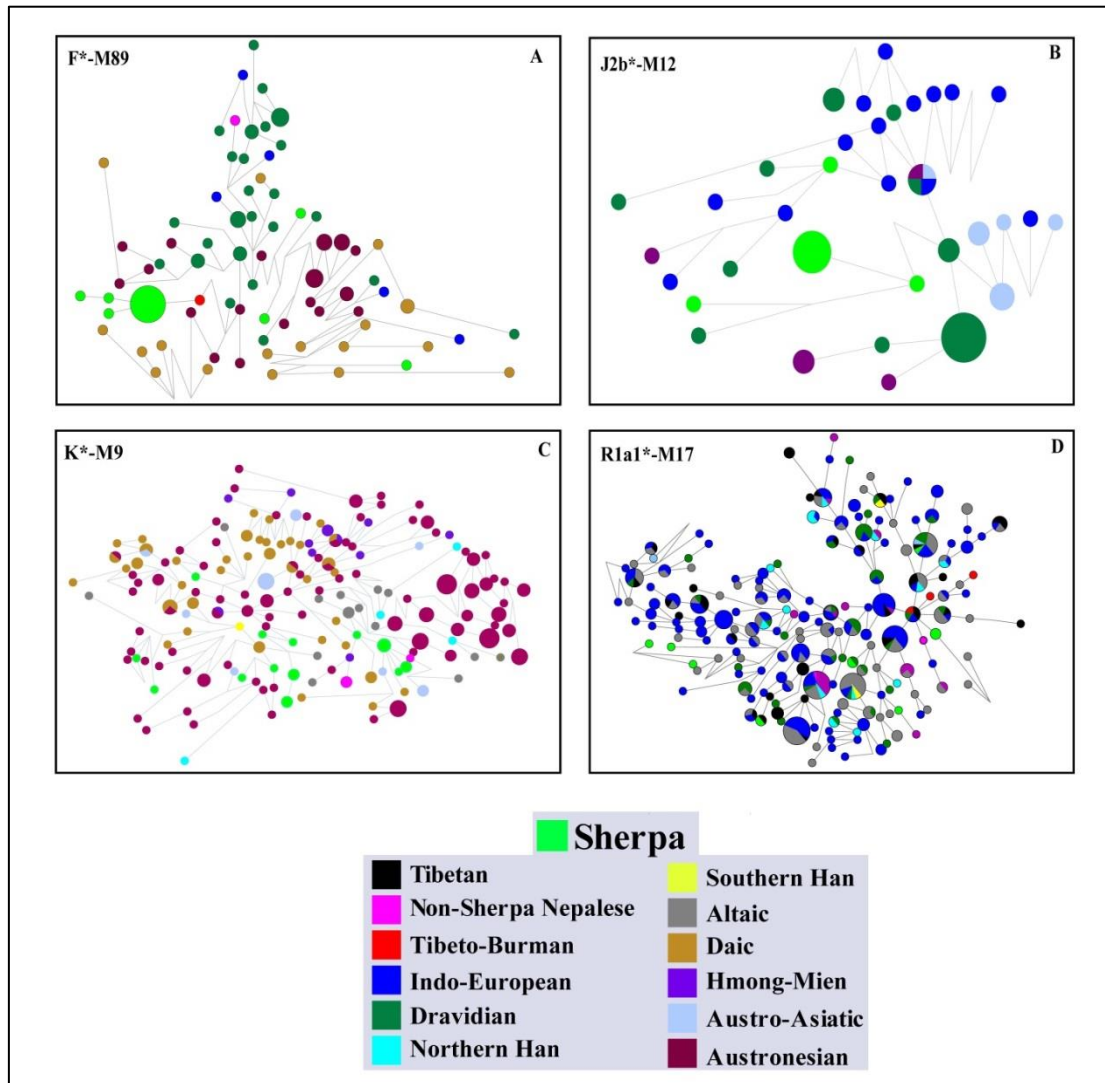
Mitochondrial DNA diversity in Sherpas

Classification of mtDNA haplogroups of 582 Sherpa samples yielded several major haplogroups including A (27.15%), M9a (24.23%), C4a (20.96%), M70 (7.22%) and D (5.84%) (Table1), which together cover 85.40% of the tested Sherpas. Among Tibetans these same groups A (14.63%), M9a (22.48%) and D (16.53%) also constituted the major haplogroups (Qi et al., 2013). M9a and its four sub-haplogroups (M9a1a, M9a1a2, and M9a1b1and M9a1a1c1b1a) are widely distributed in East Asia and Southeast Asia (M.-S. Peng et al., 2011; Soares et al., 2008), though M9a1a1c1b1a is quite rare among most Asian

populations but among Sherpas is the most common (58.16% of the total M9a individuals) and among Tibetans highly prevalent (58.06% of the total M9a individuals)(M.-S. Peng et al., 2011). The shared prevalence of M9a1a1c1b1a in Sherpas and Tibetans suggests a close maternal relationship between them. After constructing a network using 36 complete mtDNA genome sequences (34 previously published and two of present study) from 13 Sherpas and 23 Tibetans belonging to M9a1a1c1b1a, we observed that all the Sherpa individuals lie at the tip branches, suggesting they likely derived from the Tibetan core haplotype (Figure 6A). Interestingly, the star-like topology (*i.e.* a core haplotype surrounded by other haplotypes with only one or two mutation steps away from the core) suggests a relatively recent expansion of this lineage, consistent with a previous observation made on Tibetans (Qi et al., 2013). The close genetic affinity between the two populations is further supported by Haplogroup D among Sherpas (5.84%), which was also widely found in Tibetans (Qi et al., 2013). But more importantly M70—previously considered a Tibetan specific lineage (Z. Qin et al., 2010; Mian Zhao et al., 2009)—was also present among Sherpas (7.22%) (Table 1). Much like the observed pattern of the Y-haplogroups, the Sherpa mtDNA haplogroup composition bears strong similarity with that observed among Tibetans.

Interestingly, two Sherpa-specific sub-haplogroups were not detected in Tibetans or other Asian populations: A15c1 (17.27%), a sub-lineage under Haplogroup A15c, and C4a3b1 (21.82%), a sub-lineage under C4a3b (Kang et al., 2013). These two Sherpa-specific sub-haplogroups together account for nearly 40% of the maternal lineages of Sherpas. Sequencing the entire mitochondrial genomes of 79 Sherpa individuals belonging to these two sub-haplogroups showed an interesting sequence variation pattern. Consistent with the previous report (Kang et al., 2013), A15c1 defined by four mutations (T4216C, A9052G, T13111C and A15924G), and C4a3b1 by other four mutations (G3745A, 5899insC, C11155T, A13563G) only exist in Sherpas, but their mother haplogroups (A15c and C4a3b) are present

in both Tibetans and Han Chinese (Chandrasekar et al., 2009; Liu et al., 2012) (Figure 7). Further network analysis showed that A15c1 and C4a3b1 both form a star-like structure, denoting recent lineage expansion among Sherpas (Figure 6B, 6C). We estimated the coalescence ages of A15c1 (1500 years ago) and C4a3b1 (940 years ago), both of which are similar with the previous estimates (Kang et al., 2013) and consistent with the proposed recent expansion. The young ages of these two Sherpa-specific sub-haplogroups hint at a recent bottleneck during migration and latter population expansion of Sherpas, which fits our observation of a shallow divergence of the Y chromosome haplogroup D3*-P99 between Sherpas and Tibetans.

Of the other rare mtDNA haplogroups in Sherpas—Z (2.75%), F (2.58%), M13 (1.72%) and U (1.37%), as well as those with <1% frequencies (M3, M5, W, G, M10, H, M62, M11a, M38, M61 and M74)—a majority including G, Z, F, M62, M10, M11a, M13, and F have been seen among Tibetans. By contrast, the presence of the other haplogroups M5, M3, W, M38, M61, M74 and U among Sherpas may indicate minor gene flows from surrounding South Asian populations (see Table 1).

Table 1. Distribution of mtDNA haplogroups among Sherpas and Tibetans.

Populations	Haplogroup frequency (%)			
	Sherpa (Nepal)	Sherpa (Tibet)	Sherpa (all)	Tibetan(Qi et al., 2013)
Sample Size	350	232	582	6,109
A	24.00	31.90	27.15	14.63
M9a	27.14	19.83	24.22	22.48
M8	28.00	17.24	23.71	7.71
M70	11.14	1.29	7.22	0.16
D	4.57	7.76	5.84	16.53
F	1.14	4.74	2.58	11.44
M13	0.57	3.45	1.72	4.22
U	0.29	3.02	1.37	1.65
M5	0.29	1.72	0.86	0.05
W	0.57	1.29	0.86	0.05
M3	0.29	1.72	0.86	0.00
G	0.57	0.86	0.69	8.22
M10	0.29	1.29	0.69	1.06
M62	0.00	1.29	0.52	2.35
M11a	0.29	0.86	0.52	0.79
H	0.00	1.29	0.52	0.26
M38	0.29	0.43	0.34	0.00
M61	0.29	0.00	0.17	0.75
M74	0.29	0.00	0.17	0.00
B	0.00	0.00	0.00	3.76
M*	0.00	0.00	0.00	0.79
TJ	0.00	0.00	0.00	0.72
R*	0.00	0.00	0.00	0.52
N*	0.00	0.00	0.00	0.47
M7	0.00	0.00	0.00	0.25
N10	0.00	0.00	0.00	0.25
M25	0.00	0.00	0.00	0.23
M20	0.00	0.00	0.00	0.20
N9a1	0.00	0.00	0.00	0.20
N11	0.00	0.00	0.00	0.18
M49	0.00	0.00	0.00	0.07
M12	0.00	0.00	0.00	0.02
Y	0.00	0.00	0.00	0.02
M12	0.00	0.00	0.00	0.02

Figure 6. Networks of three mtDNA sub-haplogroups (A: M9a1a1c1b1a1; B: A15c1; C: C4a3b1) among Sherpas and Tibetans. The star-like networks suggest recent population expansion. The complete mtDNA genome sequences were used to construct the networks.
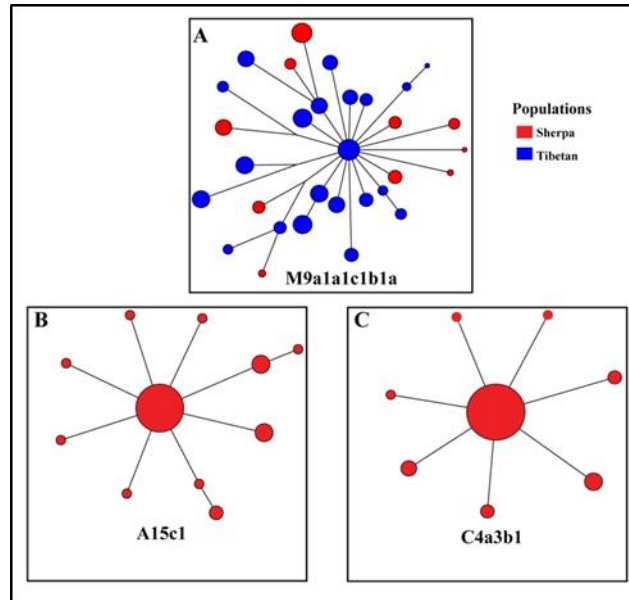
Figure 7. The mtDNA phylogenetic tree based on 215 mitochondrial whole genome sequences, including 165 Sherpas (89 from present study and 76 from previous study) and 50 non-Sherpas (44 Tibetans, 4 Han Chinese and 2 Naxi from previous studies).
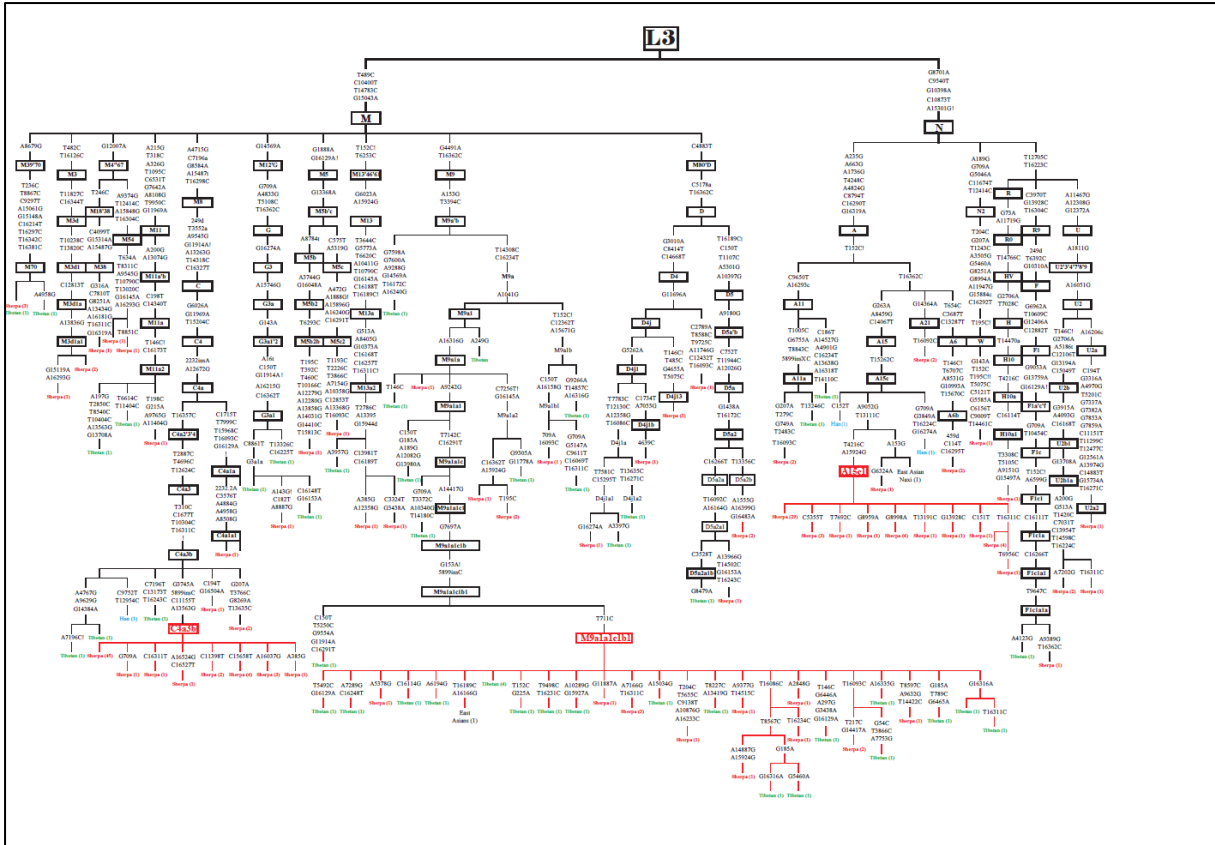
Genetic relationship of Sherpas with other Asian populations

