Genetic variation and adaptation in the Quechua, a high altitude native population.

by

James L. Rupert

B.Sc., The University of Guelph, 1984
M.Sc., University of Toronto, 1988

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Department of Zoology

We accept this thesis as conforming to the required standard.

THE UNIVERSITY OF BRITISH COLUMBIA
April 2000
© James L. Rupert
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of **Zoology**

The University of British Columbia  
Vancouver, Canada

Date  
25/04/2000
ABSTRACT

Human adaptation to high altitude involves a suite of physiological and anatomical characteristics designed to facilitate the uptake, transport and utilization of oxygen, thereby compensating for the diminished oxygen availability inherent to high altitude. While many of these characteristics may be acclimatory or developmental in nature, there is evidence that some are influenced by the genetic background of the population. The studies presented in this thesis are designed to examine the influence of genotype by comparing the frequencies of variants in a number of candidate genes between the Quechua, a high altitude population indigenous to Andean altiplano, and two low altitude populations: Na-Dene from the west coast of Canada and Caucasians of western European descent. Genes were chosen because they had previously characterized polymorphisms with phenotypes that a priori were judged to be of potential benefit at altitude and include: 1) the β2-adrenergic receptor, the principle pulmonary catecholamine receptor; 2) angiotensin converting enzyme, a regulator of systemic blood flow; 3) β-fibrinogen, a common plasma protein and primary determinant of plasma viscosity; 4) erythropoietin, a hormone involved in the synthesis of red blood cells and 5) hypoxia inducible factor 1-alpha, a major component in oxygen-dependent gene regulation. Given the theoretical requirements of high altitude populations and the phenotypes associated with these alleles of these genes, predictions were made as to which alleles should have been selected for, and therefore currently be over-represented, in the Quechua.

As expected, given the diverse origins of the populations examined, there were significant differences in haplotype structure and allele frequencies. While some of these differences may have resulted from the random divergence of separated populations over time, others, such as the under-representation of alleles associated with increased fibrinogen may reflect an adaptive change in the Quechua.

Successful expansion into a demanding environment will often be reflected in the genetic make-up of a population. Previous studies investigating the role of genetics in high altitude adaptation have concentrated on the heritability of complex traits. The work presented in this thesis represents an alternate approach, one in which the focus is on the frequencies of variants in individual genes in an altitude-adapted population.
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td></td>
<td>ii</td>
</tr>
<tr>
<td>Table of contents</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>List of Figures</td>
<td></td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td></td>
<td>viii</td>
</tr>
<tr>
<td>Preface</td>
<td></td>
<td>ix</td>
</tr>
<tr>
<td><strong>Chapter 1</strong></td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.1 Project overview</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.2 Human high altitude adaptation in the Andes</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1.3 The Peopling of the New World</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>1.4 The South American highlands</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>1.5 The Quechua</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>1.6 Summary</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>1.7 Rationale</td>
<td>27</td>
</tr>
<tr>
<td><strong>Chapter 2</strong></td>
<td>Materials and Methods</td>
<td>29</td>
</tr>
<tr>
<td><strong>Chapter 3</strong></td>
<td>$\beta$-fibrinogen allele frequencies in the Quechua</td>
<td>35</td>
</tr>
<tr>
<td><strong>Chapter 4</strong></td>
<td>Angiotensin converting enzyme (ACE) and angiotensin 2 allele frequencies in the Quechua</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Chapter 5</strong></td>
<td>Beta$_2$-adrenergic receptor allele frequencies in the Quechua</td>
<td>60</td>
</tr>
<tr>
<td><strong>Chapter 6</strong></td>
<td>The sequence of the erythropoietin and HIF-1$\alpha$ genes in the Quechua</td>
<td>73</td>
</tr>
<tr>
<td><strong>Chapter 7</strong></td>
<td>Discussion, conclusions and future considerations</td>
<td>85</td>
</tr>
<tr>
<td>Abbreviations</td>
<td></td>
<td>93</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>94</td>
</tr>
<tr>
<td>Appendices</td>
<td></td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>i) Consent form</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>ii) Primer sequences</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td>iii) DNA sequences</td>
<td>115</td>
</tr>
</tbody>
</table>
List of Figures

Two Quechua women in Ollantaytambo, Peru.......................... vi
1.1 The relationship between altitude and atmospheric pressure........ 5
1.2 Hematocrits in Andean native populations.............................. 14
1.3 Beringia and Pleistocene glaciation patterns............................ 16
1.4 Changes in allele frequencies over 600 generations of selection.... 20
1.5 Maps of South America and Peru......................................... 21
1.6 A summary of human adaptation to altitude............................ 26
3.1 Fibrinogen concentration and viscosity................................ 36
3.2 The role of fibrinogen in the clotting cascade.......................... 37
3.3 Assay for RFLPs in the β-fibrinogen gene............................... 39
3.4 Relative plasma viscosity in Quechua and Caucasians................ 41
4.1 The ACE-mediated vasomodulation system.............................. 49
4.2 ACE I/D alleles PCR amplification assay............................... 53
4.3 Assay for polymorphisms in the ACE and AT2-R1 genes.............. 55
5.1 The structure of the β2-adrenergic receptor........................... 62
5.2 Assay for RFLPs in the β2-AR gene................................... 65
6.1 Erythropoietin and the production of red blood cells................. 75
6.2 Regulatory regions of the erythropoietin gene........................ 76
6.3 Location of erythropoietin gene PCR primers.......................... 77
6.4 Assay for RFLPS in the erythropoietin gene............................ 79
6.5 Polymorphic sites in erythropoietin gene.............................. 80
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Hematocrit values for Andean populations</td>
<td>12</td>
</tr>
<tr>
<td>1.2</td>
<td>Early archeological sites in the Andes</td>
<td>18</td>
</tr>
<tr>
<td>1.3</td>
<td>Predicted change in allele frequencies after 600 and 1650 generations</td>
<td>25</td>
</tr>
<tr>
<td>2.1</td>
<td>Age and hematocrit of the Quechua who donated blood samples</td>
<td>31</td>
</tr>
<tr>
<td>3.1</td>
<td>β-fibrinogen PCR primers</td>
<td>40</td>
</tr>
<tr>
<td>3.2</td>
<td>β-fibrinogen genotype and allele frequencies</td>
<td>42</td>
</tr>
<tr>
<td>3.3</td>
<td>Linkage disequilibrium between the β-fibrinogen alleles</td>
<td>43</td>
</tr>
<tr>
<td>4.1</td>
<td>ACE and AT2-R1 genotype and allele frequencies</td>
<td>54</td>
</tr>
<tr>
<td>4.2</td>
<td>Published allele frequencies for the ACE I/D polymorphism</td>
<td>56</td>
</tr>
<tr>
<td>5.1</td>
<td>Location and product size of β₂-AR PCR primers</td>
<td>64</td>
</tr>
<tr>
<td>5.2</td>
<td>β₂-AR genotype and allele frequencies</td>
<td>68</td>
</tr>
<tr>
<td>5.3</td>
<td>Linkage disequilibrium between the β₂-AR alleles</td>
<td>69</td>
</tr>
<tr>
<td>6.1</td>
<td>Erythropoietin allele genotype and frequencies</td>
<td>81</td>
</tr>
</tbody>
</table>
Two Quechua women in Ollantaytambo, Peru.
The foundations of the great city of Cusco once lay on a vast, open pampas. As Inkariy toiled to build his city, the wind howled and blew across the plains, buffeting him as he labored and blowing over the walls as fast as he could put them up. Angry, Inkariy locked away the winds in a great corral. As he did Inka Qulla, who owned the winds appeared and demanded to know why his winds were locked away. Inkariy told him that, if he was to build his city, the winds could not be free. Inka Qulla agreed to this, but only for one day. The cunning Inkariy then tied a rope around the sun and kept it from sailing across the sky so that the day he built Cusco was very long indeed. When he finished, his wise wife told him to raise great mountains around his new city because, when they were freed, the angry winds would surely try to knock it down. Inkariy did and so Cusco, the great capital of the Inca Empire, lies high in the mountains, safe from the winds.¹

¹ A Quechua tale adapted Andean Lives. (Ricardo Valderrama Fernandez and Carmen Escalante Gutierrez eds., The University of Texas Press, 1996). For the purposes of this thesis it will be assumed that the mountains antecede the people.
Acknowledgments

First and foremost, I would like to thank Dr. Peter Hochachka. In his capacity as my Ph.D. supervisor, his suggestions during the conception, execution, analysis and publication of the work presented in this thesis were invaluable. In his capacity as a friend and colleague, he was a constant reminder of how much we know, how little we know and how much fun science can be.

I would also like to thank my co-supervisor, Dr. Dana Devine, who initially suggested that I consider β-fibrinogen as a candidate gene and encouraged me to look to the Heart and Stroke Foundation of Canada for support; and Drs. David Jones, Linda Matsuuchi, John Gosline, Don McKenzie and Bill Milsom, all of whom served at some point on my supervisory and/or examination committees.

For their contributions to this project, I would like to thank Dr. Vicky Monsalve for doing much of the work involved in obtaining the Quechua blood samples, to Sarah Baldry for maintaining my tissue culture lines and for preparing the DNAs and RNAs that were integral to the work presented in chapter 6, to Dr. Wendy Robinson for statistical advice, to Dr. Maria Issa for patiently teaching me how to quantitate fibrinogen; and to Drs. Don Brooks and Johen Jansen for apparatus and advice regarding viscometry. Also, thanks are due to our Peruvian associates, Charo Tapia and Dr. Maria Rivera Chira, for their assistance in arranging the blood gathering expedition to the Andean altiplano.

All of the U.B.C. departments in which I was fortunate to spend time over the last five years are full of fun and interesting people, many of whom contributed to both the work and play that comprised my tenure as a graduate student. This list is long and includes, among others, Drs. (or Drs. to be) Amanda Southwood, Manuela Gardner, Russ Andrews, Sheila Thornton, Gary Burness, Grant McClelland, Kevin Campbell, Charles Darveau, Beth Zimmer, Lowell McPhail, Susan Shinn and Thuan Nguyen.

This project could never have been completed, and likely never would have been initiated (at least by me), without the support of my learned and extremely competent wife, Dr. Carolyn Brown who, in unofficial capacities, served variously as advisor, assistant, reviewer, editor, critic, proof-reader, de-bunker and supporter; and who was always willing to turn a blind eye to my frequent use of her lab (including, but not limited to, equipment, personnel, supplies and resources).

Lastly, thanks to Nimue, who patiently sitting through (and occasionally on) the preparation of yet another thesis, kept me company during those late night sessions of composition and decomposition.

This work was supported by the Natural Science and Engineering Research Council in the form of operating grants to Peter Hochachka and as a scholarship to myself. I was also fortunate to receive a Heart and Stroke Foundation of Canada Research Traineeship over the last two years of my tenure as a graduate student. To both of these organizations, I would like to extend my thanks for their support.

Jim Rupert
20/4/00
Preface

Portions of the material presented in this thesis have previously been published as the following:


This work was performed in Dr. Peter Hochachka's lab in the Department of Zoology at The University of British Columbia. Drs. Devine and Monsalve (Department of Pathology and Laboratory Medicine, U.B.C.) were involved in the procurement of some of the samples and made conceptual contributions to the project.

Experimental design and execution and as well as data analysis and manuscript preparation were the responsibility of Jim Rupert, a Ph.D. candidate under the supervision of Dr. Hochachka.