We explored the genetic relationship between Sherpas and other Asian populations, and principal component analysis (PCA) of both mtDNA and Y chromosome haplotype frequencies displayed a close relationship between Sherpas and Tibetans (Figure 8A, 8B). In particular, the PCA map of the Y chromosome shows Sherpas mingled together with Tibetan populations and forming a separate cluster from other Asian populations (Figure 8B). The PCA map of mtDNA further places the Sherpas relatively closer to Tibetans compared to other Asian populations but form a separate cluster, implying a shallow divergence from Tibetans likely due to the recent expansions of the two Sherpa-specific mtDNA sub-haplogroups (A15c1 and C4a3b1) (Figure 8A). By contrast, non-Sherpa Nepalese are only distantly related to Sherpas, clustering with populations from the Indian subcontinent and

East Asia, suggesting limited gene flows between Sherpas and non-Sherpas in the markedly small area of Nepal nestled between India and China.

Figure 8. Maps of principal component analysis (PCA) based on mtDNA and Y-Chromosome haplogroup frequencies among Sherpas, Tibetans and other Asian populations.
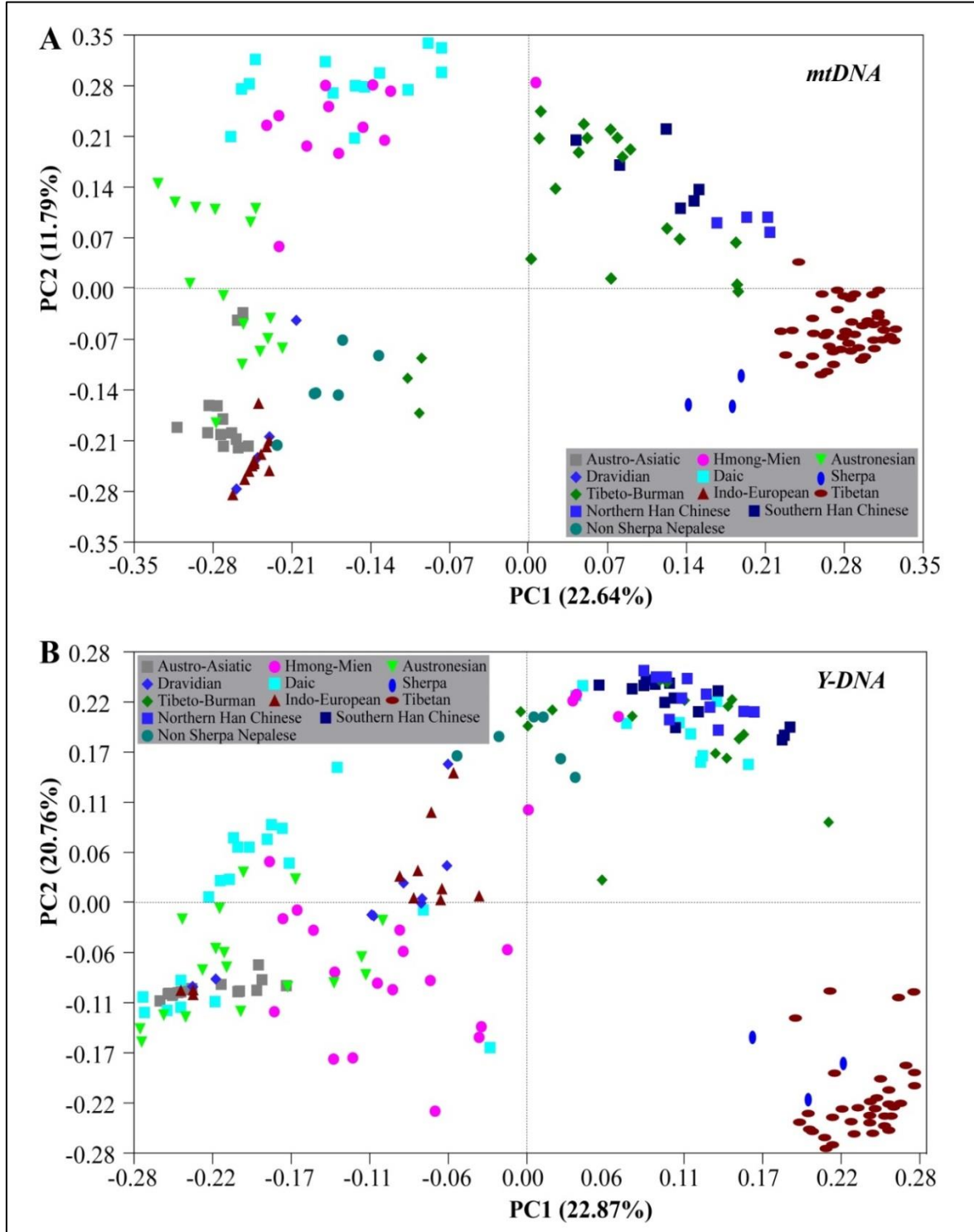
Discussions on population history of Sherpas

Our investigation of the Sherpas' origins via analysis of Y-chromosome and mtDNA diversity analysis supports a close genetic relationship between Sherpas and Tibetans. Furthermore, the detailed genetic diversity pattern revealed in the Y-STR network and the mtDNA phylogenetic trees indicate that Sherpas represent a recently derived lineage from Tibetans dated to less than 1,500 years ago. Further BSP analysis sketches a general picture of a recent bottleneck of the Sherpa population size less than 2,000 years ago (Figure 9), consistent with the proposed recent migration of the Sherpa ancestors from Tibet. These genetic findings further bolster the established linguistic affiliation among the speakers of the Tibeto-Burman language subfamily within the Sino-Tibetan family (the other subfamily is Han). Collectively, these results concur with the previous hypothesis positing the migration of the Baric branch of the Tibeto-Burman languages from the Tibetan plateau to Nepal following their dispersal from the basin of Yellow River in China(Su, Xiao, et al., 2000).

Figure 9. The BSP plot showing population dynamics of Sherpas in history. A population bottleneck around 2,000 years ago was observed. The HVS-I sequences of Sherpas were used in the BSP analysis.
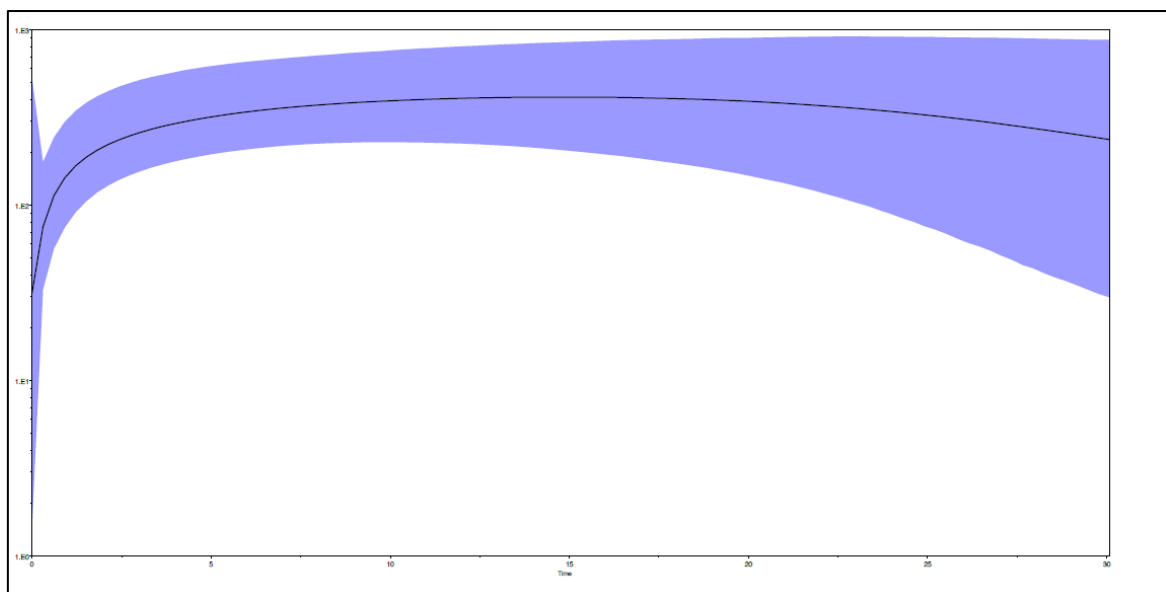
Curiously though, a previous mtDNA study of Zhangmu Sherpa reported considerable (8%-17%) South Asian genetic component among Sherpas(Kang et al., 2014), but our results only show it in a minor frequency (0.34%-2.53%). Likely, this negligible gene flow from the Indian subcontinent towards the Sherpa is due to the natural barrier of the Himalayas, that have effectively firewalled the populations in the mountainous, low-oxygen environments that comprise the Sherpa homeland (Cordaux et al., 2004; T. Gayden et al., 2007; Qi et al., 2013).

Our observation of the large amount of East Asian genetic influence on the Nepalese population (Sherpas and non-Sherpas) suggests a recent unidirectional gene flow from East Asia to Nepal, not the opposite (Gayden et al., 2013; H.-W. Wang et al., 2012). By the same token, our present data does not support the putative ancestry of the Sherpas and Han Chinese as progenitors of Tibetans (Jeong et al., 2014), and instead strongly supports that the Tibetans are ancestors of the Sherpas. While two Sherpa-specific mtDNA sub-lineages were found during our study, both are young in origin (<1500 years ago), and originated rather recently from the mother haplogroups of Tibetans, in line with recent historical records over the last 500 years ago that documents Sherpa migrations from eastern Tibet to the barren lands of Nepal's Khumbu region by crossing the Nangpa La Pass (5,716m) in the Himalayan region (Figure 3)(Oppitz, 1974).

As for the Sherpa-specific sub-haplogroups, a previous mtDNA study(Kang et al., 2013) proposed that the ND1 variants G3745A (C4a3b1) and T4216C (A15c1) may form the molecular basis of the Sherpas' adaptation to hypoxic high-altitude. After conducting a selection test by calculating the non-synonymous vs. synonymous substitution (Ka/Ks) ratio of ND1, we did not observe a signal of selection (Ka/Ks = 1.0, Nei-Gojobori model). We also conducted Tajima's D test using both the ND1 and HVS-I sequences, and no deviation from neutral expectation was detected (Tajima's D = -1.271 for ND1, P>0.1; Tajima's D = -1.168

for HVS-I, P>0.1), suggesting a neutral effect of these variants. Hence, the prevalence of the two Sherpa-specific mtDNA sub-lineages may better be explained by recent population expansion as reflected by the star-like phylogeny and their tip positions in the mtDNA phylogenetic tree (supplementary Figure S1). Accordingly, we propose that the ND1 variants may not have any direct role in hypoxic adaptation among Sherpas; indeed, our results suggest that the ancestors of Sherpas were already high altitude dwellers in the Tibetan plateau before they migrated from Tibet rather recently, which by extension predicts the same set of genes responsible for high-altitude previously identified in Tibetans (*e.g.* EPAS1 and EGLN1) should also be present among the Sherpas. In essence, the genetic adaption to high altitude in Sherpas were not likely to have developed locally in Nepal, but were instead inherent from their Tibetan ancestors who acquired the adaptive traits through their long stay (> 18,000 years)(Peng et al., 2011) in the harsh environment of the highland Tibetan plateau.

5.2   Sherpa high altitude adaptation

Population genetic analysis- genotyping of EPAS1

We re-sequenced the entire 94 kb EPAS1 gene in 50 Tibetans and identified several single nucleotide polymorphisms (SNPs) having deep divergence (F_{ST} >0.45) between Tibetans and Han Chinese in previous studies (Peng et al., 2011).We construct LD plot using those 82 SNPs which were having greater than 0.45 F_{ST} Value in between Tibetans and Han Chinese. Majority of these SNPs lies mainly in three different major blocks of EPAS1 (Figure 10).We have chosen 29 SNPs from these 82 SNPs, which probably covers different blocks of EPAS1. Since SNPs lying in the same block are tightly linked (r^2=1) to each other, we only include one or two SNPs from each haplotype block and exclude others. We genotyped these 29 EPAS1 SNPs in 50 Sherpa samples and found Sherpa have high derived allele frequencies (>49%-82%) as like in Tibetans in these SNPs.(Table 2) Interestingly, we

observed these genetic variants of EPAS1 were presence at very low frequencies(<10%) in other populations of 1000 Genomes Project(1000 Genomes Consortium, 2012) data set.

Figure 10. LD map constructed using EPAS1 SNPs which were having F_{ST} value greater than 0.45 in Tibetans.

Table 2. Allele frequency and F_{ST} value of 29 EPAS1 SNPs in different populations

Position(v37)	SNP	Allele		Frequency of derived allels			F_{ST} Sherpa			
		Derived	Ancestral	Sherpa(50)	Tibetans(50)	CHB(97)	CHB	JPT	CEU	YRI
46550132	rs116088026	C	A	0.51	0.58	0.00	0.60	0.59	0.58	0.46
46552202	rs149594770	A	T	0.54	0.51	0.01	0.62	0.61	0.61	0.56
46552352	rs140067727	C	T	0.49	0.51	0.01	0.57	0.57	0.56	0.52
46553044	rs113305133	G	A	0.51	0.58	0.01	0.59	0.59	0.58	0.53
46565091	rs12614710	T	G	0.52	0.42	0.07	0.48	0.52	0.00	0.51
46567916	rs115321619	A	G	0.67	0.74	0.01	0.74	0.73	0.71	0.45
46568680	rs73926263	G	A	0.64	0.73	0.01	0.71	0.71	0.70	0.60
46569017	rs73926264	G	A	0.67	0.73	0.01	0.74	0.73	0.73	0.63
46569770	rs73926265	A	G	0.73	0.73	0.01	0.79	0.79	0.78	0.68
46570342	rs55981512	A	G	0.75	0.73	0.01	0.81	0.80	0.80	0.75
46571017	rs149306391	G	C	0.75	0.73	0.00	0.81	0.80	0.80	0.80
46575388	rs4953354	G	A	0.76	0.79	0.12	0.67	0.62	0.58	0.56
46576918	rs76242811	C	T	0.75	0.75	0.01	0.80	0.80	0.80	0.67
46577251	rs188801636	C	T	0.74	0.75	0.01	0.80	0.80	0.79	0.79
46577299	rs6544889	G	A	0.82	0.81	0.14	0.71	0.75	0.20	0.66
46577797	SNP155	C	T	0.76	0.75	—	—	—	—	—
46583581	rs189807021	A	G	0.74	0.75	0.01	0.79	0.80	0.79	0.79
46588019	rs150877473	G	C	0.74	0.75	0.01	0.79	0.80	0.79	0.79
46588331	rs142826801	C	G	0.74	0.75	0.01	0.79	0.80	0.79	0.79
46589032	rs74898705	T	C	0.74	0.75	0.01	0.79	0.80	0.75	0.57
46592807	rs61151542	T	C	0.69	0.74	0.01	0.75	0.75	0.70	0.48
46594122	rs141366568	G	A	0.69	0.74	0.01	0.75	0.75	0.75	0.75
46597756	rs116062164	C	A	0.69	0.76	0.01	0.75	0.75	0.69	0.75
46598025	rs141426873	G	C	0.69	0.74	0.01	0.75	0.75	0.75	0.75
46600030	rs116611511	G	A	0.7	0.74	0.01	0.76	0.76	0.76	0.62
46600661	rs58160876	C	A	0.7	0.74	0.01	0.76	0.76	0.76	0.61
46600894	rs12467821	C	T	0.79	0.78	0.15	0.66	0.67	0.17	0.71
46609966	rs11690951	T	A	0.68	0.73	0.20	0.47	0.41	0.05	0.66
46615955	rs56161503	A	G	0.68	0.73	0.08	0.64	0.60	0.08	0.68

We further construct haplotype network using these 28 SNPs genotype data of Sherpa from current study and Tibetans data from previous study (Peng et al., 2011). In addition, we include data from 1000 Genomes Project for other populations: Han Chinese in Beijing, China (CHB), Japanese in Tokyo, Japan (JPT), Utah Residents with Northern and Western European Ancestry (CEU), and Yoruba in Ibadan, Nigeria (YRI) for constructing the haplotype network. We found Sherpa and Tibetans shared similar haplotype pattern for EPAS1, and which is strikingly different from haplotypes seen in Han and other populations (shown in Figure 11). Thus, the sharing of similar allele frequencies and haplotype pattern in Sherpa and Tibetans implied there might be same adaptive genetic variants of EPAS1 occurred among them as consistent with previous studies (Hanaoka et al., 2012; Jeong et al., 2014).

Figure 11. The haplotype network of EPAS1 intronic SNPs in Sherpa and other populations (CHB, CEU, JPT, YRI, Tibetans). A total of 28 SNPs located in different intronic region were used in network construction.
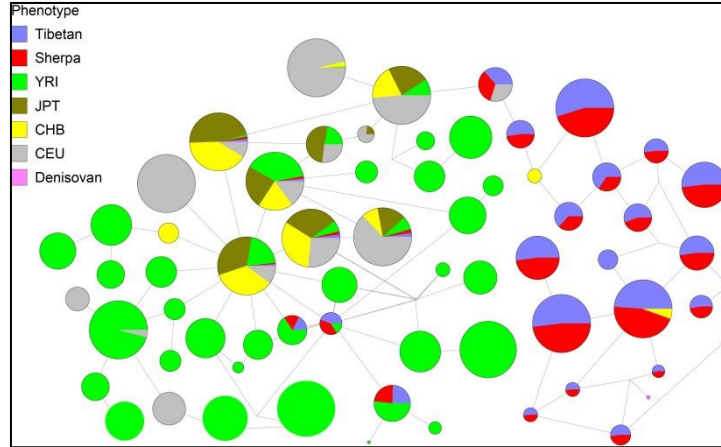
Genotyping of TED

A 3.4-kb copy number deletion at 80 kb downstream of EPAS1 gene is considered as Tibetan enriched deletion (TED) and it was initially reported by *Lou et al* (Lou et al., 2015) in 90% of Tibetans. TED is also regarded as possible candidate for HAA in Tibetans(Lou et al., 2015).We genotyped TED in 582 Sherpa samples and found TED in 93.44% of total Sherpa samples. There occurs 72.86% homozygous deletion (Zero copy) and 20.58% heterozygous deletion (one copy) in Sherpa. Intriguingly, the non-deletion carriers (two copy samples) were seen at high frequency (around 97% of total 2,792 worldwide samples) in other lowlander populations(Lou et al., 2015) but extremely rare (>10%) in Tibetans(Lou et al., 2015) and Sherpa. Thus, this high genetic divergence of deletion polymorphism in highlanders (Sherpa and Tibetans) and Lowlander population clearly suggested its prominent role for HAA.
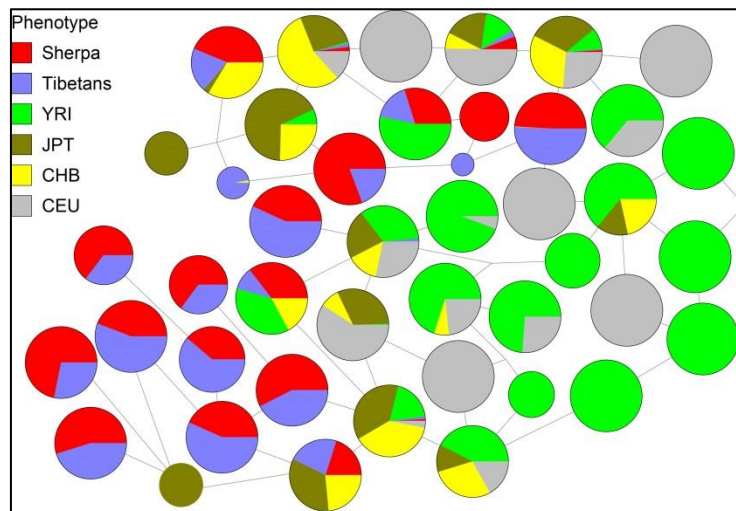
Genotyping of EGLN1

Several studies (Lorenzo et al., 2014; Xiang et al., 2013b) in EGLN1 have identified two missense mutation (rs12097901G, rs186996510C) having functional role in HAA of Tibetans. We genotyped these SNPs in 582 Sherpas and found the derived allele frequency of

rs12097901G and rs186996510C were 67% and 61% respectively. Similarly, the frequency of rs12097901G and rs186996510C in Tibetans were reported previously (Felipe R Lorenzo et al., 2014) as >80% and >63.26% respectively. There is higher frequency (>60%) of these adaptive alleles in Sherpa and Tibetans whereas other lowlander populations from 1000 Genome Project data sets have lower frequency (<20%).

We also re-sequenced the 2.5 kb upstream and 3 kb downstream region of these two SNPs (rs12097901G and rs186996510C) in 50 Sherpa samples. We used available 17 SNPs genotype information in that region to construct haplotype network in Sherpa including Tibetan data from previous studies (Xiang et al., 2013b) and other population's data (CHB, JPT, CEU, YRI) from 1000 Genomes Project. The haplotype network analysis clearly shows Sherpa and Tibetans sharing distinct haplotype of EGLN1 gene, and which is different from others population haplotype (Figure 12). Thus, Sherpa and Tibetans share similar genetic pattern in case of EGLN1 also as it was seen in EPAS1 before.

Figure.12. The haplotype network of EGLN1 gene in Sherpa and other populations (CHB, CEU, JPT, YRI, Tibetans). A total of 17 SNPs located within 5.5kb region surrounding two major adaptive EGLN1 SNPs (rs12097901G and rs186996510C) were used in network construction.

Genetic association analysis

We conducted association studies between phenotypic (hemoglobin level and degree of blood oxygen saturation level) and genotyping data of EPAS1, EGLN1 and TED in Sherpa. In case of EPAS1, we selected 11 highly differentiated SNPs from other different block of EPAS1 and genotyped it in 297 Nepalese Sherpas (126 Male and 171 Female).We found several SNPs (rs113305133G, rs116611511G and rs12467821C) having significant association (corrected p-value <0.05) with Hemoglobin in Male Sherpa samples. The significant SNPs: rs116611511G and rs12467821C) were highly linked ($0.83 < r^2 < 0.93$) with the Denisovan introgressed 2.5 kb haplotype motif (AGGAA) (Huerta-Sánchez et al., 2014).The rs12467821C, which was reported previously (Beall et al., 2010) for lowering hemoglobin concentration in Tibetans, was also found replicated in Sherpa studies. Sherpa individuals having derived allele of these significantly associated SNPs were having lower hemoglobin concentrations than individuals with ancestral allele (Figure 13). The derived allele frequencies of rs113305133G, rs116611511G and rs12467821C in 582 Sherpa are 51%, 70% and 74% respectively.

We found TED having significant association with hemoglobin concentration in both male and female Sherpa (corrected p-value <0.05). TED was showing even more significant association in lowering hemoglobin than EPAS1 intronic SNPs in Sherpas. Sherpa individuals having deletion (homozygous or heterozygous) were having lower hemoglobin concentration than non-deletion carriers. Similarly, Sherpa having derived allele of EGLN1 SNPs (rs12097901G and rs186996510C) was showing lower hemoglobin concentration than ancestral allele carrier (Figure 14).

The degree of blood oxygen saturation level was not found significantly associated with genotype data of EPAS1, EGLN1 and TED in Sherpa. Thus, lower hemoglobin trait seems

adaptive response of high altitude adaptation in Sherpas due to polygenic effect of EPAS1, EGLN1 and TED.