Peter Hochachka
Department of Zoology
U.B.C. 22/4/00
Chapter 1
Introduction

1.1 Project overview

Where order in variety we see,
And where, though all things differ, all agree.
(Alexander Pope, 1713)

Human beings have lived on the high plateaus of the Andes Mountains for hundreds of generations. The sturdy, barrel chested Andean villager has become an oft-cited paradigm of the ability of man to adapt to harsh and seemingly inhospitable environments. The nature of these adaptations has long intrigued researchers and, over the last century, much has been learned about the physiological and anatomical characteristics of these people. Much of the work has focused on adaptations that serve to maximise the uptake and utilisation of oxygen which, at the altitude of the Peruvian altiplano (3200-4000 m), is reduced by a third.

One issue that often arises in these studies is whether the characteristic can be accounted for solely by acclimatory or developmental responses within an individual or whether it is influenced by genetic background. Heritability studies have shown that, in some cases, ethnic heritage contributes to the acquisition of an adaptive trait and thus suggest that the Andean natives, whose ancestors first ascended to the high plateaus millennia ago, have evolved, at first perhaps just to survive but eventually to thrive, in the cold, thin air of the altiplano.

This project describes further studies into the role of genetics in high altitude adaptation using association analysis. The frequencies of variants in candidate genes, chosen because they potentially effect the uptake, transport or utilisation of oxygen, were compared between high altitude and low altitude populations to determine whether any of the alleles were over-represented in the highlanders. Such a correlation would be consistent with selection favouring that particular variant (or one to which it is tightly linked) in that population and may reflect adaptive changes at the population level. This is essentially a variation of molecular association analysis, a methodology commonly used to look for genetic contributions to the aetiology of complex diseases.
Each of the four data chapters describes an independent set of experiments using similar techniques and samples but assaying alleles of different genes. The genes and polymorphisms examined are described briefly below and in detail in the ensuing chapters. They include:

1) *β*-fibrinogen

β-fibrinogen is a component of fibrinogen, a common plasma protein that is the precursor to fibrin. Three polymorphisms were examined in the *β*-fibrinogen gene, all of which had previously been shown to have alleles associated with reduced fibrinogen concentrations. Two are upstream of the coding sequence and are believed to affect transcription of the gene; the third is a missense mutation within the coding sequence of the gene. It was hypothesised that the alleles associated with increased levels of protein would have been selected against in the Quechua as fibrinogen concentration is correlated with plasma viscosity and increased viscosity may exacerbate the rheological and physiological problems associated with elevated hematocrits.

2) angiotensin converting enzyme (ACE)

ACE is a component of the renin-angiotensin system that acts proteolytically to convert angiotensin 1 into the potent vasoconstrictor angiotensin 2. The locus examined in this gene was the commonly studied *Alu* insertion/deletion polymorphism in intron 16. The insertion allele, which correlates with lower ACE levels, has been associated with reduced susceptibility to cardiovascular disease, increased physical performance and, of particular relevance to this study, has been shown to be over-represented in elite British climbers. Because of this last observation, and because of its enhancing effect on maximum oxygen uptake, this allele was predicted to be over-represented in the Quechua. A common polymorphism in the angiotensin 2 receptor type 1 gene was also assayed because genotype at this locus was reported to affect the phenotype of the ACE insertion/deletion alleles.

3) the β₂-adrenergic receptor.

This receptor is the primary pulmonary catecholamine receptor and influences the flow of both air and blood through the lungs. Five polymorphisms were examined in this gene, all of which are in the coding sequence of the gene. Three of these are missense mutations that alter receptor structure and have been associated with multiple phenotypes including altered receptor
sensitivity, susceptibility to congestive heart failure and body fat deposition. As reduced
barometric pressure is the defining characteristic of the high altitude environment, alleles with the
potential to increase air flow were hypothesised to be over-represented in the Quechua. The
other two polymorphisms were silent and were analysed as potential markers for functional loci.

4) erythropoietin (Epo) and hypoxia inducible factor 1-alpha (HIF-1α)

Epo is a regulatory protein that is essential for the maturation of red blood cells. HIF-1α
is a component of the transcriptional regulatory factor HIF-1, which is involved in oxygen-
dependent gene regulation. Increased red blood cell production is a hallmark of hypoxia
acclimation and may (or may not) play a role in adaptation. HIF-1 is involved in the oxygen-
mediated expression of Epo as well as other genes. There were no reports in the literature of
functional mutations in either of these genes in humans, although a downstream polymorphism
near the HIF-1 binding region had been found in Epo. While this variant may not be functional
per se, it may be a marker for one that is, and was therefore included in the analysis. Both of
these genes were considered to be strong candidate genes for a role in high altitude adaptation
and were therefore sequenced in three Quechua in a search for polymorphisms common in that
population. Such polymorphisms could represent either amplification of extremely rare
polymorphisms that had not yet been detected in other populations or variants that had arisen in
the ancestral Quechua.

Genotypes were determined for each polymorphism and, in cases where there were
multiple polymorphic loci in the same gene, the possibility of the alleles being in linkage
disequilibrium was addressed.

The high altitude natives who participated in this study are Quechua-speaking
Amerindians of the Andean altiplano. Blood and buccal specimens collected from informed
volunteers in rural villages lying between 3400 and 4200 meters in the Central Peruvian
highlands. DNA was prepared from these samples and genotypes at the polymorphic loci
determined by polymerase chain reaction amplification, followed by diagnostic restriction
endonuclease digestion if necessary. In addition, five Quechua derived lymphoblast cell lines
were obtained from the Coriell Cell Repository. The Quechua, who are the descendants of the
once vast Inca empire, are believed to have lived high in the Andes for over 12,000 years. This is sufficient time for even a minor selective advantage conferred by an allele to manifest itself as a measurable increase in frequency in the extant population. For comparative purposes, allele frequencies were also determined in two low altitude populations: Athabaskan speaking Na-Dene Amerindians of the Canadian northwest and Caucasians of western European descent. While these populations are not closely related to the Quechua, they serve, along with previously published data, to provide some indication of the frequencies of these genetic variants in populations that are presumed not to have adapted to life in high altitude environments.

1.2 Human high altitude adaptation in the Andes

Reduced barometric pressure and the resultant hypoxia (figure 1.1) is one of the few unalterable environmental stresses encountered by humans during the course of our highly successful expansion across the face of the planet. Fluctuations in temperature, availability of water and food, predation and climatic extremes could largely be overcome by our burgeoning technology but, beyond small and expensive pressure chambers, atmospheric pressure remains beyond our control. Despite the inevitable discomfort experienced by migrants to high altitude, humans have adapted to the thin air and as they did, settlements thrived on, and eventually empires ruled, the great plateaus of the Andes and the Himalaya. Currently, about one percent of the world’s population lives permanently between 3300 m and 4500 m (Samaja, 1997). Beyond 4500 m hypoxic stress precludes long term habitation although about 260 people live at between 4850-5450 m in Phala, Tibet Autonomous Region, Peoples Republic of China and there are mining communities, such as Morococha (4540 m) and Mina Aquilar (4515 m) high in the Andes.

There is an extensive body of literature documenting the investigation into human adaptation to high altitude (see reviews by Monge and León-Velarde, 1991; Moore et al., 1994; Ramirez et al., 1999). It is beyond the scope of this introduction to summarize the hundreds of papers that have been published since 1889, when Paul Bert first encouraged François Viault to monitor his own hematocrit as he traveled from Lima, on the coast of Peru, to Morococha at
4540 m in the Andes mountains. Most of these studies are either in Andean natives or in Himalayan populations such as the Sherpa of Nepal or Tibetans. The research portion of this thesis centers on the Quechua of the Peruvian altiplano and therefore this introduction will focus on the research done on Andean populations. The primary goal of this introduction is to provide an overview of what is known about high altitude adaptation in Andean natives with a particular emphasis on the characteristics for which there is evidence of a genetic contribution.

Figure 1.1. The relationship between altitude and atmospheric pressure. Altitude and atmospheric pressure are shown on the right. Partial pressure of oxygen is shown on the left. The shaded box indicates the range over which high altitude human populations live in the Andes. The asterisk indicates the point above which long term human habitation is considered impossible. Various high points are indicated for reference. 1: Denver, USA (1600 m); 2: Cusco, Peru (home to almost a half million people, 3200 m); 3) Mt. Baker (a perennially snow capped peak visible from U.B.C., 3320 m); 4: The Matterhorn (4480 m); 5: Mt. Logan, highest mountain in Canada (6050 m); 6: Aconcagua, highest mountain in the Andes (6960 m) and 7: Mt. Everest, highest mountain in the world (8848 m).
Acclimation, development or adaptation?

The flow of oxygen from the external environment through the body to its ultimate destination in the mitochondria occurs through a series of stepwise movements from a region of higher partial pressure of oxygen (PO$_2$) to a region of lower PO$_2$. As the pressure of atmospheric oxygen is reduced, the driving force for this transfer is correspondingly diminished. Most of the characteristics associated with high altitude existence serve to overcome this by either increasing the PO$_2$ in the lungs or by facilitating the transport of oxygen from the lungs to the tissues. Some of these changes occur within minutes of exposure whereas others take days or months to appear. The more profound changes seem to require prolonged exposure during development and, in some cases, may represent phylogenetic adaptations characteristic of the population. The literature is sometimes inconsistent in the terms used to categorize these changes. For the purposes of this thesis, acclimation will refer exclusively to transient, reversible changes whereas adaptation, in which changes are permanent, will be subdivided into developmental and genetic, the latter referring to heritable characteristics.

One often discussed issue in human high altitude biology is whether there is a genetic component to adaptation or whether all of the changes manifested in the high altitude native can be accounted for by the plasticity of the human phenotype. Acclimation occurs in lowlanders traveling to altitude and developmental changes will occur to some extent regardless of genetic origin (e.g. increased chest dimensions reported in de Meer et al., 1995). Phenotypic plasticity can mediate the effects of selective pressure on genotype. If a beneficial characteristic can be acquired upon exposure, then the relative advantage of being genetically predisposed to the trait will be blunted, the extent of which will depend on the relative cost of acquiring the characteristic. Further confounding this is the possibility that the heritable potential may not be for the trait per se but for the developmental ability to acquire the trait. For a discussion of the interplay of genotype and phenotypic plasticity (see Hochachka et al., 1999).

Pulmonary function and ventilation

In the 1920s, Jourdannet, an early researcher in the Andes, described high altitude
natives as having a "vast chest [that] makes him comfortable in the midst of this thin air" (quoted by Houston, 1987). This is one of the earliest descriptions of what may the most commonly cited characteristics of high altitude natives: the "barrel chest". Hurtado (1932) commented on this and postulated that the enlarged chest might allow for increased lung volumes, thereby increasing oxygen intake. Numerous anthropometric studies have been undertaken to determine to what extent this putative adaptation is genetically determined. Several thoracic parameters were measured in Peruvian Aymara Indians native to 3,900 m (Kramer, 1992). Sternal length, anterior-posterior depth and chest circumference had heritable components accounting for approximately 30% of the variation observed for these parameters (by comparison, about 50% of the variation in stature was shown to inherited).

Melton (1992) compared development in two groups of children of high altitude heritage (a mix of Aymara and Quechua) who were born in either Tacna (560 m) or Puna (3900 m), Peru. The parents of the low altitude children were recent migrants from Puna. Despite having been raised at a relative low altitude, the Tacna children had the lengthened sternum characteristic of high altitude populations and chests as large as those of the children raised at 3900 m. This study is particularly informative as it examined a population within one generation of movement, eliminating the possibility of lowland admixture.