Table 3. Association studies of 12 EPAS1 SNPs, 2 EGLN1 SNPs and TED, with hemoglobin and blood oxygen saturation in Sherpa

	CHR	SNP	Position	Effect Allele	Hemoglobin level				Degree of blood oxygen saturation			
					BETA	SE	P	Bonf Correction	BETA	SE	P	Bonf Correction
Male N=126	2	rs149594770	46552202	A	0.6291	0.2568	0.01571	0.2042	-0.00219	0.003168	0.4903	/
	2	rs140067727	46552352	C	-0.6709	0.2745	0.01646	0.214	5.67E-05	0.003044	0.9852	/
	2	rs113305133	46553044	G	-0.9014	0.2439	0.00033	0.004293	-0.00195	0.003014	0.5179	/
	2	rs149306391	46571017	G	0.6171	0.2604	0.01938	0.252	0.002628	0.003287	0.4255	/
	2	rs4953354	46575388	G	0.4683	0.283	0.1006	/	-0.000074	0.003453	0.983	/
	2	rs188801636	46577251	C	0.6104	0.2427	0.0132	0.1716	0.000678	0.002955	0.8189	/
	2	rs6544889	46577299	G	0.5988	0.2387	0.01342	0.1744	-0.00019	0.002907	0.947	/
	2	SNP155	46577797	C	0.1959	0.1927	0.3113	/	-0.00012	0.002299	0.9594	/
	2	rs116611511	46600030	G	0.722	0.2334	0.002447	0.0318	0.000256	0.002878	0.9291	/
	2	rs58160876	46600661	C	0.6646	0.2388	0.00623	0.08099	0.000294	0.002924	0.92	/
	2	rs12467821	46600894	C	0.8058	0.2522	0.001771	0.02302	-0.00058	0.003116	0.8538	/
	2	TED	46694276	CNVs-0 copy	0.9303	0.2615	0.000537	0.006977	0.001224	0.003285	0.71	/
	1	rs186996510	231557255	C	-0.1733	0.2507	0.4908	0.9816	0.001517	0.003028	0.6173	/
	1	rs12097901	231557623	G	-0.1537	0.252	0.543	/	0.000158	0.003045	0.9587	/
Female N=171	2	rs149594770	46552202	A	0.02261	0.2132	0.9157	/	-0.00226	0.002178	0.3017	/
	2	rs140067727	46552352	C	0.08298	0.2333	0.7227	/	0.003986	0.002198	0.07154	0.9301
	2	rs113305133	46553044	G	-0.02837	0.2172	0.8962	/	0.004809	0.002194	0.02981	0.3875
	2	rs149306391	46571017	G	0.6521	0.2437	0.008187	0.1064	-0.00148	0.002562	0.5648	/
	2	rs4953354	46575388	G	-0.1278	0.2483	0.6074	/	-0.00115	0.002553	0.6538	/
	2	rs188801636	46577251	C	0.4574	0.2369	0.0552	0.7176	0.00047	0.002454	0.8484	/
	2	rs6544889	46577299	G	0.3586	0.2406	0.138	/	-0.00089	0.00248	0.7195	/
	2	SNP155	46577797	C	0.2463	0.1764	0.1646	/	-0.00129	0.001816	0.4782	/
	2	rs116611511	46600030	G	0.2857	0.2354	0.2266	/	0.000507	0.002422	0.8345	/
	2	rs58160876	46600661	C	0.3499	0.2337	0.1362	/	0.000514	0.00241	0.8315	/
	2	rs12467821	46600894	C	0.4061	0.2474	0.1026	/	0.000322	0.002555	0.8999	/
	2	TED	46694276	CNVs-0 copy	1.004	0.2757	0.000363	0.00472	-0.00103	0.002933	0.7267	/
	1	rs186996510	231557255	C	0.3234	0.1923	0.09424	0.1885	0.000198	0.002079	0.9243	/
	1	rs12097901	231557623	G	0.333	0.208	0.111	0.222	-0.001048	0.002351	0.6563	/
All N=297	2	rs149594770	46552202	A	0.2933	0.1831	0.1102	/	-0.00232	0.001837	0.2073	/
	2	rs140067727	46552352	C	-0.2705	0.1967	0.1704	/	0.002352	0.001823	0.1978	/
	2	rs113305133	46553044	G	-0.4298	0.1816	0.0186	0.2418	0.001988	0.00182	0.2754	/
	2	rs149306391	46571017	G	0.7216	0.1979	0.000316	0.004106	0.000176	0.002044	0.9315	/
	2	rs4953354	46575388	G	0.2105	0.2077	0.3117	/	-0.00094	0.002088	0.6539	/
	2	rs188801636	46577251	C	0.6738	0.1853	0.000325	0.004228	0.000105	0.001894	0.9557	/
	2	rs6544889	46577299	G	0.6295	0.1851	0.000702	0.009936	-0.00102	0.001886	0.5892	/
	2	SNP155	46577797	C	0.3379	0.1421	0.01809	0.2352	-0.0011	0.001434	0.4426	/
	2	rs116611511	46600030	G	0.6434	0.1818	0.000466	0.006051	-7.94E-05	0.001856	0.9659	/
	2	rs58160876	46600661	C	0.6399	0.1829	0.000541	0.007032	-4.15E-05	0.001867	0.9823	/
	2	rs12467821	46600894	C	0.75	0.1935	0.000131	0.001697	-0.0006	0.001983	0.7613	/
	2	TED	46694276	CNVs-0 copy	1.032	0.2088	1.32E-06	0.00001711	-0.000046	0.002188	0.9832	/
	1	rs186996510	231557255	C	0.05076	0.1656	0.7594	/	0.000849	0.001755	0.6289	/
	1	rs12097901	231557623	G	0.02928	0.1739	0.8664	/	-0.000304	0.001887	0.8721	/

Figure 13. The box plot of two missense mutation (rs12097901G, rs186996510C) of

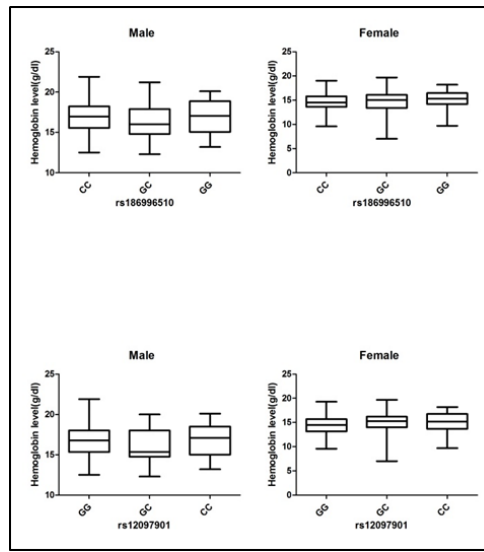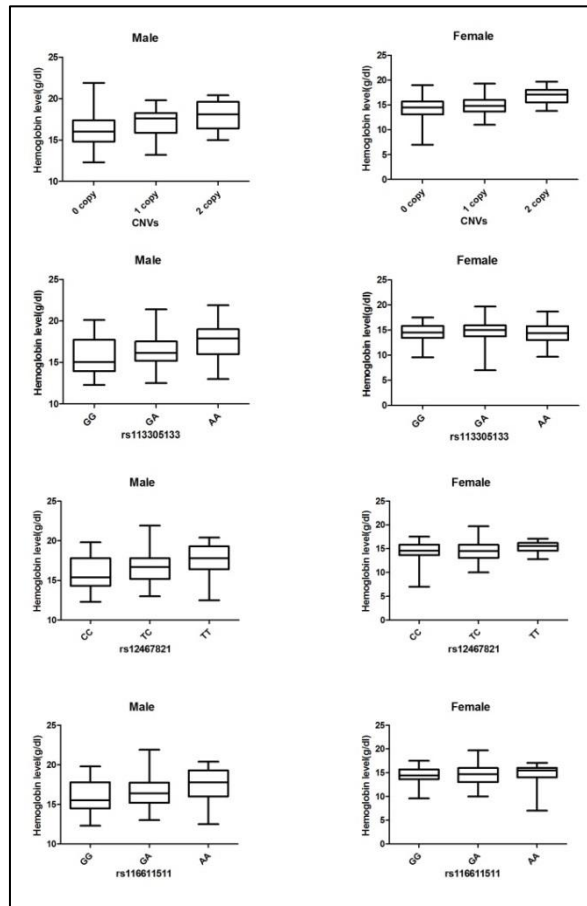EGLN1gene in Sherpa comparing with their genotype and hemoglobin level.

Figure 14.The box plot showing comparison of hemoglobin levels with CNVs (0 copy, 1

copy and 2 copy) and three different genotypes of these EPAS1 SNPs (rs113305133G,

rs116611511G and rs12467821C) in Male and Female Sherpa .

Discussion on Sherpa high altitude adaptation

The EPAS1 and EGLN1 are two top candidate genes reported in several studies (C. M. Beall et al., 2010; F. R. Lorenzo, 2010; Felipe R Lorenzo et al., 2014; Y. Peng et al., 2011; L. B. Scheinfeldt & Tishkoff, 2013; Simonson, 2010; Xiang et al., 2013b; Xu et al., 2011; Yi et al., 2010) for HAA in Tibetans. Beside these, recent studies identified TED for having higher possibilities in Tibetans adaptation (Lou et al., 2015). Since Sherpa are more closer to Tibetans (Kang et al., 2013) we expect they might have similar gene governed for adaptation in Himalayas, so we genotyped mainly EPAS1, EGLN1 and TED region in Sherpa and found similar genetic variants on these genes between these two populations. Thus, we observed these genes(EPAS1 , EGLN1 and TED) might have more pronounced effect in high altitude hypoxic Sherpa adaptation in contrary to ND1 variants G3745A (C4a3b1) and T4216C (A15c1) reported previously(Kang et al., 2013).Our comparative studies between these two highlanders populations (Sherpa and Tibetans) searched for those genetic variants shared only between them and different from other lowlanders modern human populations. We found the genetic variants of EPAS1, EGLN1 and TED are shared only among Tibetans and Sherpas suggesting these genes might be promising candidate for HAA in Himalayan plateau.

The 2.5 kb EPAS1 haplotype motif (AGGAA) in Tibetans was recently proposed to occur as the result of Denisovan introgression. Sherpa also contain similar haplotype as it was seen in Tibetans and they might also have gained it from Denisovan gene flow but it is difficult to pinpoint that this 2.5 kb introgressed haplotype motif of EPAS1 was the adaptive haplotype for Tibetan adaptation. Since, there are around 90 SNPs located in different intronic region of EPAS1 having F_{ST} value greater than 0.40 between Tibetans and Han Chinese. Moreover, the absence of 3.4-kb copy number deletion at 80 kb downstream of EPAS1 gene and adaptive EGLN1 haplotype in Denisovan suggested Denisovan introgression was not only the sole mechanism for Tibetan adaptation. The TED and EGLN1

adaptive haplotype in Tibetans and Sherpa might have occurred due to ongoing natural selection after the introgression. Interestingly, we noticed TED was more significant than EPAS1 intronic SNPs for lowering hemoglobin concentration in both Male and Female Sherpa. However, association studies just give us a hint; we need further functional studies to know the role of TED and EPAS1 SNPs for HAA in Sherpa. The functional studies of EGLN1 missense mutation already shows these SNPs were responsible for reducing the erythropoietic response to hypoxia (Lorenzo et al., 2014). Since Sherpa also contain this missense mutation at high frequency (>60%) they are also protected from polycythemia at high altitude as it was seen in Tibetans (Lorenzo et al., 2014).

Although it is difficult to define specific high altitude hypoxic adaptation pathway, it is indeed higher possibilities that polygenic (EPAS1, EGLN1 and TED) effect might have occurred for producing high altitude phenotype in Sherpas and these genetic variations in Sherpa might have occurred due to positive selection at high Himalayas. However, further functional studies are needed to reveal detail molecular mechanism for high altitude adaptation.

5. CONCLUSION

Our mtDNA and Y-chromosome analyses of Sherpas indicate that Tibetans were the ancestors of Sherpas. So, Sherpas inherited high altitude adaptive features from their ancestors (Tibetans) who have long history of staying on the Tibetan plateau. The long term colonization (around 30,000 years ago) of Tibetans has undergone positive selection mainly in EPAS1, EGLN1 and TED. We tested these genes in Sherpa and found similar genetic variants in Sherpa as it was in Tibetans which occurred due to inheritance from their Tibetan ancestors. We found that the Sherpa-specific mtDNA haplotype originated due to population expansion after the ancestors of Sherpa migrated to Nepal before 1500 years ago and it has no role in high altitude hypoxic adaptation. The similar adaptive traits of lower hemoglobin level seen between Sherpa and Tibetans further supports the occurrence of similar genes for high altitude hypoxic adaptation. Hence, Sherpa is a recently derived Tibetan population containing similar genes with Tibetans for adaptation at the Himalayas.

REFERENCES

Adams, W. H., & Shresta, S. M. (1974). Hemoglobin levels, vitamin B12, and folate status in a Himalayan village. *The American journal of clinical nutrition, 27*(2), 217-219.

Adams, W. H., & Strang, L. J. (1975). Hemoglobin levels in persons of Tibetan ancestry living at high altitude. *Experimental Biology and Medicine, 149*(4), 1036-1039.

Aldenderfer, M. S. (2003). Moving Up in the World Archaeologists seek to understand how and when people came to occupy the Andean and Tibetan plateaus. *Am Sci, 91*(6), 542-550.

Alkorta-Aranburu, G., Beall, C. M., Witonsky, D. B., Gebremedhin, A., Pritchard, J. K., & Di Rienzo, A. (2012). The genetic architecture of adaptations to high altitude in Ethiopia. *PLoS genetics, 8*(12), e1003110.

Allen, P., Matheson, G., Zhu, G., Gheorgiu, D., Dunlop, R., Falconer, T., . . . Hochachka, P. (1997). Simultaneous 31P MRS of the soleus and gastrocnemius in Sherpas during graded calf muscle exercise. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 273*(3), R999-R1007.

Anderson, S., Bankier, A. T., Barrell, B. G., De Bruijn, M., Coulson, A. R., Drouin, J., . . . Sanger, F. (1981). Sequence and organization of the human mitochondrial genome.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetics, 23*(2), 147-147.

Aquadro, C. F., & Greenberg, B. D. (1983). Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics, 103*(2), 287-312.

Bandelt, H.-J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution, 16*(1), 37-48.

Bandelt, H.-J., Richards, M., & Macaulay, V. (2006). *Human mitochondrial DNA and the evolution of Homo sapiens* (Vol. 18): Springer Science & Business Media.

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics, 21*(2), 263-265.

Bastien, G. J., Schepens, B., Willems, P. A., & Heglund, N. C. (2005). Energetics of load carrying in Nepalese porters. *Science, 308*(5729), 1755-1755.

Beall, C., & Reichsman, A. (1984). Hemoglobin levels in a Himalayan high altitude population. *American journal of physical anthropology, 63*(3), 301-306.

Beall, C. M. (2007a). Detecting natural selection in high-altitude human populations. *Respiratory physiology & neurobiology, 158*(2), 161-171.

Beall, C. M. (2007b). Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proceedings of the National Academy of Sciences, 104*(Suppl 1), 8655-8660.

Beall, C. M., Brittenham, G. M., Macuaga, F., & Barragan, M. (1990). Variation in hemoglobin concentration among samples of high-altitude natives in the Andes and the Himalayas. *American Journal of Human Biology, 2*(6), 639-651.

Beall, C. M., Cavalleri, G. L., Deng, L., Elston, R. C., Gao, Y., Knight, J., . . . McCormack, M. (2010). Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders. *Proceedings of the National Academy of Sciences, 107*(25), 11459-11464.

Beall, C. M., Laskowski, D., & Erzurum, S. C. (2012). Nitric oxide in adaptation to altitude. *Free Radical Biology and Medicine, 52*(7), 1123-1134.

Beall, C. M., Song, K., Elston, R. C., & Goldstein, M. C. (2004). Higher offspring survival among Tibetan women with high oxygen saturation genotypes residing at 4,000 m. *Proceedings of the National Academy of Sciences of the United States of America, 101*(39), 14300-14304.

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J. M., Mei, R., . . . Herráez, D. L. (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS genetics, 6*(9), e1001116.

Bigham, A. W., Mao, X., Mei, R., Brutsaert, T., Wilson, M. J., Julian, C. G., . . . Shriver, M. D. (2009). Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. *Human genomics, 4*(2), 79.

Birky, C. W., Demko, C. A., Perlman, P. S., & Strausberg, R. (1978). Uniparental inheritance of mitochondrial genes in yeast: dependence on input bias of mitochondrial DNA and preliminary investigations of the mechanism. *Genetics, 89*(4), 615-651.

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research, 27*(8), 1767-1780.

Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., . . . Atkinson, Q. D. (2012). Mapping the origins and expansion of the Indo-European language family. *Science, 337*(6097), 957-960.

Boyer, S. J., & Blume, F. D. (1984). Weight loss and changes in body composition at high altitude. *Journal of Applied Physiology, 57*(5), 1580-1585.

Breton, S., Beaupre, H. D., Stewart, D. T., Hoeh, W. R., & Blier, P. U. (2007). The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? *Trends in Genetics, 23*(9), 465-474.

Brooks, G., Butterfield, G., Wolfe, R., Groves, B., Mazzeo, R., Sutton, J., . . . Reeves, J. (1991). Increased dependence on blood glucose after acclimatization to 4,300 m. *Journal of Applied Physiology, 70*(2), 919-927.

Brown, W. M., George, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences, 76*(4), 1967-1971.

Brutsaert, T. D. (2008). Do high-altitude natives have enhanced exercise performance at altitude? *Applied Physiology, Nutrition, and Metabolism, 33*(3), 582-592.

Butler, J. M., Schoske, R., Vallone, P. M., Kline, M. C., Redd, A. J., & Hammer, M. F. (2002). A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Science International, 129*(1), 10-24.

Butterfield, G. E., Gates, J., Fleming, S., Brooks, G. A., Sutton, J. R., & Reeves, J. T. (1992). Increased energy intake minimizes weight loss in men at high altitude. *Journal of Applied Physiology, 72*(5), 1741-1748.

Cann, R., & Wilson, A. (1983). Length mutations in human mitochondrial DNA. *Genetics, 104*(4), 699-711.

Cann, R. L. (1994). mtDNA and Native Americans: a Southern perspective. *American journal of human genetics, 55*(1), 7.

Cann, R. L. (2001). Genetic clues to dispersal in human populations: retracing the past from the present. *Science, 291*(5509), 1742-1748.

Cann, R. L., Brown, W. M., & Wilson, A. C. (1984). Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics, 106*(3), 479-499.

CAVALLI-SFORZA, L. (1995). The History and Geography of Human Genes. Princeton University et al: Press.

Cavalli-Sforza, L. L. (1996). The spread of agriculture and nomadic pastoralism: Insights from genetics, linguistics, and archaeology. *The origins and spread of agriculture and pastoralism in Eurasia*, 51-69.

Chandrasekar, A., Kumar, S., Sreenath, J., Sarkar, B. N., Urade, B. P., Mallick, S., . . . Basu, D. (2009). Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PLoS One, 4*(10), e7447.

Chaubey, G., Singh, M., Crivellaro, F., Tamang, R., Nandan, A., Singh, K., . . . Sharma, V. (2014). Unravelling the distinct strains of Tharu ancestry. *European Journal of Human Genetics*.

Chene, P., Cechowska-Pasko, M., & Bankowski, E. (2006). The effect of hypoxia on the expression of 150 kDa oxygen-regulated protein (ORP 150) in HeLa cells. *Cellular Physiology and Biochemistry, 17*(1-2), 89-96.

Consortium, G. P. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature, 491*(7422), 56-65.

Coop, G., Pickrell, J. K., Novembre, J., Kudaravalli, S., Li, J., Absher, D., . . . Pritchard, J. K. (2009). The role of geography in human adaptation. *PLoS Genet, 5*(6), e1000500.

Cordaux, R., Aunger, R., Bentley, G., Nasidze, I., Sirajuddin, S. M., & Stoneking, M. (2004). Independent origins of Indian caste and tribal paternal lineages. *Current Biology, 14*(3), 231-235. doi: 10.1016/j.cub.2004.01.024

Curran, L. S., Zhuang, J., Droma, T., & Moore, L. G. (1998). Superior exercise performance in lifelong Tibetan residents of 4,400 m compared with Tibetan residents of 3,658 m. *American journal of physical anthropology, 105*(1), 21-31.

Dames, S., Chou, L. S., Xiao, Y., Wayman, T., Stocks, J., Singleton, M., . . . Mao, R. (2013). The development of next-generation sequencing assays for the mitochondrial genome and 108 nuclear genes associated with mitochondrial disorders. *J Mol Diagn, 15*(4), 526-534. doi: 10.1016/j.jmoldx.2013.03.005

Daub, J. T., Hofer, T., Cutivet, E., Dupanloup, I., Quintana-Murci, L., Robinson-Rechavi, M., & Excoffier, L. (2013). Evidence for polygenic adaptation to pathogens in the human genome. *Molecular biology and evolution*, mst080.

Denko, N. C. (2008). Hypoxia, HIF1 and glucose metabolism in the solid tumour. *Nature Reviews Cancer, 8*(9), 705-713.

Droma, Y., Hanaoka, M., Basnyat, B., Arjyal, A., Neupane, P., Pandit, A., . . . Katsuyama, Y. (2008). Adaptation to high altitude in Sherpas: association with the insertion/deletion polymorphism in the Angiotensin-converting enzyme gene. *Wilderness & environmental medicine, 19*(1), 22-29.

Droma, Y., Hanaoka, M., Basnyat, B., Arjyal, A., Neupane, P., Pandit, A., . . . Katsuyama, Y. (2006). Genetic contribution of the endothelial nitric oxide synthase gene to high altitude adaptation in sherpas. *High altitude medicine & biology, 7*(3), 209-220.

Drummond, A. J., Nicholls, G. K., Rodrigo, A. G., & Solomon, W. (2002). Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics, 161*(3), 1307-1320.

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular biology and evolution, 22*(5), 1185-1192.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution, 29*(8), 1969-1973.

Endicott, P., Ho, S. Y., Metspalu, M., & Stringer, C. (2009). Evaluating the mitochondrial timescale of human evolution. *Trends Ecol Evol, 24*(9), 515-521.

Erzurum, S., Ghosh, S., Janocha, A., Xu, W., Bauer, S., Bryan, N., . . . Stuehr, D. (2007). Higher blood flow and circulating NO products offset high-altitude hypoxia among Tibetans. *Proceedings of the National Academy of Sciences, 104*(45), 17593-17598.

Foll, M., Gaggiotti, O. E., Daub, J. T., Vatsiou, A., & Excoffier, L. (2014). Widespread signals of convergent adaptation to high altitude in Asia and america. *Am J Hum Genet, 95*(4), 394-407. doi: 10.1016/j.ajhg.2014.09.002

Formenti, F., Constantin-Teodosiu, D., Emmanuel, Y., Cheeseman, J., Dorrington, K. L., Edwards, L. M., . . . McNamara, C. J. (2010). Regulation of human metabolism by hypoxia-inducible factor. *Proceedings of the National Academy of Sciences, 107*(28), 12722-12727.

Fornarino, S., Pala, M., Battaglia, V., Maranta, R., Achilli, A., Modiano, G., . . . Santachiara-Benerecetti, S. A. (2009). Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC evolutionary biology, 9*(1), 154.

Forster, P., Harding, R., Torroni, A., & Bandelt, H.-J. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. *American journal of human genetics, 59*(4), 935.

Forsythe, J. A., Jiang, B. H., Iyer, N. V., Agani, F., Leung, S. W., Koos, R. D., & Semenza, G. L. (1996). Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1. *Molecular and Cellular Biology, 16*(9), 4604-4613.

Garrido, E., Rodas, G., Javierre, C., Segura, R., Estruch, A., & Ventura, J. L. (1997). Cardiorespiratory response to exercise in elite Sherpa climber transferred to sea level. *Medicine and science in sports and exercise, 29*(7), 937-942.

Garrido, E., Segura, R., Capdevila, A., Pujol, J., Javierre, C., & Ventura, J. L. (1996). Are Himalayan Sherpas better protected against brain damage associated with extreme altitude climbs? *Clinical Science, 90*(1), 81.

Gayden, T., Cadenas, A. M., Regueiro, M., Singh, N. B., Zhivotovsky, L. A., Underhill, P. A., . . . Herrera, R. J. (2007). The Himalayas as a directional barrier to gene flow. *The American Journal of Human Genetics, 80*(5), 884-894.

Gayden, T., Cadenas, A. M., Regueiro, M., Singh, N. B., Zhivotovsky, L. A., Underhill, P. A., . . . Herrera, R. J. (2007). The Himalayas as a directional barrier to gene flow. *American Journal of Human Genetics, 80*(5), 884-894. doi: 10.1086/516757

Gayden, T., Chennakrishnaiah, S., La Salvia, J., Jimenez, S., Regueiro, M., Maloney, T., . . . Stojkovic, O. (2011). Y-STR diversity in the Himalayas. *International journal of legal medicine, 125*(3), 367-375.

Gayden, T., Mirabal, S., Cadenas, A. M., Lacau, H., Simms, T. M., Morlote, D., . . . Herrera, R. J. (2009). Genetic insights into the origins of Tibeto-Burman populations in the Himalayas. *Journal of human genetics, 54*(4), 216-223.

Gayden, T., Perez, A., Persad, P. J., Bukhari, A., Chennakrishnaiah, S., Simms, T., . . . Herrera, R. J. (2013). The Himalayas: Barrier and conduit for gene flow. *American journal of physical anthropology, 151*(2), 169-182.

Ge, R.-L., Simonson, T. S., Cooksey, R. C., Tanna, U., Qin, G., Huff, C. D., . . . Prchal, J. T. (2012). Metabolic insight into mechanisms of high-altitude adaptation in Tibetans. *Molecular genetics and metabolism, 106*(2), 244-247.

Ge, R. L., Chen, Q.-H., Wang, L.-H., Gen, D., Yang, P., Kubo, K., . . . Takeoka, M. (1994). Higher exercise performance and lower VO2max in Tibetan than Han residents at 4,700 m altitude. *Journal of Applied Physiology, 77*(2), 684-691.

Gelfi, C., De Palma, S., Ripamonti, M., EBERINI, I., WAIT, R., Bajracharya, A., . . . Cerretelli, P. (2004). New aspects of altitude adaptation in Tibetans: a proteomic approach. *The FASEB Journal, 18*(3), 612-614.

Gilbert-Kawai, E. T., Milledge, J. S., Grocott, M. P., & Martin, D. S. (2014). King of the mountains: Tibetan and Sherpa physiological adaptations for life at high altitude. *Physiology, 29*(6), 388-402.

Giles, R. E., Blanc, H., Cann, H. M., & Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences, 77*(11), 6715-6719.