Parameters of lung function such as forced vital capacity (FVC) are correlated with thoracic dimensions (Whittaker, 1992). In an extensive study of lung function, Mueller et al., (1978) compared lung function in Aymara, non-Aymara and Mestizo (mixed) natives living at three altitude ranges (coastal, 2500-3500 m and 4000-4500 m) in the Department of Arica in Chile. They concluded that, while altitude of residence clearly affected lung function, the role of ethnicity was unclear. In male children living at 4000 m, Aymara had significantly higher values than did Mestizos; however; the opposite was seen in female children. In adults, there was evidence for ethnic effects but again the effect was gender dependent with significance not seen in males. Despite the ambiguity of the data, the authors concluded that ethnicity was important in adult lung function variation in the Aymara although this effect was secondary to the developmental effects of long term hypoxic exposure. In the same population, (Greksa et al.,
1988) showed that Aymara children had a greater forced vital capacity (relative to stature) than did European children raised at the same altitude (3800 m). Although these studies support the hypothesis that genetic heritage contributes to enhanced lung development in high altitude natives, others do not. For example, Frisancho (1975) reported that sea level natives who had acclimated to altitude during growth developed forced vital capacities (adjusted for stature) equal to those of high altitude natives.

In a study specifically designed to test whether there was genetic contribution to lung capacity in the Aymara, Greksa et al., (1996) used skin reflectance as an indirect measure of genetic composition. Skin reflectance is highly heritable and is useful in estimating the degree of admixture between two populations, especially when they strongly differ in complexion. In the case of Aymara and Europeans, the darker the skin, the greater the Aymara contribution to the admixture (Greksa, 1992). The results of the 1996 study show a significant association between skin colour and total lung capacity leaving the authors to conclude that, while environmental forces contributed to enhanced lung capacity in the Aymara, genetic factors also influence the development of this characteristic.

**Performance**

Maximum rate of oxygen consumption (\(\dot{V}O_{2\text{max}}\)) is an indicator of aerobic capacity and is a strong index of capacity for sustained work. The \(\dot{V}O_{2\text{max}}\) for “average” males is about 45 ml kg\(^{-1}\) min\(^{-1}\) although values can exceed 90 ml kg\(^{-1}\) min\(^{-1}\) in elite endurance athletes (Newsholme et al., 1994). Although \(\dot{V}O_{2\text{max}}\) values vary considerably with age and training, several researchers have demonstrated a heritable component contributing to this important physiological parameter (Bouchard et al., 1998). \(\dot{V}O_{2\text{max}}\) drops with exposure to altitude at a rate of about 10% per 1000 m ascended beyond 1500 m (Dua and Sen Gupta, 1980) and remains depressed even after years of exposure. This reduction in fitness is less pronounced in high altitude natives. In 1969, Baker assessed \(\dot{V}O_{2\text{max}}\) values in Quechua students and age/sex matched “white” students. Both groups had lived all of their lives at 3830 m. Average \(\dot{V}O_{2\text{max}}\) (in ml kg\(^{-1}\) min\(^{-1}\)) was about 10% higher in the Quechua than in the Caucasians. In the same study, \(\dot{V}O_{2\text{max}}\) values measured
in Quechua native to 4000 m were close to those of “reasonably fit white researchers” (49.1 ml kg\(^{-1}\) min\(^{-1}\) vs. 50.4 ml kg\(^{-1}\) min\(^{-1}\)) when the latter were at sea level and much higher when the Caucasians were at 4000 m (49.1 ml kg\(^{-1}\) min\(^{-1}\) vs. 38.1 ml kg\(^{-1}\) min\(^{-1}\)). The author concluded that the “Quechua heritage confers a special capacity for oxygen consumption at 4000 meters”. This capacity seems to be maintained even after six weeks of deacclimation at sea level suggesting that it is either developmentally or genetically influenced (Hochachka et al., 1991).

Additional evidence for a genetic component to superior physical performance at altitude comes from studies of Quechua, born or raised at sea level, who were transported to high altitude (4000 m). The resultant reduction in \(\dot{V}O_{2\text{max}}\) was less than half of that experienced by Caucasians exposed to the same conditions (Buskirk, 1976).

A variant in the gene encoding angiotensin converting enzyme (ACE) has been reported to be associated with increased \(\dot{V}O_{2\text{max}}\) and with physical performance at extreme altitude. The potential role of this gene in altitude adaptation and the frequency of the so called “performance” allele in the Quechua are discussed in chapter 4.

**Ventilation**

Exposure to reduced oxygen availability triggers an increase in ventilation. In sea level natives translocated to 3500 m, total ventilation increased by ~40% over the first two weeks at altitude, primarily due to an increase in ventilatory rate (Chiodi, 1950). The same study found that ventilation was about 10% greater in long term non-native residents. The initial ventilatory response to hypoxia is mediated by the carotid and aortic bodies, which detect the resulting drop in blood oxygen content and signal the respiratory control center in the medulla to increase rate of inspiration. Prolonged exposure to hypobaric conditions appears to blunt the hypoxic ventilatory response (HVR) to further hypoxia in some populations, resulting in a relative hypoventilation compared to sea level natives under the same conditions (Chiodi, 1957). This phenomena is more pronounced in Andeans than in Himalayan populations (Moore et al., 1992) and some researchers have postulated that this is evidence for superior adaptation in the high altitude populations in Tibet (e.g. Beall et al., 1997; Moore et al., 1992).
A blunted hypoxic response may be maladaptive. Hypoventilation can lead to low arterial O₂ pressure (PaO₂) that will, in turn, stimulate erythropoiesis. While the resultant increase in hematocrit (% packed cell volume of blood) may facilitate oxygen transport, the concomitant increase in blood viscosity interferes with blood flow and can contribute to hypertension and other medical complications (Dintinfass, 1981). The potential benefits of reducing factors other than hematocrit that contribute to blood viscosity in high altitude populations are discussed in chapter 3.

Twin studies have demonstrated that there are heritable aspects of the ventilatory response to reduced O₂. A significant correlation was reported for the respiratory response to reduced O₂ (but not increased CO₂) in monozygotic twins that was absent in sex matched dizygotic twins (Collins et al., 1978). In a comparative study of high altitude natives from the Himalayas (Tibetan) and the Andes (Aymara), Beall et al., 1997 reported that there was a heritable contribution to HVR variance in the Aymara (21%) but not in their resting ventilatory rate. In contrast, both parameters were genetically influenced in the Himalayans. Heritable factors accounted for 35% of the variance in resting ventilatory rate (which was 50% higher than that of the Aymara) and 31% of that in HVR. As heritability is required for phylogenetic change, the authors conclude that there is more potential for evolution to modify these characteristics in Tibetans than in the Andeans.

Ventilatory response to acute hypoxia was compared in Aymara neonates (< 5 d.o.) and was similar to that of lowland babies (primarily mestizo). No substantial difference was reported between highlanders and lowlanders until mid-childhood (12 y.o) (Lahiri et al., 1976) suggesting that there is a significant developmental effect.

**Pulmonary diffusion**

Once into the lungs, oxygen passes from the alveoli to the blood by passive diffusion across the pulmonary-vasculature interface. This flow can be defined by the Fick equation as:

\[ \text{VO}_2 = \text{GLO}_2 (\text{PAO}_2 - \text{PaO}_2) \]
where \( \text{VO}_2 \) is the flow, \( \text{GLO}_2 \) is the diffusing capacity and \((\text{PAO}_2 - \text{PaO}_2)\) is the partial pressure differential between the lungs and the blood. When the latter is reduced due to a lower atmospheric \( \text{PO}_2 \), the flow can maintained by increasing the diffusion capacity. By determining the diffusion rate for carbon monoxide (CO), Vincent et al., (1978) demonstrated that diffusion was higher across a range of inspired \( \text{PO}_2 \) values in high altitude natives compared to acclimatized sea level natives due to an increased membrane diffusing component. In a study of Quechua natives exercising at very high altitude (6000 m), Schoene et al., (1990), showed that subjects maintained a relatively high level of blood oxygen saturation despite the low \( \text{PAO}_2/\text{PaO}_2 \) ratio. The ventilatory response was similar to sea level natives exposed to similar conditions however carbon monoxide diffusion rates were significantly higher and the authors concluded that the Andean natives had an increased pulmonary diffusing capacity for oxygen.

**Hematocrit and erythropoiesis**

Oxygen is transported through the blood stream bound to hemoglobin packaged in circulating erythrocytes, or red blood cells (RBCs). Both the blood hemoglobin concentration and the percentage of the blood volume occupied by RBCs (hematocrit) are elevated following exposure to high altitude. During the initial acclimatory response, hematocrits will climb from the average (male) of about 44% to over 50% in about 20 days (Hannon et al., 1969). This is due to increased release of erythropoietin (Epo) by the kidney in response to the reduction in \( \text{PaO}_2 \). Epo is a glycoprotein hormone that accelerates red cell maturation. As acclimation continues to improve oxygen delivery, the \( \text{PaO}_2 \) begins to rise, inhibiting Epo synthesis and therefore limiting further erythrocyte proliferation. However, as the life span of a RBC is about 120 days, the effects of hyperbaric exposure can be evident long after a return to normoxia. This phenomena is sometimes exploited by athletes who, with the hope of increasing their oxygen carrying capacity, sleep at altitude while training at sea level (Bailey and Davies, 1997).

Numerous measurements of hematocrits and/or hemoglobin concentrations have been made in high altitude populations (table 1.1, figure 1.2). An initial impression is that high altitude populations have elevated hematocrits (polycythemia), as do acclimated lowlanders.
While polycythemia may increase the oxygen carrying capacity of the blood, it also increases blood viscosity (Dintinfass, 1985). Increased viscosity impedes blood flow and has been implicated in the development of hypertension, cardiovascular disease and cancer (Dintinfass, 1981). Excessive polycythemia is the basis of chronic mountain sickness (Monge’s disease), which has been described as the loss of adaptation to high altitude (Monge and León-Velarde, 1991).

Table 1.1. Hematocrits in Andean native populations. A summary of values reported for Quechua and Aymara populations living at various altitudes.

<table>
<thead>
<tr>
<th>Altitude (m)</th>
<th>hematocrit (%)</th>
<th>Sex</th>
<th>Population</th>
<th>Location</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>450</td>
<td>38.1</td>
<td>M,F</td>
<td>Aymara</td>
<td>Azapa, Chile</td>
<td>1</td>
</tr>
<tr>
<td>450</td>
<td>42.1</td>
<td>M</td>
<td>Quechua</td>
<td>Santa Cruz, Bolivia</td>
<td>2</td>
</tr>
<tr>
<td>450</td>
<td>41.6</td>
<td>M</td>
<td>Aymara</td>
<td>Santa Cruz, Bolivia</td>
<td>2</td>
</tr>
<tr>
<td>550</td>
<td>41.0</td>
<td>M,F</td>
<td>Aymara</td>
<td>Lluta, Chile</td>
<td>1</td>
</tr>
<tr>
<td>3200</td>
<td>45.6</td>
<td>M,F</td>
<td>Aymara</td>
<td>Tignamar, Chile</td>
<td>1</td>
</tr>
<tr>
<td>3600</td>
<td>50.5</td>
<td>M</td>
<td>Quechua</td>
<td>La Paz, Bolivia</td>
<td>2</td>
</tr>
<tr>
<td>3600</td>
<td>52.0</td>
<td>M</td>
<td>Aymara</td>
<td>La Paz, Bolivia</td>
<td>2</td>
</tr>
<tr>
<td>3700</td>
<td>43.3</td>
<td>F</td>
<td>Quechua</td>
<td>Central Peru</td>
<td>3</td>
</tr>
<tr>
<td>3700</td>
<td>45.7</td>
<td>M</td>
<td>Quechua</td>
<td>Central Peru</td>
<td>3</td>
</tr>
<tr>
<td>3700</td>
<td>52.2</td>
<td>M</td>
<td>Quechua</td>
<td>Ollague, Chile</td>
<td>4</td>
</tr>
<tr>
<td>4000</td>
<td>53.1</td>
<td>M</td>
<td>Quechua</td>
<td>Nuñoa, Peru</td>
<td>5</td>
</tr>
<tr>
<td>4100</td>
<td>49.9</td>
<td>M,F</td>
<td>Aymara</td>
<td>Visviri, Chile</td>
<td>1</td>
</tr>
<tr>
<td>4200</td>
<td>51.4</td>
<td>M</td>
<td>Quechua</td>
<td>Nuñoa, Peru</td>
<td>5</td>
</tr>
<tr>
<td>4260</td>
<td>51.8</td>
<td>M,F</td>
<td>Aymara</td>
<td>Guallatira, Chile</td>
<td>1</td>
</tr>
<tr>
<td>4355</td>
<td>52.9</td>
<td>M</td>
<td>Aymara</td>
<td>Northern Chile**</td>
<td>6</td>
</tr>
<tr>
<td>4355</td>
<td>50.6</td>
<td>F</td>
<td>Aymara</td>
<td>Northern Chile**</td>
<td>6</td>
</tr>
<tr>
<td>4400</td>
<td>53.1</td>
<td>M</td>
<td>Quechua</td>
<td>Macusani, Peru</td>
<td>7</td>
</tr>
<tr>
<td>4460</td>
<td>54.3</td>
<td>M,F</td>
<td>Aymara</td>
<td>Parinacota, Chile</td>
<td>1</td>
</tr>
<tr>
<td>4500</td>
<td>59.5</td>
<td>M</td>
<td>*</td>
<td>Mina Aguilar, Arg.</td>
<td>8</td>
</tr>
<tr>
<td>4540</td>
<td>59.5</td>
<td>M</td>
<td>Quechua</td>
<td>Morococha, Peru</td>
<td>9</td>
</tr>
<tr>
<td>4540</td>
<td>71.1</td>
<td>M</td>
<td>“Indian”</td>
<td>Morococha, Peru</td>
<td>10</td>
</tr>
<tr>
<td>4600</td>
<td>54.3</td>
<td>M,F</td>
<td>Aymara</td>
<td>Caquena, Chile</td>
<td>1</td>
</tr>
</tbody>
</table>