Gill, M., & Pugh, L. (1964). Basal metabolism and respiration in men living at 5,800 m (19,000 ft). *Journal of Applied Physiology, 19*(5), 949-954.

gounder Palanichamy, M., Sun, C., Agrawal, S., Bandelt, H.-J., Kong, Q.-P., Khan, F., . . . Zhang, Y.-P. (2004). Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *The American Journal of Human Genetics, 75*(6), 966-978.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., . . . Fritz, M. H.-Y. (2010). A draft sequence of the Neandertal genome. *Science, 328*(5979), 710-722.

Greenberg, B. D., Newbold, J. E., & Sugino, A. (1983). Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene, 21*(1), 33-49.

Gubin, A. N., & Miller, J. L. (2001). Human erythroid porphobilinogen deaminase exists in 2 splice variants. *Blood, 97*(3), 815-817.

Gupta, R., & Basu, A. (1991). Altitude and growth among the Sherpas of the eastern Himalayas. *American Journal of Human Biology, 3*(1), 1-9.

Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B., & Keightley, P. D. (2008). Direct estimation of the mitochondrial DNA mutation rate in Drosophila melanogaster. *PLoS Biol, 6*(8), e204.

Hackett, P. H., Reeves, J. T., Reeves, C. D., Grover, R. F., & Rennie, D. (1980). Control of breathing in Sherpas at low and high altitude. *Journal of Applied Physiology, 49*(3), 374-379.

Hammer, M. F., Karafet, T., Rasanayagam, A., Wood, E. T., Altheide, T. K., Jenkins, T., . . . Zegura, S. L. (1998). Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Molecular biology and evolution, 15*(4), 427-441.

Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M., & Horai, S. (2006). Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *Journal of human genetics, 51*(1), 47-58.

Hanaoka, M., Droma, Y., Basnyat, B., Ito, M., Kobayashi, N., Katsuyama, Y., . . . Ota, M. (2012). Genetic variants in EPAS1 contribute to adaptation to high-altitude hypoxia in Sherpas. *PLoS One, 7*(12), e50566.

Havryk, A. P., Gilbert, M., & Burgess, K. R. (2002). Spirometry values in Himalayan high altitude residents (Sherpas). *Respiratory physiology & neurobiology, 132*(2), 223-232.

<center>References</center>

Henn, B. M., Cavalli-Sforza, L. L., & Feldman, M. W. (2012). The great human expansion. *Proceedings of the National Academy of Sciences, 109*(44), 17758-17764.

Henn, B. M., Gignoux, C. R., Jobin, M., Granka, J. M., Macpherson, J. M., Kidd, J. M., . . . Feldman, M. W. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci U S A, 108*(13), 5154-5162. doi: 10.1073/pnas.1017511108

Hochachka, P., Beatty, C., Burelle, Y., Trump, M., McKenzie, D., & Matheson, G. (2002). The lactate paradox in human high-altitude physiological performance. *Physiology, 17*(3), 122-126.

Hochachka, P., Clark, C., Holden, J., Stanley, C., Ugurbil, K., & Menon, R. (1996). 31P magnetic resonance spectroscopy of the Sherpa heart: a phosphocreatine/adenosine triphosphate signature of metabolic defense against hypobaric hypoxia. *Proceedings of the National Academy of Sciences, 93*(3), 1215-1220.

Holden, J., Stone, C., Clark, C., Brown, W., Nickles, R., Stanley, C., & Hochachka, P. (1995). Enhanced cardiac metabolism of plasma glucose in high-altitude natives: adaptation against chronic hypoxia. *Journal of Applied Physiology, 79*(1), 222-228.

Holloway, C. J., Montgomery, H. E., Murray, A. J., Cochlin, L. E., Codreanu, I., Hopwood, N., . . . Tyler, D. J. (2011). Cardiac response to hypobaric hypoxia: persistent changes in cardiac mass, function, and energy metabolism after a trek to Mt. Everest Base Camp. *The FASEB Journal, 25*(2), 792-796.

Hoppeler, H., & Vogt, M. (2001). Muscle tissue adaptations to hypoxia. *Journal of Experimental Biology, 204*(18), 3133-3139.

Hoppeler, H., Vogt, M., Weibel, E. R., & Flück, M. (2003). Response of skeletal muscle mitochondria to hypoxia. *Experimental Physiology, 88*(1), 109-119.

Hornbein, T. F., Townes, B. D., Schoene, R. B., Sutton, J. R., & Houston, C. S. (1989). The cost to the central nervous system of climbing to extremely high altitude. *New England Journal of Medicine, 321*(25), 1714-1719.

Huerta-Sánchez, E., DeGiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., . . . Nielsen, R. (2013). Genetic Signatures Reveal High-Altitude Adaptation in a Set of Ethiopian Populations. *Molecular biology and evolution, 30*(8), 1877-1888. doi: 10.1093/molbev/mst089

Huerta-Sánchez, E., Jin, X., Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., . . . Ni, P. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature, 512*(7513), 194-197.

Hughes, J. F., Skaletsky, H., Pyntikova, T., Graves, T. A., van Daalen, S. K., Minx, P. J., . . . Friedman, C. (2010). Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature, 463*(7280), 536-539.

Ingman, M., Kaessmann, H., PaÈaÈbo, S., & Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature, 408*(6813), 708-713.

Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., . . . Kaelin Jr, W. G. (2001). HIFα targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science, 292*(5516), 464-468.

Jablonski, N. G., & Chaplin, G. (2000). The evolution of human skin coloration. *Journal of human evolution, 39*(1), 57-106.

Jablonski, N. G., & Chaplin, G. (2010). Colloquium paper: human skin pigmentation as an adaptation to UV radiation. *Proc Natl Acad Sci U S A, 107 Suppl 2*, 8962-8968. doi: 10.1073/pnas.0914628107

Jansen, G. F., Krins, A., Basnyat, B., Bosch, A., & Odoom, J. A. (2000). Cerebral autoregulation in subjects adapted and not adapted to high altitude. *Stroke, 31*(10), 2314-2318.

Jansen, G. F., Krins, A., Basnyat, B., Odoom, J. A., & Ince, C. (2007). Role of the altitude level on cerebral autoregulation in residents at high altitude. *Journal of Applied Physiology, 103*(2), 518-523.

Jazin, E., Soodyall, H., Jalonen, P., Lindholm, E., Stoneking, M., & Gyllensten, U. (1998). Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nature genetics, 18*(2), 109-110.

Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D. B., Pritchard, J. K., . . . Di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. *Nature communications, 5*.

Jobgen, W. S., Fried, S. K., Fu, W. J., Meininger, C. J., & Wu, G. (2006). Regulatory role for the arginine–nitric oxide pathway in metabolism of energy substrates. *The Journal of nutritional biochemistry, 17*(9), 571-588.

Jobling, M. A., & Tyler-Smith, C. (1995). Fathers and sons: the Y chromosome and human evolution. *Trends in Genetics, 11*(11), 449-456.

Jobling, M. A., & Tyler-Smith, C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics, 4*(8), 598-612.

Julian, C. G., Wilson, M. J., Lopez, M., Yamashiro, H., Tellez, W., Rodriguez, A., . . . Vargas, E. (2009). Augmented uterine artery blood flow and oxygen delivery protect Andeans

from altitude-associated reductions in fetal growth. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 296*(5), R1564-R1575.

Kang, L., Wang, C.-C., Chen, F., Yao, D., Jin, L., & Li, H. (2014). Northward genetic penetration across the Himalayas viewed from Sherpa people. *Mitochondrial DNA*(0), 1-8.

Kang, L., Zheng, H.-X., Chen, F., Yan, S., Liu, K., Qin, Z., . . . Wang, X. (2013). mtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. *Molecular biology and evolution*, mst147.

Kansakar, T. R. (1996). Multilingualism and the language situation in Nepal. *Linguistics of the Tibeto-Burman Area, 19*(2), 17-30.

Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., & Hammer, M. F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome research, 18*(5), 830-838.

Kayser, B., Hoppeler, H., Claassen, H., & Cerretelli, P. (1991). Muscle structure and performance capacity of Himalayan Sherpas. *Journal of Applied Physiology, 70*(5), 1938-1942.

Kayser, B., Marconi, C., Amatya, T., Basnyat, B., Colombini, A., Broers, B., & Cerretelli, P. (1994). The metabolic and ventilatory response to exercise in Tibetans born at low altitude. *Respiration physiology, 98*(1), 15-26.

Kennedy, S. L., Stanley, W. C., Panchal, A. R., & Mazzeo, R. S. (2001). Alterations in enzymes involved in fat metabolism after acute and chronic altitude exposure. *Journal of Applied Physiology, 90*(1), 17-22.

Keyl, C., Schneider, A., Greene, R. E., Passino, C., Spadacini, G., Bandinelli, G., . . . Bernardi, L. (2000). Effects of breathing control on cardiocirculatory modulation in Caucasian lowlanders and Himalayan Sherpas. *European journal of applied physiology, 83*(6), 481-486.

Kim, E. J., Yoo, Y. G., Yang, W. K., Lim, Y. S., Na, T. Y., Lee, I. K., & Lee, M. O. (2008). Transcriptional activation of HIF-1 by RORalpha and its role in hypoxia signaling. *Arterioscler Thromb Vasc Biol, 28*(10), 1796-1802. doi: 10.1161/ATVBAHA.108.171546

Kivisild, T., Bamshad, M. J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., . . . Dixon, M. E. (1999). Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Current Biology, 9*(22), 1331-1334.

Kivisild, T., Kaldma, K., Metspalu, M., Parik, J., Papiha, S., & Villems, R. (1999). The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World *Genomic diversity* (pp. 135-152): Springer.

Kivisild, T., Shen, P., Wall, D. P., Do, B., Sung, R., Davis, K., . . . Torroni, A. (2006). The role of selection in the evolution of human mitochondrial genomes. *Genetics, 172*(1), 373-387.

Kivisild, T., Tolk, H.-V., Parik, J., Wang, Y., Papiha, S. S., Bandelt, H.-J., & Villems, R. (2002). The emerging limbs and twigs of the East Asian mtDNA tree. *Molecular biology and evolution, 19*(10), 1737-1751.

Kong, Q.-P., Sun, C., Wang, H.-W., Zhao, M., Wang, W.-Z., Zhong, L., . . . Cheng, Y.-T. (2011). Large-scale mtDNA screening reveals a surprising matrilineal complexity in east Asia and its implications to the peopling of the region. *Molecular biology and evolution, 28*(1), 513-522.

Kraaijenbrink, T., Parkin, E. J., Carvalho-Silva, D. R., van Driem, G., Barbujani, G., Tyler-Smith, C., . . . de Knijff, P. (2009). Genetic and linguistic borders in the Himalayan Region. *Becoming Eloquent: Advances in the Emergence of Language, Human Cognition and Modern Cultures*, 181-201.

Lahiri, S., & Milledge, J. (1967). Acid-base in Sherpa altitude residents and lowlanders at 4880 m. *Respiration physiology, 2*(3), 323-334.

Lahiri, S., Milledge, J., Chattopadhyay, H., Bhattacharyya, A., & Sinha, A. (1967). Respiration and heart rate of Sherpa highlanders during exercise. *Journal of Applied Physiology, 23*(4), 545-554.

Lewis, N., Bailey, D. M., duManoir, G. R., Messinger, L., Lucas, S. J., Cotter, J. D., . . . Stembridge, M. (2014). Conduit artery structure and function in lowlanders and native highlanders: relationships with oxidative stress and role of sympathoexcitation. *The Journal of physiology, 592*(5), 1009-1024.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics, 25*(11), 1451-1452. doi: 10.1093/bioinformatics/btp187

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics, 25*(11), 1451-1452.

Liu, J., Wang, L.-D., Sun, Y.-B., Li, E.-M., Xu, L.-Y., Zhang, Y.-P., . . . Kong, Q.-P. (2012). Deciphering the signature of selective constraints on cancerous mitochondrial genome. *Molecular biology and evolution, 29*(4), 1255-1261.

Lorenzo, F. R. (2010). Novel PHD2 mutation associated with Tibetan genetic adaptation to high altitude hypoxia. *ASH 52nd Annual Meeting*.

Lorenzo, F. R., Huff, C., Myllymäki, M., Olenchock, B., Swierczek, S., Tashi, T., . . . McClain, D. A. (2014). A genetic mechanism for Tibetan high-altitude adaptation. *Nature genetics*.

Lou, H., Lu, Y., Lu, D., Fu, R., Wang, X., Feng, Q., . . . Kang, L. (2015). A 3.4-kb Copy-Number Deletion near EPAS1 Is Significantly Enriched in High-Altitude Tibetans but Absent from the Denisovan Sequence. *The American Journal of Human Genetics, 97*(1), 54-66.

Lutz, S., Weisser, H.-J., Heizmann, J., & Pollak, S. (1998). Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *International journal of legal medicine, 111*(2), 67-77.

Macaulay, V., Richards, M., & Sykes, B. (1999). Mitochondrial DNA recombination-no need to panic. *Proceedings of the Royal Society of London B: Biological Sciences, 266*(1433), 2037-2039.