Data is presented graphically in figure 1.2.
* referred to only as “high altitude natives” ** average for four villages
This condition primarily arises in Andean males and is characterized by an elevated hematocrit that in extreme cases can exceed 70%. Sufferers complain of a variety of physical symptoms including dyspnea, insomnia, dizziness and headaches as well as impaired mental faculties (Winslow and Monge, 1987). Congestive heart failure can develop in older patients. Treatment can involve phlebotomy but usually requires that the affected individual move to lower altitudes.

Several studies on high altitude natives have reported very high hematocrits (table 1.1; figure 1.2) and some researchers believe that values this high no longer represent an adaptive response but rather are a pathological manifestation of hypoxic stress exacerbated by lifestyle and environment. Ballew et al., (1989) point out several factors that may have mislead researchers into concluding that very high hematocrits are a characteristic of high altitude populations: 1) many studies are done in mining centers which tend to be populated by transient workers, often from lower altitudes; 2) smoking and chronic respiratory diseases, both of which reduce PaO₂ and contribute to increased hematocrits, are common in urban centers such as Cusco and La Paz and may influence values measured in people who work or live close in these vicinities. Hematocrits taken in traditional, agropastoralist communities, both Andean and Himalayan, are not substantially higher than low altitude populations (Ballew et al., 1989) suggesting that polycythemia may not be an inevitable result of prolonged altitude exposure. It is likely that it is a trend toward sea level hematocrit values, despite living in hypoxic conditions, that reflects adaptation. The possible role of variants in genes involved in the erythropoietic response to hypoxia is discussed in chapter 6.

Other red blood cell indices in Quechua living at 4000 - 4400 m (including mean corpuscular volume (MCV), mean corpuscular hemoglobin content (MCH) and mean corpuscular hemoglobin concentration (MCHC)) have been shown to be consistent with sea level standards (Garruto, 1976). This differs from high altitude animals such as llamas and vicuñas that have a low MCV and MCH (the resultant MCHC remains close to that of humans) (Heath and Williams, 1995). These animals have a low hematocrit (% packed cell volume) but a high RBC count. This allows a greater surface area for oxygen exchange without the viscosity
problems inherent with high hematocrits. Curiously, this ostensible adaptation appears to be a primitive feature of the camilid family and not an evolutionary change in their high altitude relatives.

High O₂ affinity hemoglobin variants have been reported in some hypoxia-adapted animals such as Bar-headed geese (Jessen et al., 1991); however, while similar human variants do exist (e.g. Hb Andrew Minneapolis; Hebbel et al., 1978), there is no evidence that they are common in high altitude human natives.

![Figure 1.2. Hematocrits in Andean native populations. This is a graphic depiction of the published data summarized in table 1.1. Solid squares are male, solid circles are female and open squares are mixed male and female. Standard sea level values (Thomas, 1993) for males and females are indicated by dashed lines and the corresponding range is bracketed.](image)

**Metabolism and fuel preference**

Oxygen is required to sustain metabolic processes. Most of the aforementioned adaptations alleviate the effects of reduced oxygen availability by improving the efficacy of its uptake. Another possible response is to reduce oxygen requirements. Most people traveling to high altitude are advised to rest and refrain from exerting themselves for a couple of days, thereby reducing the metabolic demand for oxygen while they acclimate. Studies of fuel consumption suggest that the brain, which is highly hypoxia sensitive, maintains a significantly lower metabolic rate in high altitude natives compared with sea level natives (Hochachka et al.,
This constitutive hypometabolism is maintained for weeks after exposure to sea level oxygen levels suggesting that it may be influenced by either developmental or genetic factors.

It is also possible to minimize the impact of reduced oxygen availability by preferentially utilizing fuels with a higher ATP yield per oxygen consumed. While the oxidization of fatty acids produces more energy per gram burned, it requires 25-50% more oxygen than catabolizing carbohydrates (Daut and Elzinga, 1989). In myocardial tissue, which is highly metabolically active, the preferred fuel source after a 6-20 hour fast are fatty acids; however in Quechua there appears to be a significantly greater cardiac reliance on glucose as fuel (Holden et al., 1995). Although there was a decrease in cardiac glucose utilization after a three-week exposure to normoxic conditions, the rate of glucose uptake relative to plasma glucose levels remained significantly different in Quechua compared to lowlanders. This may reflect an adaptive change in myocardial fuel preference; however whether this is due to developmental or genetic factors (or both) is currently unknown. It may also be a reflection of life-style. Dietary analyses of Quechua living either at altitude or at sea level found that the highlanders ate a greater percentage of carbohydrates than did the lowland group and had higher serum triglycerides (Watt et al., 1976). Calorie consumption and percent body fat was the same for both groups. The frequencies of genotypes that effect fat utilization in the Quechua are discussed in chapter 5.

1.3 The Peopling of the New World

There is a general consensus that the earliest human inhabitants of the New World were nomadic tribes who had migrated from northeastern Siberia into Alaska across what is now known as the Bering Strait. This crossing was made possible by the exposure the continental shelf due to glacial effects on sea level. This land bridge, known as Beringia, has surfaced numerous times over the last 2 million years (figure 1.3a), most recently during the last major glacial advance that occurred between 13,000 and 23,000 ybp (years before present). During this period an ice free corridor existed between the Laurentide glacier, which covered much of the northern half of the continent east of the Rocky Mountains, and the Cordillarian glacier,
Figure 1.3. Beringia and Pleistocene glaciation patterns. a) The ebb and flow of glacial ice during the last 800 centuries. Dark blocks are periods of glacial advance during which Beringia was exposed and overland travel between Asia and North America was possible. Adapted from Jennings, 1978. b) The approximate extent of glaciation (stippled areas) in North and South America 15,000 years before present (ybp). The gap between the northern ice fields is thought to have served as a passageway to the inviting south. Adapted from Cavalli-Sforza et al., 1996 and Jennings, 1983.
which spread down from the mountains of the west coast (figure 1.3b). One theory is that the migrants moved down this corridor, into the Western plains and then spread south and east, occupying much of the New World in a relatively short period of time. An alternate theory is that the initial migration followed along the west coast by boat. This latter theory could account for the rapidity of the spread; however, as much of the coastline that they would have followed is now submerged, it is unlikely that evidence of such a passage will ever be uncovered.

The exact time of this migration is unknown and is subject to some debate in the paleoanthropological community. Jennings (1983) concludes that the evidence supports an initial colonization at around 35,000 ybp (over a Beringia exposed by the second Mid-Wisconsin advance) that was followed by movement south about 25,000 years ago. Apparent bone tools dating from 27,000 +/- 3000 ybp have been found in western Alaska at Old Crow flats. Several artifacts dating to around 20,000 ybp have been recovered from the Meadowcraft Shelter in Pennsylvania and dates exceeding 20,000 years ybp have been proposed for sites in Peru, Chile and Brazil including artifacts allegedly being as old as 30,000 years. The great antiquity of some of these sites is contentious and some authors, such as Cavalli-Sforza maintain that the incontrovertible evidence of human habitation does not antecede 15,000 ybp (Cavalli-Sforza et al., 1994), consistent with a crossing from Asia during the last glacial period (a series of advances and retreats that lasted about 10,000 years and ended about 13,000 ybp).

Concomitantly, the ice free passage to the inviting south was open and there are numerous archeological sites dating between 14,000 and 12,000 ybp throughout the Americas including the Wilsall (Anzick) site in Montana where carbon-14 dating has dated human remains to 10,000 +/- 300 ybp (Greenberg et al., 1986). In addition, dental divergence supports an Amerindian/Asian separation at about 14,000 ybp (Greenberg et al., 1986) with all native Americans descending from a single northeast Asian ancestral stock. Recent analysis of mitochondrial DNA (mtDNA) variation (Starikovskaya et al., 1998); however, supports an early expansion of Asiatic populations into Beringia. Divergence between mtDNA haplogroups A, C and D suggests that a separation of Amerindians and Siberian aboriginal populations (Chukchi and Siberian Eskimos) occurred approximately 34,000 ybp, a period when both Beringia and the Alberta corridor were
ice free. A more recent divergence time for haplogroup B (13,500 - 17,700 ybp) is proposed to be evidence for a second wave of paleo-Indian influx. Diversity within haplotype A indicates that the north-west coast Amerindians (including the Na-Dene), the Eskimo and the Chukchi separated approximately 7000 -13,000 ybp, probably coinciding with the most recent exposure of Beringia.

The initial migrants, the first people to colonize what long after came to be called the New World, were the Amerindian ancestors of most extant Native American populations. Dental, linguistic and genetic data suggest that this occupation was only the first of three pre-Columbian waves of colonization (reviewed by Greenberg et al., 1986) and that a second migration, occurring shortly after the first, introduced the ancestors of the Na-Dene speaking natives to the New World. The Na-Dene language family includes Athabaskan speaking groups such as the Dogrib and the Slave; non-Athabaskan speakers such as the Haida and Tlingit of the north west coast and Southern Na-Dene such as the Apache and the Navajo of the American south-west. A third wave followed around 10,000 years ago, spreading across the north as far as Greenland, giving rise to the Inuit and, to the west of Alaska, the Aleut.

*Table 1.2. Early archeological sites in the Andes*

<table>
<thead>
<tr>
<th>Site</th>
<th>Proposed age (ybp)</th>
<th>Altitude</th>
<th>Location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pikimachay</td>
<td>22,000*</td>
<td>2850 m</td>
<td>South Central Peru</td>
<td>1</td>
</tr>
<tr>
<td>Huargo Cave</td>
<td>13,460*</td>
<td>4050 m</td>
<td>Central Peru</td>
<td>2</td>
</tr>
<tr>
<td>Pachamachay</td>
<td>11,800</td>
<td>4300 m</td>
<td>West Central Peru</td>
<td>3</td>
</tr>
<tr>
<td>Jayamachay</td>
<td>10,750</td>
<td>3400 m</td>
<td>Central Peru</td>
<td>1</td>
</tr>
<tr>
<td>Pikamachay</td>
<td>12,000</td>
<td>2850 m</td>
<td>South Central Peru</td>
<td>4</td>
</tr>
<tr>
<td>Panaulauca Cave</td>
<td>10,000</td>
<td>4150 m</td>
<td>West Central Peru</td>
<td>2</td>
</tr>
<tr>
<td>Lauricocha</td>
<td>9,500</td>
<td>4050 m</td>
<td>Northern Peru</td>
<td>2</td>
</tr>
<tr>
<td>El Inga</td>
<td>9,030</td>
<td>2800 m</td>
<td>Northern Equador</td>
<td>5</td>
</tr>
</tbody>
</table>

For the location of some of these sites, see inset map in figure 1.5. *earliest evidence of occupancy these dates are not universally excepted (see Lynch, 1990) References: 1) MacNeish (1971); 2) Bruhns (1994); 3) Lynch (1990); 4) Wenke (1984) and 5) MacNeish (1976).

The occupation of South America is believed to have commenced shortly after the arrival of Amerindians into the New World (table 1.2). It may have been by a single group of paleo-
Indians (see Rothhammer and Silva, 1989); however there is no evidence of a major genetic bottleneck occurring during the movement across the Isthmus of Panama (Monsalve et al., 1994). Once again, there is little consensus as to when this occurred (see Lynch (1990) for a critical evaluation of the evidence for early (>12,000 ybp) occupancy of South America).

MacNeish (1976) reviewed the archeological evidence from the Ayacucho Valley and postulated this region as "being the earliest stage in man’s appearance in South America". He proposed that artifacts from cave sites in this valley span a continuous period from 22,000 to 1500 ybp, encompassing the entire pre-Columbian history of the Peruvian highlands from the earliest cultures to the imperial reign of the Inca. Other authors (i.e. Lynch, 1990) are skeptical of both the dating on which these claims are based and on the authenticity of some of the early artifacts, such as the putative lithic tools from the Ayacucho Valley, as man made objects.

Martin (1973) proposed an interesting model of the occupation of the Americas. He theorised that humans crossed the isthmus of Beringia into Alaska about 11,700 ybp. They then traveled down the ice free corridor to the area presently occupied by Edmonton, arriving there about two centuries later. From there they expanded rapidly, spreading south across both continents at a rate of about 16 kilometers per year. This rapid rate of expansion was possible because of an abundance of "naive" megafauna that provided the invaders with ample food. The population density behind this wave front was low (0.04 person/km^2). According to Martin’s theory, this rapid expansion coupled with low density would account for the dearth of paleo-Indian sites dating to the period of earliest expansion. Eventually, the megafauna grew wary (or extinct) and the now disperse paleo-Indians began to settle in regions of better hunting and to develop the classic big game hunting cultures, such as the Clovis point tradition, that suddenly appear on the prehistoric horizon around 10,000 years ago, scattered throughout North America. Supporters of the earlier appearance of man (e.g. MacNeish, 1976) question many of the assumptions made in this model, especially the date of first arrival. However, the model is consistent with most of the undisputed archeological data.

It is beyond the scope of this thesis (and the ability of its author) to even begin to critically examine the data, theories and disputes that comprise current American
paleoanthropology. For the purpose of this thesis, a conservative date of occupation of the Andean altiplano of 12,000 ybp has been chosen as it best accommodates the undisputed archeological data (table 1.2). Assuming a generation time of 20 years, these populations have had approximately 600 generations to adapt to the conditions found at these altitudes, including the approximately 30% reduction in available oxygen. As shown in figure 1.4, this is sufficient time for even small advantages to influence the genetic makeup of a population.

![Graphs](image.png)

**Figure 1.4.** Predicted changes in allele frequencies over 600 generations (approximately 12,000 years) in response to selective advantage. Initial frequencies of allele A are 0.10, 0.25, 0.50 or 0.75. Fitness of genotypes: a) AA = 1.0, Aa = 0.995 and aa = 0.99; b) AA = 1.0, Aa = 0.9975 and aa = 0.995. Model assumes random breeding in an unlimited population. Models generated with PopBio.

### 1.4 The South American highlands

The Andes, which extend from Central America to Terra del Fuego, is the world's longest mountain range. Much of Andean prehistory centers around the broad plateaus of the central region of this vast chain of mountains that stretch from southern Columbia to the northern reaches of Chile (figure 1.5). In Equador, the Andes split into two parallel ranges, the Cordillera Oriental to the west and the towering, snowcapped Cordillera Occidental to the east. In between, at between 2000 and 3600 m, lie the cuencas, fertile highland valleys watered by glacial run-off from the adjoining peaks. As the mountains progress south into Peru, topography grows more complex as a third range, the Cordillera Central, splits the central valley, eventually joining the Occidental range at Cerro de Pasco. The slopes of this central range and the valleys separating it
Figure 1.5. South America and Peru (insert). The lightly shaded areas on the continental map represent the Andes mountain range. Some of the Andean archaeological sites and sites of research discussed in the text are indicated. 1: Jayamachay; 2: Pikimachay; 3: Lauricoch; 4: Huargo Cave; 5 Pachamachay and Panaulaucu Cave. Quechua samples used for DNA preparation in this study were collected from several villages near Ollantaytambo.
from the other two ranges are temperate, well watered and fertile. In these high basins (which include the valley of Ayacucho (figure 1.5), site of many of the early archeological sites) and the *altiplano* into which they merge are found major centers of indigenous population, both contemporary and historic, including the Inca capital of Cusco. The natural vegetation of this region consists of a combination of tropical mountain and *paramo*. The former, depending on altitude, ranges from sub-tropical forests to shrub above the 3500 m tree line while the *paramo* are high altitude grasslands. Much of the natural vegetation in the regions has been replaced by agriculture, with extensive terracing that allows improbably steep fields to ascend high above the green valley floors. To the east of the mountains the land falls away to the Amazon rain forest, to the west, to the dry coastal plains.

South of the central Andes, where the ranges are the most separated, is the flat, arid *altiplano*. Lying between 3600 and 4000 m, this vast plateau stretches from Southern Peru into Bolivia and includes Lake Titicaca and La Paz, the capitol of Bolivia (both at 3800 m). Further south, this plateau transforms into the arid expanse of desert and salt flats called the *puna*, a thinly inhabited, inhospitable plain that extends south to Argentina. At this point the two Andean ranges coalesce into a single range which extends south to eventually merge with the highlands of Patagonia.

### 1.5 The Quechua

The native cultures of South America have been classified into five distinct sociocultural systems. The Quechua and the Aymara represent the Central Andean system, which is characterized by domestication of animals and plants, irrigation, class structure, state formation, elaborate buildings and road construction (see Salzano and Callegari-Jacques, 1988). They are also grouped together by their languages, both of which are classified as South Central Andean however the genetic distance between the two groups (Cavalli-Sforza *et al.*, 1994) suggests that they may be more closely related to other Central and South American populations despite their current proximity.
Quechua was the language of the Inca, the divinely self-appointed ruling class who governed *Tawantinsuyu*, the vast Inca empire that grew out of the fertile highland valleys of South Central Peru. Irrigation and terracing greatly improved the productivity of the region and, in conjunction with the great variety of available resources, supported rapid population growth. In 1438, in a battle near their capital city of Cusco, the Inca defeated their neighboring state and began expanding into an empire that eventually stretched from Southern Colombia to South Central Chile and encompassed 12 million people.

The Inca empire lasted only about 100 years before falling to Pizarro and his 260 *conquistadores* who, more through political cunning than by force of arms, managed to defeat an empire larger than their homeland Spain. In 1535 Pizarro founded Lima, the capital of The Viceroyship of Peru. Despite several Inca rebellions against the European conquerors, the Spanish maintained control of the highlands but, as the high altitude made them uncomfortable and because their primary interest was shipping the wealth out of the colony, the Spanish tended to stay on the coast and, as Lima’s importance grew, Cusco’s faded. To some extent this may have served to preserve the native Andean heritage as the influx of European blood into their region was perhaps less than it would have been if the *altiplano* had been more hospitable. The indigenous people of the Andes, though decimated by the diseases imported, along with gun powder and steel, by the *conquistadores*, survived the onslaught of European civilization better than their North American counterparts. As of 1990, there were an estimated 6.2 million Quechua speaking people living primarily in the highlands of Equador, Peru, Bolivia and Argentina (Caviedes and Knapp, 1995). The extent of the Inca empire is still seen today. Five hundred years after its demise half of the speakers of native American languages speak languages classified as Andean (Cavalli-Sforza *et al.*, 1994) and the Quechua are the largest linguistically defined native population in the New World. Additionally, there are about 1.6 million (as of 1990) Aymara speakers, living primarily between Lake Titicaca and La Paz, Bolivia. Currently, many of the Andean Indians are agropastoralists, herding llamas and cultivating potatoes, quinoa and some grain crops; however some have left this lifestyle for work in the mines or in the cities.
The Quechua are defined by language and may not necessarily represent a homogeneous population. As J. A. Vellard succinctly expressed it: "There is no existing race that corresponds to the actual distribution of the Quechua language" (quoted in Comas, 1971). As the Inca empire expanded, it was the practice of the conquerors to distribute the vanquished throughout the imperial realm, although it is recorded that the Inca were aware of the problems in transplanting lowlanders to the high reaches of the empire. The result of this practice is that the extant Quechua may have heterogeneous ancestors including lowland natives. Subsequent interbreeding with Europeans contributed to this heterogeneity, although it is believed that a combination of physical and cultural factors may have limited the influx of Caucasian genes into the native Andean gene pool. A recent estimate of average Caucasian admixture in contemporary Quechua is 0.247 (Salzano and Callegari-Jacques, 1988).

As was discussed earlier, the Quechua show considerable adaptation to high altitude. Their ancestors have inhabited the high plateaus of the Andes for over 12,000 years, sufficient time to incur detectable changes in allele frequencies even with minor differential fitness (figure 1.4). There is some belief that the Himalayans may be better adapted to altitude than the Andeans (Moore et al., 1992), perhaps due to their longer history of exposure (surmised to be about 30,000-35,000 years or almost thrice that of the Andeans). In most cases, the change in allele frequencies in response to selection is rapid at first (especially when the initial frequency of the favoured allele is low) and, while changes would be more pronounced after 1650 generations, they should still be evident after six hundred (table 1.3). Our principle reason for choosing Andean natives over Himalayan highlanders (e.g. Tibetans and Sherpas) for this study was a practical consideration pertaining to the acquisition of samples.

A confounding issue in identifying adaptive changes in the genetic composition of a population is the background of changes resulting from stochastic forces such as genetic drift. Genetic drift occurs when populations are sufficiently small that there is a significant probability that alleles will be lost due to chance alone (Cavalli-Sforza et al., 1994). Founder effect, in which the genetic composition of a extant population reflects changes that occurred by chance in a small progenitor population, is an example of this.
The resultant loss of alleles will reduce genetic variation in a population and, as preferential selection of such variants is the basis of evolution, population homogeneity will slow the evolutionary process. Given that the amount of time available for evolution to have occurred in the Andean populations is, by evolutionary standards, brief, the arguments for such changes having occurred will be strengthened if there is evidence for heterogeneity in the founding populations(s). Kidd *et al.*, (1991) concluded that, given the heterogeneity of nuclear polymorphisms, there was not an excessive loss of heterozygosity during the populating of the Americas. Furthermore, there is mtDNA evidence that a significant genetic bottleneck did not occur during the human expansion into South America (Monsalve *et al.*, 1994) arguing that this variability may have been maintained as humans spread toward the Andes.