Malacrida, S., Katsuyama, Y., Droma, Y., Basnyat, B., Angelini, C., Ota, M., & Danieli, G. (2007). Association between human polymorphic DNA markers and hypoxia adaptation in Sherpa detected by a preliminary genome scan. *Annals of human genetics, 71*(5), 630-638.

Marconi, C., Marzorati, M., Grassi, B., Basnyat, B., Colombini, A., Kayser, B., & Cerretelli, P. (2004). Second generation Tibetan lowlanders acclimatize to high altitude more quickly than Caucasians. *The Journal of physiology, 556*(2), 661-671.

Masuyama, S., Kimura, H., Sugita, T., Kuriyama, T., Tatsumi, K., Kunitomo, F., . . . Watanabe, S. (1986). Control of ventilation in extreme-altitude climbers. *Journal of Applied Physiology, 61*(2), 500-506.

Merriwether, D. A., Clark, A. G., Ballinger, S. W., Schurr, T. G., Soodyall, H., Jenkins, T., . . . Wallace, D. C. (1991). The structure of human mitochondrial DNA variation. *Journal of Molecular Evolution, 33*(6), 543-555.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., . . . de Filippo, C. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science, 338*(6104), 222-226.

Milledge, J., & Lahiri, S. (1967). Respiratory control in lowlanders and Sherpa highlanders at altitude. *Respiration physiology, 2*(3), 310-322.

Minetti, A. E., Formenti, F., & Ardigò, L. P. (2006). Himalayan porter's specialization: metabolic power, economy, efficiency and skill. *Proceedings of the Royal Society B: Biological Sciences, 273*(1602), 2791-2797.

Mishmar, D., Ruiz-Pesini, E., Golik, P., Macaulay, V., Clark, A. G., Hosseini, S., . . . Brown, M. D. (2003). Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences, 100*(1), 171-176.

Moncada, S., Palmer, R., & Higgs, E. (1991). Nitric oxide: physiology, pathophysiology, and pharmacology. *Pharmacological reviews, 43*(2), 109-142.

Montgomery, H., Marshall, R., Hemingway, H., Myerson, S., Clarkson, P., Dollery, C., . . . World, M. (1998). Human gene for physical performance. *Nature, 393*(6682), 221-222.

Moore, L. G. (2003). Fetal growth restriction and maternal oxygen transport during high altitude pregnancy. *High altitude medicine & biology, 4*(2), 141-156.

Moore, L. G., Charles, S. M., & Julian, C. G. (2011). Humans at high altitude: hypoxia and fetal growth. *Respiratory physiology & neurobiology, 178*(1), 181-190.

Moore, L. G., Zamudio, S., Zhuang, J., Sun, S., & Droma, T. (2001). Oxygen transport in Tibetan women during pregnancy at 3,658 m. *American journal of physical anthropology, 114*(1), 42-53.

Morpurgo, G., Arese, P., Bosia, A., Pescarmona, G., Luzzana, M., & Modiano, G. (1976). Sherpas living permanently at high altitutde: a new pattern of adaptation. *Proceedings of the National Academy of Sciences, 73*(3), 747-751.

Moseley, M. E. (1997). *The Incas and their ancestors: the archaeology of Peru*: Thames and Hudson.

Murray, A. J. (2009). Metabolic adaptation of skeletal muscle to high altitude hypoxia: how new technologies could resolve the controversies. *Genome Med, 1*(12), 117-117.

Nair, C., Malhotra, M., & Gopinath, P. (1971). Effect of altitude and cold acclimatisation on the basal metabolism in man. *Aerospace medicine, 42*(10), 1056-1059.

Nass, M. M., & Nass, S. (1963). Intramitochondrial fibers with DNA characteristics I. Fixation and electron staining reactions. *The Journal of cell biology, 19*(3), 593-611.

Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular biology and evolution, 3*(5), 418-426.

Neiman, M., & Taylor, D. R. (2009). The causes of mutation accumulation in mitochondrial genomes. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb. 2008.1758.

Niu, W., Wu, Y., Li, B., Chen, N., & Song, S. (1995). Effects of long-term acclimatization in lowlanders migrating to high altitude: comparison with high altitude residents. *European journal of applied physiology and occupational physiology, 71*(6), 543-548.

Ohno, S. (2013). *Sex chromosomes and sex-linked genes* (Vol. 1): Springer-Verlag.

Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Al-Zahery, N., . . . Dugoujon, J.-M. (2006). The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science, 314*(5806), 1767-1770.

Olivo, P. D., Van de Walle, M. J., Laipis, P. J., & Hauswirth, W. W. (1983). Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop.

Oppenheimer, S. (2012). A single southern exit of modern humans from Africa: Before or after Toba? *Quaternary International, 258*, 88-99.

Oppitz, M. (1974). Myths and facts: Reconsidering some data concerning the clan history of the Sherpas.

Otsuka, K., Norboo, T., Otsuka, Y., Higuchi, H., Hayajiri, M., Narushima, C., . . . Wada, T. (2005). Chronoecological health watch of arterial stiffness and neuro-cardio-pulmonary function in elderly community at high altitude (3524 m), compared with Japanese town. *Biomedicine & pharmacotherapy, 59*, S58-S67.

Papandreou, I., Cairns, R. A., Fontana, L., Lim, A. L., & Denko, N. C. (2006). HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. *Cell metabolism, 3*(3), 187-197.

Patitucci, M., & Lugrin, D. (2009). Angiogenic/lymphangiogenic factors and adaptation to extreme altitudes during an expedition to Mount Everest. *Acta physiologica, 196*(2), 259-265.

Pawson, I. (1977). Growth characteristics of populations of Tibetan origin in Nepal. *American journal of physical anthropology, 47*(3), 473-482.

Pawson, I. G. (1976). Growth and development in high altitude populations: a review of Ethiopian, Peruvian, and Nepalese studies. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 194*(1114), 83-98.

Peng, M.-S., Palanichamy, M. G., Yao, Y.-G., Mitra, B., Cheng, Y.-T., Zhao, M., . . . Wang, W.-Z. (2011). Inland post-glacial dispersal in East Asia revealed by mitochondrial haplogroup M9a'b. *BMC biology, 9*(1), 2.

Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X., . . . Shi, H. (2011). Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. *Molecular biology and evolution, 28*(2), 1075-1081.

Portman, M. A., Standaert, T. A., & Ning, X. (1996). Developmental changes in ATP utilization during graded hypoxia and reoxygenation in the heart in vivo. *American Journal of Physiology-Heart and Circulatory Physiology, 270*(1), H216-H223.

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology, 20*(4), R208-R215.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., . . . de Filippo, C. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature, 505*(7481), 43-49.

Pugh, L., Gill, M., Lahiri, S., Milledge, J., Ward, M., & West, J. (1964). Muscular exercise at great altitudes. *Journal of Applied Physiology, 19*(3), 431-440.

Pugh, L. G. C. E. (1962). Physiological and medical aspects of the Himalayan Scientific and Mountaineering Expedition. *BMJ, 2*(5305), 621-627.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575.

Qi, X., Cui, C., Peng, Y., Zhang, X., Yang, Z., Zhong, H., . . . Su, B. (2013). Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the tibetan plateau. *Mol Biol Evol, 30*(8), 1761-1778. doi: 10.1093/molbev/mst093

Qin, P., & Stoneking, M. (2015). Denisovan Ancestry in East Eurasian and Native American Populations. *bioRxiv*, 017475.

Qin, Z., Yang, Y., Kang, L., Yan, S., Cho, K., Cai, X., . . . Fei, D. (2010). A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. *American journal of physical anthropology, 143*(4), 555-569.

Racimo, F., Sankararaman, S., Nielsen, R., & Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics, 16*(6), 359-371.

Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K. E., Rasmussen, S., Albrechtsen, A., . . . Jombart, T. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science, 334*(6052), 94-98.

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., . . . Johnson, P. L. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature, 468*(7327), 1053-1060.

Renfrew, C. (1996). Language families and the spread of farming. *The origins and spread of agriculture and pastoralism in Eurasia*, 70-92.

Richards, M., Macaulay, V., Torroni, A., & Bandelt, H.-J. (2002). In search of geographical patterns in European mitochondrial DNA. *The American Journal of Human Genetics, 71*(5), 1168-1174.

Roberts, A., Butterfield, G., Cymerman, A., Reeves, J., Wolfel, E., & Brooks, G. (1996). Acclimatization to 4,300-m altitude decreases reliance on fat as a substrate. *Journal of Applied Physiology, 81*(4), 1762-1771.

Rozen, S., Skaletsky, H., Marszalek, J. D., Minx, P. J., Cordum, H. S., Waterston, R. H., . . . Page, D. C. (2003). Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature, 423*(6942), 873-876.

Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.-J., & Nørby, S. (2000). mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *The American Journal of Human Genetics, 67*(3), 718-726.

Samaja, M., Mariani, C., Prestini, A., & Cerretelli, P. (1997). Acid-base balance and O2 transport at high altitude. *Acta physiologica scandinavica, 159*(3), 249-256.

Samaja, M., Veicsteinas, A., & Cerretelli, P. (1979). Oxygen affinity of blood in altitude Sherpas. *Journal of Applied Physiology, 47*(2), 337-341.

Sanson, M., Ingueneau, C., Vindis, C., Thiers, J.-C., Glock, Y., Rousseau, H., . . . Salvayre, R. (2008). Oxygen-regulated protein-150 prevents calcium homeostasis deregulation and apoptosis induced by oxidized LDL in vascular cells. *Cell Death & Differentiation, 15*(8), 1255-1265.

Santolaya, R., Lahiri, S., Alfaro, R., & Schoene, R. (1989). Respiratory adaptation in the highest inhabitants and highest Sherpa mountaineers. *Respiration physiology, 77*(2), 253-262.

Schadt, E. E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P. Y., . . . Suver, C. (2008). Mapping the genetic architecture of gene expression in human liver. *PLoS Biol, 6*(5), e107.

Scheinfeldt, L. B., Soi, S., Thompson, S., Ranciaro, A., Meskel, D., Beggs, W., . . . Belay, G. (2012). Genetic adaptation to high altitude in the Ethiopian highlands. *Genome Biol, 13*(1), R1.

Scheinfeldt, L. B., Soi, S., & Tishkoff, S. A. (2010). Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. *Proceedings of the National Academy of Sciences, 107*(Supplement 2), 8931-8938.

Scheinfeldt, L. B., & Tishkoff, S. A. (2013). Recent human adaptation: genomic approaches, interpretation and insights. *Nat Rev Genet, 14*(10), 692-702. doi: 10.1038/nrg3604

Schlattl, A., Anders, S., Waszak, S. M., Huber, W., & Korbel, J. O. (2011). Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions. *Genome research, 21*(12), 2004-2013.

Schneider, A., Greene, R. E., Keyl, C., Bandinelli, G., Passino, C., Spadacini, G., . . . Boiardi, A. (2001). Peripheral arterial vascular function at altitude: sea-level natives versus Himalayan high-altitude natives. *Journal of hypertension, 19*(2), 213-222.

Schoene, R., Lahiri, S., Hackett, P., Peters, R., Milledge, J., Pizzo, C., . . . Maret, K. (1984). Relationship of hypoxic ventilatory response to exercise performance on Mount Everest. *Journal of Applied Physiology, 56*(6), 1478-1483.

Scortegagna, M., Ding, K., Oktay, Y., Gaur, A., Thurmond, F., Yan, L.-J., . . . Richardson, J. A. (2003). Multiple organ pathology, metabolic abnormalities and impaired homeostasis of reactive oxygen species in Epas1−/− mice. *Nature genetics, 35*(4), 331-340.

Semenza, G. L. (1999). Perspectives on oxygen sensing. *Cell, 98*(3), 281-284.

Sengupta, S., Zhivotovsky, L. A., King, R., Mehdi, S., Edmonds, C. A., Chow, C.-E. T., . . . Ramesh, A. (2006). Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *The American Journal of Human Genetics, 78*(2), 202-221.

Shi, H., Dong, Y.-l., Wen, B., Xiao, C.-J., Underhill, P. A., Shen, P.-d., . . . Su, B. (2005). Y-chromosome evidence of southern origin of the East Asian–specific haplogroup O3-M122. *The American Journal of Human Genetics, 77*(3), 408-419.

Shi, H., Zhong, H., Peng, Y., Dong, Y.-L., Qi, X.-B., Zhang, F., . . . Xiao, C.-J. (2008). Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC biology, 6*(1), 45.

Simonson, T. S. (2010). Genetic evidence for high-altitude adaptation in Tibet. *Science, 329*, 72-75.

References

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., . . . Bieri, T. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature, 423*(6942), 825-837.

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. *American Journal of Human Biology, 9*(6), 751-763.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical research, 23*(01), 23-35.

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., . . . Richards, M. B. (2009). Correcting for purifying selection: an improved human mitochondrial molecular clock. *The American Journal of Human Genetics, 84*(6), 740-759.

Soares, P., Trejaut, J. A., Loo, J.-H., Hill, C., Mormina, M., Lee, C.-L., . . . Macaulay, V. (2008). Climate change and postglacial human dispersals in Southeast Asia. *Molecular biology and evolution, 25*(6), 1209-1218.