*Table 1.3. A comparison of predicted change in allele frequencies after 600 and 1650 generations (approximately 12,000 and 33,000 years respectively).*

<table>
<thead>
<tr>
<th>Initial frequency of A</th>
<th>Final frequency of A after 600 generations</th>
<th>Final frequency of A after 1650 generations</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>0.69</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>0.25</td>
<td>0.87</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>0.50</td>
<td>0.95</td>
<td>1.0</td>
</tr>
<tr>
<td>0.75</td>
<td>0.98</td>
<td>1.0</td>
</tr>
<tr>
<td>0.90</td>
<td>0.99</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Based a co-dominant inheritance pattern with genotypic fitness of 0.99 for aa, 0.995 for Aa and 1.0 for AA. Model assumes unlimited population and an inbreeding coefficient = 0.0.

1.6 Summary

There is little doubt that humans can successfully adapt to the hypoxic conditions manifest at high altitudes. Millions of people live between 3000 and 4500 m in the high Andean plateaus and have done so for centuries. While acclimation will allow lowlanders to function comfortably at these altitudes, it is evident that having been born and/or raised in the highlands confers a substantial advantage. While it has not been conclusively proven that high altitude Andean populations are genetically adapted to hypoxia, there is substantial evidence for heritable components contributing to some of the anatomical and physiological changes that are believed to facilitate the uptake, transport and utilization of oxygen (summarized figure 1.6).
The Quechua are an excellent population in which to search for evidence of selective pressures altering allele frequencies in genes that may be involved in high altitude adaptation. They live high in the mountains, breathing air with 33% less oxygen than that to which sea level natives are accustomed. The archeological evidence supports an occupation of the altiplano for at least 600 generations, ample time for selection to favour alleles conferring even a nominal advantage, and molecular evidence suggests that the ancestral population was genetically heterogeneous so that there was variation on which selection could act.
1.7 Rationale

For a species that has spread through virtually every terrestrial environment on the planet humans are remarkably invariant, minor changes in pigmentation and stature not withstanding. We have however, always been more interested in our slight differences rather than our great commonalities and, for both good and bad, the quantification of human diversity has a long history in science. Interest in variety accelerated in the late nineteenth century with the publication of Darwin’s theories of natural selection as people began to realise that given evolutionary time scales, apparently trivial differences in individuals could eventually alter the makeup of entire populations, provided that the variation could be passed on to subsequent generations. Trans-generational transmission separates heritable characteristics from those acquired during an individual’s lifetime thus establishing the dichotomy of nurture and nature that, ultimately, is the source of all biological diversity.

The work presented in this thesis represents one approach to understanding the role of genetics in the adaptation of humans to the hypoxic conditions found at high altitude (3000 - 4000 m). While previous work in this field has focused on the heritable component of continuous traits, such as chest morphology and blood oxygen saturation, this project focuses on changes in the frequencies of variations within single genes. Given the established (or potential) phenotypes of these variants and the characteristics, or theoretical requirements, of high altitude populations, it is possible to make testable predictions about these frequencies in these populations. Essentially, this is a variation of molecular association analysis, a technique that has been used successfully to identify potential genetic factors in the development of complex diseases such as Alzheimer’s disease and myocardial infarction as well as numerous conditions that have been associated with human leukocyte antigen (HLA) genotypes (see Lander and Schork, 1994). Clinical association studies test whether a particular allele occurs at a higher frequency among affected than unaffected individuals (Lander and Schork, 1994). In this study, the comparison is between individuals who represent a population that is proposed to be adapted to a particular environment and individuals from populations not exposed to those conditions. If an association is observed it may be due to: 1) the over-represented allele having been selected
for in the population because it confers some benefit; 2) an allele to which the over-represented variant is in linkage disequilibrium has been selected for; or 3) random variation in the population. The latter possibility is discussed further in the final chapter of this thesis.

Testing large numbers of polymorphisms raises the issue of observing chance associations. This can be corrected for but may involve such low \( P \) values (Xu et al., 1999) that unrealistically large sample sizes are required to detect significant changes. One was to circumvent this is to chose candidate genes that, by definition, have a prior probably of being involved in the genotype (Cox and Bell, 1989).

The candidate genes for this study were chosen because they have variants associated with phenotypes of potential benefit at altitude or because they encode proteins directly involved in the response to changes in oxygen availability. They are: 1) the \( \beta \)-fibrinogen gene, the product of which is a major determinant of plasma viscosity; 2) the angiotensin converting enzyme (ACE) gene, variants of which have been associated with high altitude performance and 3) the \( \beta_2 \)-adrenergic receptor gene (\( \beta_2 \)-AR), which encodes a major pulmonary catecholamine receptor. All of these genes have previously characterised polymorphisms and will be discussed further in subsequent chapters. In addition, the genes encoding erythropoietin and hypoxia inducible factor 1\( \alpha \) were examined. As of this writing no reports of functional polymorphic sites in these genes have been published; however, their functions in the production of red blood cells and in oxygen sensitive gene regulation make them prime candidates for a role in the development of hypoxia tolerance.

The high altitude population chosen for study are Quechua Indians of the central Andes. As previously discussed, these hardy highland people have inhabited the altiplano for more than 120 centuries and show considerable adaptation to life on the high plateaus. Additionally, having been the subject of scientific investigation for almost a century, they are a relatively well studied population and there is a large body of literature describing the anatomical and physiological characteristics that seem to mitigate the deleterious effects of lifelong exposure to hypobaric hypoxia.
Chapter 2

Materials and Methods\(^1\)

*General methodology*

DNA samples representing three populations 1) Quechua speakers living at above 3200 m on the Peruvian *altiplano*, 2) Na-Dene speaking Amerindians from the coastal regions of British Columbia and 3) Caucasians of Western European descent, were prepared from peripheral blood leukocytes and/or buccal epithelial cells. These samples were genotyped for known polymorphisms by polymerase chain reaction (PCR) amplification usually followed by digestion with a diagnostic restriction endonuclease. The genes assayed were selected for analysis because they encode products that are involved in the uptake, transport or utilization of oxygen and, in most cases, they contained known polymorphisms with previously characterized phenotypes. Allele frequencies at these loci were determined and compared between the populations. In cases where there were multiple polymorphisms in the same gene, it was determined if the alleles at these loci were inherited independently of one another or whether they were in linkage disequilibrium and therefore were moving between generations as haplotypes (sets of co-segregating alleles at linked loci) rather than as independent alleles.

It was hypothesised that, due to selection favouring the maintenance, and therefore the inter-generational transfer of any allele that contributed to an advantageous phenotype at altitude, such alleles would be over-represented in the Quechua. Association between genotype and phenotype has been used for years to characterize the role of genetics in the development of physical characteristics, and it has been proposed that such association analysis is the most effective method of unraveling the genetic basis of complex diseases (Risch and Merikangas, 1996). In the research presented herein, the phenotype in question is adaptation to high altitude existence in the Quechua, and the associations being tested are between this and the genotype at a

\(^1\) The materials and methods described in the following section apply to all subsequent chapters. Specific assay techniques, including PCR conditions are detailed in each chapter as necessary. All PCR primers were prepared by N.A.P.S. at U.B.C. Primer sequences are given in appendix ii.
number of candidate genes selected for their role (albeit sometimes indirect) in oxygen utilization.

**Sample collection**

*The Quechua*

Blood and/or buccal epithelial samples were obtained with informed consent from Quechua living in four small communities (Huiloc, Patacancha, Platerylloq and Qqelcqanqa) located between 3200-4200 m above sea level near the town of Ollantaytambo, Peru (see figure 1.5). The first two sets of blood samples (9 in October, 1997 and 13 in April, 1998) were collected near Patacancha by Charo Tapia (Dpto. de Ciencias Fisiológicas, Universidad Cayetano Heredia, Lima, Peru) and shipped to the University of British Columbia (U.B.C.) in Vancouver, B.C., Canada for DNA preparation and analysis. In October of 1998, a small expedition, consisting of the author and Dr. Vicky Monsalve of the Department of Anthropology (U.B.C.) traveled to Peru where they met Ms. Tapia and collected an additional 66 blood samples and 17 buccal samples. These samples were stored at 4°C until they could be shipped through Lima to Vancouver. The age and gender of each donor was recorded as well as information pertaining to smoking history and recent travel to higher or lower altitudes (see appendix i). Subject relatedness was established by interview or by surname comparison. The initial blood samples were drawn from the anticubidal vein into 20 ml syringes. The samples taken in October of 1998 were initially intended to be drawn into Vacutainer tubes (Becton-Dickinson, Franklin Lakes, NJ) but instead were collected in 20 ml syringes and then transferred to the Vacutainer tubes (containing either Na-Citrate or EDTA as an anti-coagulant). This step proved to be essential as, at 4200 m, the pressure differential between ambient and the interior of the Vacutainer tube was insufficient to draw blood effectively. Buccal samples were obtained using the sugar mouthwash procedure described in (Spitz et al., 1996). Subjects were asked to vigorously swish a 4% sucrose solution in their mouth for a minute before spitting it into a sterile collection cup. In exchange for their time and cooperation, blood donors were given
$10.00 (U.S.) worth of foodstuffs (e.g. pasta, cooking oil, flower, rice and beans) that had been purchased in Ollantaytambo.

Hematocrits were determined by centrifugation of 10 μl of blood in a hematocrit capillary tube for five minutes at approximately 12,000 g followed by measurement of the packed cell volume and the total volume. The percent-ratio of these two values is the hematocrit. For 13 of the initial samples, this was performed in Dr. Dana Devine’s lab in the Department of Pathology and Laboratory Medicine at U.B.C. For the samples collected in October of 1998, hematocrits were determined in the clinic in Ollantaytambo using equipment generously provided by the staff.

Gender frequencies, average age and average hematocrit for Quechua blood donors are given in table 2.1. Not all samples were used for genotype analysis due to lack of family history or because the donor had a first-degree relative already represented in the analysis.

Table 2.1. Age and hematocrit of the Quechua who donated blood samples.

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>Average age (years)</th>
<th>Average hematocrit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>61</td>
<td>31.4 ± 13.5</td>
<td>43.3% ± 5.0%</td>
</tr>
<tr>
<td>Male</td>
<td>21</td>
<td>33.2 ± 11.4</td>
<td>45.7% ± 4.4%</td>
</tr>
<tr>
<td>Total</td>
<td>82</td>
<td>31.8 ± 13.0</td>
<td>44.3% ± 5.4%</td>
</tr>
</tbody>
</table>

Table includes individuals who were excluded from subsequent genotype analysis due to lack of family history or because they had a first degree relative already included in the analysis. ± standard deviation

In addition to the blood and buccal samples, five Quechua lymphoblast cell lines (GM11197, GM11198, GM11199, GM11200, GM11201) were purchased from the Coriell Cell Repository Human Diversity Collection (Coriell Cell Repositories, Camden, N.J.). These cell lines, which were established from Quechua living in central Peru, were originally deposited by Dr. Ken Kidd (Yale University, CT) and were among those described by Barr and Kidd (1993). The original blood samples from which the cell lines were derived were obtained from unrelated individuals (L. A. Giuffra, Washington University, personal communication).
The Na-Dene

Na-Dene is a family of native American languages comprising Haida, Tlingit, Eyak and the Athabaskan family. The first three are single languages of natives from the Northwest coast of Canada and the southern coast of Alaska. The Athabaskan languages are considerably more disperse and are spoken by many bands in Alaska and western Canada as well as by the Apache and Navajo of the American Southwest. Na-Dene is related to Ket, a central Siberian language (Ruhlen, 1998), and not to other native American languages. It is believed to have derived from the language spoken by the second wave of Asian migrants who entered North America approximately 10,000 years ago. The DNA samples used in this project were provided by Dr. M. V. Monsalve (Department of Anthropology, U.B.C.) and have been previously described (Monsalve et al., 1998).

Caucasians

Blood and/or buccal samples were obtained with consent from various unrelated members of the university community at Stanford University, Case Western Reserve University and U.B.C. The subjects were primarily American or Canadian of Western European descent however detailed family histories were not obtained.

DNA and RNA preparation

Lymphocytes were prepared from the blood samples by hypotonic cell lysis. Five volumes of hypotonic RBC lysis buffer (130 mM NH₄Cl, 0.9 mM NH₄CO₃, pH 8.0) was combined with the blood sample in a 50 ml tube and centrifuged for 5 min. at 6000 g in a Jouan MR 1812 centrifuge. The pellet (primarily white blood cells) was washed with 5 ml of lysis buffer, centrifuged again, rinsed with 5 ml of isotonic saline (0.85% NaCl) and, following a final centrifugation, were resuspended in 3 ml of saline/EDTA (75 mM NaCl, 24 mM EDTA), 1% sodium dodecyl sulfate and 1 mg proteinase K (E. Merck Darmstadt, Germany) and incubated overnight at 56°C. DNA was isolated by phenol/chloroform extraction as follows. The lysate
was mixed thoroughly with one volume of Tris-buffered phenol (pH 8.0) and centrifuged at maximum speed in a International Clinical Centrifuge (model CL) for 5 min. The aqueous phase (upper) was then transferred to a fresh tube and this process repeated once with a 50:50 mixture of phenol/chloroform and twice with chloroform. Two volumes of ethanol were added to the final lysate to precipitate the DNA that was then removed either by manual spooling around the tip of a Pasteur pipette or by centrifugation. The DNA was allowed to dry and then resuspended in 400 µl H₂O. DNA was prepared from buccal samples in a similar manner (Spitz et al., 1996). Purified DNA was quantified at 260 nm on a Pharmacia BioTech spectrophotometer. DNA yield from a single tube of blood was approximately 200 µg and average recovery from the buccal samples was approximately 88 µg although this ranged widely (12 - 320 µg). Following quantification, the DNA was aliquoted into 10-100 ng/µl samples for PCR and the remainder stored at -70° C.

Lymphoblast cell lines were propagated and maintained in Dr. C. J. Brown’s lab in the Dept. of Medical Genetics, U.B.C. by Sarah Baldry. DNA, RNA and cDNA were prepared from these cell lines using established techniques (Miller et al., 1988; Chomczynski, 1989).

Chimpanzee DNA was kindly provided by Dr. C. J. Brown.

**Genotyping**

Most of the polymorphisms were assayed by restriction fragment length polymorphism (RFLP) analysis. The details of specific analyses are given in the appropriate chapters. In general, the region spanning the polymorphic site was amplified by PCR and the product digested with a diagnostic restriction enzyme. Digestion products were electrophoresed on 8% polyacrylamide gels run in Tris-borate-EDTA buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA) and visualized by ethidium bromide (EtBr) staining. The sizes of amplification and digestion products were estimated by comparison to a 100 basepair ladder (GibcoBRL, Gaithersburg, MD) run on the same gel. For data presentation, photographs were digitized using an Agfa Arcus II scanner and incorporated into Adobe Photoshop images.
DNA Sequencing

DNA sequencing was performed by the Nucleic Acid-Protein Service Unit (NAPS) at the U.B.C. PCR product was used as a template. Conditions for PCR amplification are described in subsequent chapters. Four identical 25 μl amplification reactions were pooled, electrophoresed on a 2% agarose gel run in tris-acetate-EDTA buffer (40 mM Tris acetate, 2 mM EDTA) and the product visualized by EtBr staining. The amplification product was cut out of the gel and purified over QiaQuick columns (Qiagen Inc., ON, Canada). Ninety to 200 ng of this product and an appropriate primer (3.2 pmol/μl) was then sent to NAPS for di-deoxy sequencing. DNA sequence data were analyzed using a number of sequence programs including DNA Strider 1.0 (Institut de Recherche Fondamentale, France) that was used for restriction enzyme site mapping and BLAST (Basic Local Alignment Search Tool) that was used for sequence alignment. BLAST is available through the National Center for Biological Information (NCBI) at http://www.ncbi.nlm.nih.gov/. Sequences used for comparison were obtained on-line through GenBank, which was also accessed through the NCBI site. Ambiguous bases were resolved by re-sequencing a new template preparation, usually on the other strand.

Statistical analysis

Changes in allele frequencies over time were estimated using PopBio 2.4, Population biology simulations for Macintosh (©Eric Bowman and the Reed Institute, 1990-1995, available at http://web.reed.edu/academic/departments/biology/software.html). Allele frequencies were established by gene counting and compared either by two way contingency tables using StatView SE software (Abacus Concepts Inc., Berkeley, CA) or, if expected values were less than five, by Fisher’s exact test (Siegel and Castellan Jr., 1988). Tests for Hardy Weinberg equilibrium (modified Markov-chain random walk algorithm) and, when multiple polymorphisms were examined in a single gene, linkage disequilibrium analysis (likelihood-ratio test for phase unknown genotypic data) were performed using Arlequin software (Schneider et al., 1997) available at http://anthropologie.unige.ch/arlequin/).
Chapter 3

β-fibrinogen allele frequencies in the Quechua

Introduction

Elevated hematocrits are a characteristic response to high altitude hypoxia in humans. This is true for both recently acclimated low-landers and for natives who have spent their lives in the mountains. Reported hematocrit values in Quechua men living at over 3000 m often exceed 50% (table 1.1 and figure 1.2). While these values are not outside of the normal range for males at sea level (40% - 54%), they tend to be higher than the sea level average (47%; Thomas, 1993).

The advantage to having a high hematocrit in hypoxic conditions is that the resultant increase in hemoglobin concentration increases the oxygen carrying capacity of the blood, thereby facilitating oxygen transport to the tissue. There is however a potential disadvantage to this strategy: increased hematocrits result in elevated blood viscosity (figure 3.1) which in turn has been associated with a number of cardiovascular disorders including myocardial infarction, and coronary thrombosis as well as circulatory problems such as hypertension (Dintinfass, 1981). Hyperviscosity may also contribute to the etiology of chronic mountain sickness, a polycythemic condition that occurs in high altitude natives (primarily Andean) for which the only long term cure is descent to lower altitudes (Winslow and Monge, 1987). Whole blood is a non-Newtonian fluid and its viscosity is dependent on both intrinsic factors, such as hematocrit and plasma composition as well as on extrinsic conditions such as the size of the vessels through which it is flowing. Furthermore, blood cells are flexible fluid filled structures and do not behave as rigid particles in a flowing solution (fig 3.1). These characteristics allow blood to flow at hematocrits above 65% when, if it were a suspension of rigid particles, it would, as Dintinfass expressed it, “achieve the consistency of concrete” (Dintinfass, 1985). Despite these characteristics, the increase in viscosity associated with high hematocrits impedes blood flow by

---

increasing resistance. This phenomenon may be exacerbated at altitude due to the decrease in vessel diameter resulting from hypoxic vasoconstriction.

Blood viscosity is also affected by the concentration of various plasma proteins. The major determinant of plasma viscosity is fibrinogen (Lowe et al., 1993), a common plasma protein (1.9 - 4.5 mg/ml) that is the precursor to fibrin, the primary component of blood clots (figure 3.1). Fibrinogen levels are genetically regulated, and several polymorphisms affecting these levels are located in, or near, the gene encoding the beta chain of fibrinogen (β-fibrinogen).
This polypeptide is one of three (along with an alpha and gamma) that comprise fibrinogen and is the rate determining component in holoprotein assembly (Yu et al., 1984). A guanosine (G) to adenosine (A) transition at base -455 (sometimes reported as being at base -453) is associated with an ~10% increase in fibrinogen in both A/A homozygotes and A/G heterozygotes (Thomas et al., 1991). A second upstream mutation, a cytosine (C) to thymidine (T) transition at base -148, is associated with a similar increase (+10% in T/T homozygotes, +6% in heterozygotes (Heinrich et al., 1995). Within the gene, a G to A transition at base 448, which substitutes an arginine residue for a lysine (Schmelzer et al., 1988) has been associated with a 33% increase in plasma fibrinogen concentration in pooled homozygotes and heterozygotes (Carter et al., 1997). The phenotype varies (some studies, such as Conner et al., (1992) found no significant effect) and appears to be influenced by gender (Carter et al., 1997; de Maat et al., 1995) and by environmental conditions such as exercise (Montgomery et al., 1995). The alleles associated with increased fibrinogen are fairly common in most of the populations tested thus far, ranging from 11% in the Inuit to over 30% in some Caucasian groups (de Maat et al., 1995).

Figure 3.2 The role of fibrinogen in the clotting cascade. Fibrinogen is the precursor to fibrin, the actual clot forming substance. The beta subunit of the fibrinogen multimer is the limiting component in fibrinogen assembly.
Selective pressure in high altitude populations to ameliorate the detrimental effects of higher hematocrits on blood viscosity may be reflected in a low frequency of the alleles associated with increased fibrinogen relative to populations residing at, or near, sea level. To test this hypothesis, the β-fibrinogen genotypes of 60 Quechua were determined as were the genotypes of the two lowland populations. Relative plasma viscosity was also determined for a subset of the Quechua and a small sample of Caucasians.

**Materials and Methods**

**Genotyping**

Polymorphisms in the β-fibrinogen gene were identified by restriction endonuclease digestion of PCR amplified DNA. Primer product size and diagnostic enzymes are shown in table 3.1. In addition, the genotype for a Rsa I polymorphism in the H19 gene was determined in a subset of the Quechua as a test of heterogeneity. Primer sequences are given in appendix ii. Primers Fib-B6 and Fib-B7 have previously been described (Baumann and Henschen, 1993a). The primers in H19 were a gift from Dr. C. Brown (U.B.C.) and have been previously described (Jinno et al., 1995). DNA (50-200 ng) was amplified in a Perkin Elmer DNA Thermal Cycler using 0.033 nmoles of each primer and 0.625 units Taq polymerase (GibcoBRL, Gaithersburg, MD). The final reaction mixture (25 µl) was 0.2 mM dNTPs, 1.5 mM MgCl₂, 20 mM Tris/Cl pH 8.4 and 50 mM KCl. Amplification conditions were 94°C, 1 min.; 58°C, 1 min.; 72°C, 2 min. for 40 cycles. Amplified product (10 µl) was then digested using 0.2 -1.0 units of the appropriate restriction endonuclease under the conditions prescribed by the manufacturers and separated on 8% polyacrylamide gels (figure 3.3).

Chimpanzee (*Pan troglodytes*) DNA was amplified using the same primer pairs and similar conditions to those used for human genotyping (annealing temperature was lowered to 56°C). Amplification product was purified and sequenced on both strands using the primers Fib-1A, Fib-1B, Fib-1C, Fib-1Arev, Fib-B6 and Fib-B7.
Figure 3.3. Assay for RFLPs in the β-fibrinogen gene. Representative polyacrylamide gels showing separation of digestion products for the RFLPs used to identify alleles of the β-fibrinogen gene. Genotypes are given across the top of each panel. The polymorphism and the restriction endonuclease used to characterize it are given across the bottom. The band indicated by * is a primer artefact that did not interfere in the determination of genotype. Fragment sizes are in base pairs.
Table 3.1: $\beta$-fibrinogen PCR primers. Product size and location of primers used for sequencing and polymorphism analysis of the $\beta$-fibrinogen gene.

<table>
<thead>
<tr>
<th>Primer*</th>
<th>Product size</th>
<th>Polymorphism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fib-IC : Fib-IArev</td>
<td>204 bp</td>
<td>G/A$^{455}$ (Hae III)</td>
</tr>
<tr>
<td>Fib-1A : Fib-1B</td>
<td>339 bp</td>
<td>C/T$^{448}$ (Hind III)</td>
</tr>
<tr>
<td>Fib-B6 : Fib-B7</td>
<td>314 bp</td>
<td>G/A$^{448}$ (Mnl I)</td>
</tr>
</tbody>
</table>

*Primer sequences are given in appendix ii.