Stembridge, M., Ainslie, P. N., Hughes, M. G., Stöhr, E. J., Cotter, J. D., Nio, A. Q., & Shave, R. (2014). Ventricular structure, function, and mechanics at high altitude: chronic remodeling in Sherpa vs. short-term lowlander adaptation. *Journal of Applied Physiology, 117*(3), 334-343.

Stembridge, M., Ainslie, P. N., & Shave, R. (2014). Short-term adaptation and chronic cardiac remodelling to high altitude in lowlander natives and Himalayan Sherpa. *Experimental Physiology*.

Stevens, S. F. (1996). *Claiming the High Ground: Sherppassubsistenceand Environmental Change in the Highest Himalaya*: Motilal Banarsidass Publishe.

Stoneking, M., & Soodyall, H. (1996). Human evolution and the mitochondrial genome. *Current opinion in genetics & development, 6*(6), 731-736.

Su, B., Jin, L., Underhill, P., Martinson, J., Saha, N., McGarvey, S. T., . . . Deka, R. (2000). Polynesian origins: Insights from the Y chromosome. *Proceedings of the National Academy of Sciences of the United States of America, 97*(15), 8225-8228.

Su, B., Xiao, C., Deka, R., Seielstad, M. T., Kangwanpong, D., Xiao, J., . . . Chakraborty, R. (2000). Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Human genetics, 107*(6), 582-590.

Su, B., Xiao, J., Underhill, P., Deka, R., Zhang, W., Akey, J., . . . Luo, J. (1999). Y-Chromosome evidence for a northward migration of modern humans into Eastern

Asia during the last Ice Age. *The American Journal of Human Genetics, 65*(6), 1718-1724.

Suzuki, K., Kizaki, T., Hitomi, Y., Nukita, M., Kimoto, K., Miyazawa, N., . . . Ohno, H. (2003). Genetic variation in hypoxia-inducible factor 1α and its possible association with high altitude adaptation in Sherpas. *Medical hypotheses, 61*(3), 385-389.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution, 28*(10), 2731-2739.

Taylor, C. T., & Moncada, S. (2010). Nitric oxide, cytochrome C oxidase, and the cellular response to hypoxia. *Arteriosclerosis, thrombosis, and vascular biology, 30*(4), 643-647.

Thangaraj, K., Chaubey, G., Kivisild, T., Rani, D. S., Singh, V. K., Ismail, T., . . . Reddy, A. G. (2008). Maternal footprints of southeast Asians in North India. *Human heredity, 66*(1), 1-9.

Underhill, P. A., & Kivisild, T. (2007). Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu. Rev. Genet., 41*, 539-564.

Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazon Lahr, M., Foley, R. A., . . . Cavalli-Sforza, L. L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of human genetics, 65*(01), 43-62.

Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., . . . Francalacci, P. (2000). Y chromosome sequence variation and the history of human populations. *Nature genetics, 26*(3), 358-361.

van Driem, G. (2013). East Asian Ethnolinguistic Phylogeography.

Van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human mutation, 30*(2), E386-E394.

Walker, A., Smith, S., & Smith, S. (1987). Mitochondrial DNA and human evolution. *Nature, 325*, 1-5.

Wang, B., Zhang, Y. B., Zhang, F., Lin, H., Wang, X., Wan, N., . . . Yu, J. (2011). On the origin of Tibetans and their genetic basis in adapting high-altitude environments. *PLoS One, 6*(2), e17002. doi: 10.1371/journal.pone.0017002

Wang, H.-W., Li, Y.-C., Sun, F., Zhao, M., Mitra, B., Chaudhuri, T. K., . . . Zhang, Y.-P. (2012). Revisiting the role of the Himalayas in peopling Nepal: insights from mitochondrial genomes. *Journal of human genetics, 57*(4), 228-234.

Ward, M. (1954). High altitude deterioration. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 40-42.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *evolution*, 1358-1370.

Weischenfeldt, J., Symmons, O., Spitz, F., & Korbel, J. O. (2013). Phenotypic impact of genomic structural variation: insights from and for human disease. *Nature Reviews Genetics, 14*(2), 125-138.

Weitz, C. A., & Garruto, R. M. (2007). A comparative analysis of arterial oxygen saturation among Tibetans and Han born and raised at high altitude. *High altitude medicine & biology, 8*(1), 13-26.

Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., . . . Shanmugalakshmi, S. (2001). The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences, 98*(18), 10244-10249.

West, J. (1986). *Lactate during exercise at extreme altitude.* Paper presented at the Federation proceedings.

White, D. J., Wolff, J. N., Pierson, M., & Gemmell, N. J. (2008). Revealing the hidden complexities of mtDNA inheritance. *Mol Ecol, 17*(23), 4925-4942.

Whitehouse, P., Usher, T., Ruhlen, M., & Wang, W. S.-Y. (2004). Kusunda: An Indo-Pacific language in Nepal. *Proceedings of the National Academy of Sciences of the United States of America, 101*(15), 5692-5695.

Wilson, M., Stoneking, M., Holland, M., DiZinno, J., & Budowle, B. (1993). Guidelines for the use of mitochondrial DNA sequencing in forensic science. *Crime Lab Digest, 20*(4), 68-77.

Wilson, M. H., Edsell, M. E., Davagnanam, I., Hirani, S. P., Martin, D. S., Levett, D. Z., . . . Newman, S. P. (2011). Cerebral artery dilatation maintains cerebral oxygenation at extreme altitude and in acute hypoxia—an ultrasound and MRI study. *Journal of Cerebral Blood Flow & Metabolism, 31*(10), 2019-2029.

Wilson, M. J., Lopez, M., Vargas, M., Julian, C., Tellez, W., Rodriguez, A., . . . Shriver, M. (2007). Greater uterine artery blood flow during pregnancy in multigenerational (Andean) than shorter-term (European) high-altitude residents. *American Journal of*

Physiology-Regulatory, Integrative and Comparative Physiology, 293(3), R1313-R1324.

Wink, D. A., Miranda, K. M., Espey, M. G., Pluta, R. M., Hewett, S. J., Colton, C., . . . Grisham, M. B. (2001). Mechanisms of the antioxidant effects of nitric oxide. *Antioxidants and Redox Signaling, 3*(2), 203-213.

Winslow, R. M. (2007). The role of hemoglobin oxygen affinity in oxygen transport at high altitude. *Respiratory physiology & neurobiology, 158*(2), 121-127.

Winslow, R. M., Chapman, K. W., Gibson, C., Samaja, M., Monge, C., Goldwasser, E., . . . Santolaya, R. (1989). Different hematologic responses to hypoxia in Sherpas and Quechua Indians. *Journal of Applied Physiology, 66*(4), 1561-1569.

Wu, T., Liu, F., Cui, C., Qi, X., & Su, B. (2013). A genetic adaptive pattern-low hemoglobin concentration in the Himalayan highlanders. *Zhongguo ying yong sheng li xue za zhi= Zhongguo yingyong shenglixue zazhi= Chinese journal of applied physiology, 29*(6), 481-493.

Wuren, T., Simonson, T. S., Qin, G., Xing, J., Huff, C. D., Witherspoon, D. J., . . . Ge, R. L. (2014). Shared and unique signals of high-altitude adaptation in geographically distinct Tibetan populations. *PLoS One, 9*(3), e88252. doi: 10.1371/journal.pone.0088252

Xiang, K., Peng, Y., Yang, Z., Zhang, X., Cui, C., Zhang, H., . . . Chen, H. (2013a). Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. *Molecular biology and evolution*.

Xiang, K., Peng, Y., Yang, Z., Zhang, X., Cui, C., Zhang, H., . . . Chen, H. (2013b). Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. *Molecular biology and evolution, 30*(8), 1889-1898.

Xing, J. (2010). Toward a more uniform sampling of human genetic diversity: a survey of worldwide populations by high-density genotyping. *Genomics, 96*, 199-210.

Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W., . . . Shen, Y. (2011). A genome-wide search for signals of high-altitude adaptation in Tibetans. *Molecular biology and evolution, 28*(2), 1003-1011.

Yao, Y.-G., Kong, Q.-P., Bandelt, H.-J., Kivisild, T., & Zhang, Y.-P. (2002). Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *The American Journal of Human Genetics, 70*(3), 635-651.

References

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., . . . Korneliussen, T. S. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science, 329*(5987), 75-78.

Zeller, T., Wild, P., Szymczak, S., Rotival, M., Schillert, A., Castagne, R., . . . Rossmann, H. (2010). Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS One, 5*(5), e10693-e10693.

Zhao, M., Kong, Q.-P., Wang, H.-W., Peng, M.-S., Xie, X.-D., Wang, W.-Z., . . . Tu, Y.-Q. (2009). Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proceedings of the National Academy of Sciences, 106*(50), 21230-21235.

Zhao, M., Kong, Q. P., Wang, H. W., Peng, M. S., Xie, X. D., Wang, W. Z., . . . Zhang, Y. P. (2009). Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc Natl Acad Sci U S A, 106*(50), 21230-21235. doi: 10.1073/pnas.0907844106

Zhong, H., Shi, H., Qi, X.-B., Duan, Z.-Y., Tan, P.-P., Jin, L., . . . Ma, R. Z. (2011). Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Molecular biology and evolution, 28*(1), 717-727.

LIST OF ORIGINAL PUBLICATIONS

1.1 **Sushil Bhandari**, Xiaoming Zhang, Chaoying Cui, Bianba, Shiyu Liao, Yi Peng, Hui Zhang, Kun Xiang, Hong Shi, Ouzhuluobu, Baimakongzhuo, Gonggalanzi, Shimin Liu, Gengdeng, Tianyi Wu, Xuebin Qi, Bing Su. *Genetic evidence of a recent Tibetan ancestry to Sherpas in the Himalayan region.* Scientific report 2015

1.2 Bhandari *et al...* Sherpas share genetic variations with Tibetans for high-altitude adaptation (manuscript under preparation)

ACKNOWLEDGMENTS

My deepest gratitude goes to my supervisor Professor **Su Bing** for giving me a great opportunity to work on interesting research projects in Sherpa population. His guidance, constant encouragement and constructive suggestions throughout the course of my study have made remarkable influence throughout my career. I am very fortunate to get chance to study under his direct supervision in his prestigious Lab at Kunming Institute of Zoology, Chinese Academy of Sciences, China.

I am grateful to all the people who have donated their blood for making this study possible. This work became essentially possible through the collaboration with several people who helped us to collect samples from Tibet and Nepal. I am indebted to all the participating staff from High Altitude Medical Research Center, School of Medicine, Tibetan University and Nepal Health Research Council, Nepal. I take the opportunity to thank Dr. Ashwani Kumar Sinha and Dr. Zhang Xiaoming for collecting samples from remote Sherpa villages of Nepal.

I acknowledge and wish to thank all the members of Prof. Su's laboratory for their kind cooperation and support. I also like to thank Dr. Qi Xuebin, Dr. Shi Hong, Dr. Hua Chen, Dr. Li Ming，Dr. Shi Lei and teacher Yu for providing valuable advices in my research. I truly enjoy the friendly working atmosphere in the lab. I am thankful to Dr. Peng Yi, Dr. Yang Zhao Hui, Dr. Xiang Kun and Dr. Yang Deying for teaching me different technical skills in doing experiment. Besides Lab work, different other outside activities organized by Lab; like going together out for having BBQ, lunch, dinner, singing in KTV, travelling, playing games etc. gives me homely environment in KIZ, China. I express heartfelt thanks to Miss Zhang Hui and Miss Gou Yan for providing technical assistance to conduct Lab works. I also like to thanks all other (Lin Qiang, Liu Jiewei, Yang Yandong, Yang Lixin, Zhang Xu, Dang, He
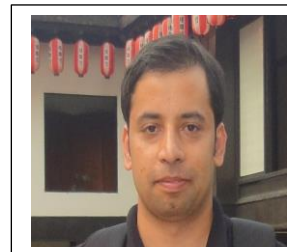
CURRICULUM VITAE

Personal Details

Name: **Sushil Bhandari**

Date of Birth: 22 November 1986

Father Name: Krishna Prasad Bhandari

Mother Name: Bhawani Bhandari

Gender: Male

Nationality: Nepalese

Permanent Address: Tupche, 2 Nuwakot, Nepal

Email: sushilnuwakot@gmail.com

Academic Background

2012-2016   PhD scholar under the supervision of Prof. Su Bing in State Key

               Laboratory of Genetic Resources and Evolution, Kunming Institute of

               Zoology, Chinese Academy of Sciences, China.

2010-2012   Worked as a Research Assistant in Nepal Academy of Science and

               Technology, Nepal in Molecular characterization of High Altitude

               Medicinal plants.

2009-2010   M. Sc. Dissertation from Center for Cellular and molecular biology

               (CCMB), Hyderabad, India. Title: "Y Chromosome and mtDNA study

               in Brahmin population on Kathmandu district of Nepal" under the

               guidance of Dr K. Thangaraj

2008-2010   Master of Science (M. Sc.) in Biotechnology. Osmania University,

               Hyderabad, India

2004-2008   Bachelor's Degree (BSc) in Microbiology. Tribhuvan University, Nepal