Viscosity measurements

Relative plasma viscosity was determined using a Canon 100 viscometer (kindly provided by Dr. D. Brooks, Dept. of Chemistry, U.B.C.) suspended in a five gallon water bath. Temperature was maintained at between 25° C and 27° C, although relative viscosity is not significantly affected by temperatures between 15°C and 40°C (Duck, 1990). After being separated from the blood cells by centrifugation (2000 rpm, 10 min.), a sample of plasma was loaded into the viscometer and allowed to equilibrate for 2 min. prior to testing. The sample was then drawn up into the viscometer reservoir by manual suction, released and timed as it flowed through the apparatus. Three separate measurements were made and averaged for each sample. Relative viscosity was determined by comparing the elapsed time of flow through the viscometer to that of distilled water measured under identical conditions. To avoid dilution of the plasma, samples for which viscosity was to be measured were collected in Vacutainer tubes with dry anticoagulant (EDTA powder). All of the viscosity measurements were made in the Regional Hospital in Cusco, Peru where laboratory space and equipment had kindly been made available by Dr. J. Ponce de Leon of the Laboratorio Clinico.

Results

$\beta$-fibrinogen genotypes

Genotype and allele frequencies for the three $\beta$-fibrinogen polymorphisms in the Quechua, Na-Dene and Caucasians are given in table 3.2. In the Quechua, the allele frequencies for all three polymorphisms differ significantly from those of both the Na-Dene and the
Caucasians after correction for number of polymorphisms tested (3) \((i.e. \, P<0.017 \, for \, 1 \, d.f.)\).
The frequencies in the Caucasian sample are not significantly different from those of the Na-
Dene or from those previously reported in the literature for Caucasians \((e.g. \, \text{Thomas et al., 1994;}\)
\(\text{Montgomery et al., 1995; Carter et al., 1997})\). Genotype frequencies were in Hardy-Weinberg
equilibrium in all three samples. There was significant linkage disequilibrium between alleles at
all three loci \(\text{table 3.3) consistent with the haplotypes \(G^{455/C^{148}}/G^{448}\) and \(A^{455/T^{148}}/A^{448}\).\ The
allele frequencies for the \(Rsa \, I\) polymorphism in the \(H19\) gene in a subset of the Quechua \(n =14\) was 0.64 \((T): 0.36 \,(C).\) Allele frequencies for the \(\beta\)-fibrinogen polymorphisms are
available online through The Allele Frequency Database (Alfred), which is maintained by Dr.
Ken Kidd's lab at Yale University (New Haven, CT). The URL for this site is:
http://alfred.med.yale.edu/alfred/index.asp.

Seven hundred and six bases of the chimpanzee \(\beta\)-fibrinogen gene were sequenced \(\text{see appendix iiia).\ The chimpanzee had a G at base -455, a T at base -148 and an A at base 448.\ The
sequences were highly conserved; aside from the human polymorphic sites there were only
two discrepancies between the sequences. Sequence data are available through GenBank
\(\text{accession Numbers AF200354, AF200355 and AF200356).}\)

The relative plasma viscosities of 34 Quechua plasma samples were determined and
compared with those of three Caucasian researchers sampled and assayed under the same
conditions \(\text{figure 3.4). The values between the two groups were similar (Quechua: 1.85, s =}
0.08; Caucasians: 1.81, s = 0.08). Both were lower than the normal mean \(2.01\) but fell within
the normal range \(1.76 - 2.35; \text{Duck, 1990).}\)

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure3_4.png}
\caption{Relative plasma viscosity of Caucasians (n = 3) and Quechua (n=34). The
normal mean (thick dashed line) and range (95%, thin dashed line) are shown (Duck, 1990).}
\end{figure}
Table 3.2. β-fibrinogen genotype and allele frequencies for the Quechua, Na-Dene and Caucasian samples

<table>
<thead>
<tr>
<th></th>
<th>G/A&lt;sup&gt;455&lt;/sup&gt; genotype</th>
<th></th>
<th>G/A&lt;sup&gt;448&lt;/sup&gt; genotype</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>G/G</td>
<td>G/A</td>
<td>A/A</td>
<td>Population</td>
</tr>
<tr>
<td>Quechua</td>
<td>58</td>
<td>2</td>
<td>0</td>
<td>Quechua</td>
</tr>
<tr>
<td>Na-Dene</td>
<td>37</td>
<td>10</td>
<td>3</td>
<td>Na-Dene</td>
</tr>
<tr>
<td>Caucasian</td>
<td>21</td>
<td>10</td>
<td>0</td>
<td>Caucasian</td>
</tr>
</tbody>
</table>

The Quechua are significantly lower than both the Caucasians and the North American natives (Na-Dene) (P<0.016 for 1 df). The North American (Na-Dene) native frequencies are not significantly different from Caucasians (P<0.05 for 1 df).
Table 3.3: Linkage disequilibrium between polymorphisms in the β-fibrinogen gene in Quechua, Na-Dene and Caucasians.

<table>
<thead>
<tr>
<th>Polymorphic locus</th>
<th>G/A&lt;sup&gt;455&lt;/sup&gt;</th>
<th>C/T&lt;sup&gt;148&lt;/sup&gt;</th>
<th>G/A&lt;sup&gt;448&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>G/A&lt;sup&gt;455&lt;/sup&gt;</td>
<td>*</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>C/T&lt;sup&gt;148&lt;/sup&gt;</td>
<td>+</td>
<td>*</td>
<td>+</td>
</tr>
<tr>
<td>G/A&lt;sup&gt;448&lt;/sup&gt;</td>
<td>+</td>
<td>+</td>
<td>*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Polymorphic locus</th>
<th>G/A&lt;sup&gt;455&lt;/sup&gt;</th>
<th>C/T&lt;sup&gt;148&lt;/sup&gt;</th>
<th>G/A&lt;sup&gt;448&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>G/A&lt;sup&gt;455&lt;/sup&gt;</td>
<td>*</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>C/T&lt;sup&gt;148&lt;/sup&gt;</td>
<td>+</td>
<td>*</td>
<td>+</td>
</tr>
<tr>
<td>G/A&lt;sup&gt;448&lt;/sup&gt;</td>
<td>+</td>
<td>+</td>
<td>*</td>
</tr>
</tbody>
</table>

Pairs denoted by + are in significant linkage disequilibrium (p<0.01).

Discussion

The frequencies of the β-fibrinogen alleles A<sup>455</sup>, T<sup>148</sup> and A<sup>448</sup> were significantly lower in the Quechua sample than in both the Na-Dene and the Caucasian samples. As these alleles are associated with higher levels of fibrinogen, this under-representation may serve to mitigate the deleterious effects of the chronic polycythemia characteristic of high altitude populations by limiting the contribution of fibrinogen to blood viscosity. Reduced fibrinogen levels would be particularly beneficial in these populations as the effects of high fibrinogen levels on blood viscosity (Mayer et al., 1966) and on erythrocyte aggregation (Tanahashi et al., 1989) are exacerbated by high red blood cell counts. High fibrinogen concentrations and, to some extent, fibrinogen genotype have been implicated in the development of cardiovascular disease (Heinrich and Assmann, 1995; Carter et al., 1997). A paucity of the alleles associated with higher concentrations of fibrinogen may also contribute to the relatively low incidence of systemic
hypertension (Heath and Williams, 1995) and heart disease (Way, 1976) reported in the Quechua.

Initially, it was a concern that the limited variation at the β-fibrinogen locus that was observed in the Quechua was simply due to genetic homogeneity in the sample population. To address this possibility, the frequency of a common polymorphism in H19, a gene that encodes a functional RNA expressed in the fetus and placenta (Zhang and Tycko, 1992) and is non-syntenic to the β-fibrinogen gene, was determined. The subset of Quechua DNAs that was analyzed was heterogeneous for this polymorphism, suggesting that the low frequency of a subset of β-fibrinogen alleles in our Quechua was not simply due to genetic homogeneity. Additionally, heterogeneity was subsequently demonstrated in two other genes (those encoding angiotensin converting enzyme and the β2-adrenergic receptor) for the same sample population (Rupert et al., 1999; see chapters 4 and 5).

Each of the alleles associated with increased fibrinogen was quite rare in our Quechua sample (two copies in 120 chromosomes). Assuming that Caucasian admixture in the sample was similar to that previously reported for contemporary Quechua (0.247; Salzano and Callegari-Jacques, 1988) and a frequency of the less common alleles in Caucasians of 0.16, one would expect approximately four rare alleles in our sample due to genetic admixture alone. This raises the possibility that these alleles were a post-Colombian addition to the Quechua gene pool or that selection is sufficient to eliminate them as they are introduced. If the latter were true, one would expect that the alleles would also be rare in Hispanics whose ancestors have been living on the altiplano. Clearly, more data are needed to address this issue, especially pertaining to admixture in the communities from which the Quechua volunteers lived and the frequency of this allele in other South American populations, both of indigenous and colonial descent.

Both our data for the Na-Dene and data published for the Inuit (de Maat et al., 1995) suggest that the β-fibrinogen allele frequencies that was observed in the Quechua are not characteristic of Native North Americans in general. It is possible that selection has acted on the Quechua, subsequent to their original migration into the highlands, to reduce the frequencies of the alleles associated with higher fibrinogen so as to mitigate the deleterious effects of elevated
hematocrits on blood viscosity. Although several studies have reported elevated hematocrits in Andean natives (see table 1.1 and figure 1.2), as discussed in the introduction and in chapter 6 some researchers believe that very high hematocrits can be accounted for by factors other than altitude (see Garrutto and Dutt, 1983; Ballew et al., 1989). The hematocrits measured for the individuals who donated blood for this study (table 2.1) were close to average sea level values. While this obviously blunts the selective pressure that we postulated to have favoured the alleles that were over-represented in the Quechua, it does not exclude the possibility of selection. It is possible other compensatory mechanisms have arisen in the Andeans that limit the need for secondary polycythemia but that, prior to this event, the reduced fibrinogen phenotype was advantageous and thus selected for.

The issue of whether the prevalence of these alleles are characteristic of the Quechua and not of South American natives in general cannot addressed without comparing the Quechua to closely related, low altitude control populations. The ancestors of both the Na-Dene and the Inuit are believed to have arrived in North America during different migration waves from that which included the Quechua antecedents (Cavalli-Sforza et al., 1994) and these two groups may genetically quite distant from the Quechua. If so, significant differences in allele frequencies due to stochastic forces would to be expected. Countering this, recent analysis of HLA genotypes (Monsalve et al., 1999) however, suggests that the colonization may have been a single event, and therefore, the divergence in the populations may not be as great as initially thought.

The mechanisms by which the polymorphisms affect B-fibrinogen synthesis are unknown. Both upstream mutations are in the putative 5’ regulatory region of the gene and have been reported to alter protein:DNA interactions in vitro suggesting that they may affect transcriptional regulation (Baumann and Henschen, 1993b; Lane et al., 1993). All three polymorphisms have been reported to be in strong linkage disequilibrium in various populations (de Maat et al., 1995; Carter et al., 1997; Thomas et al., 1994; this study) making it difficult to correlate any single allele with an observed phenotype. It is possible that there is an additive effect and phenotype is determined more by haplotype than by individual alleles; however, as both our data and that of de Maat et al., (1995) are consistent with complete linkage
disequilibrium in Native Americans, this may not be a source of phenotypic variation in these populations. It is also possible that none of the alleles at these loci are responsible for the phenotype. There may be an as yet uncharacterized variant elsewhere in the gene that is affecting fibrinogen levels. Linkage disequilibrium between the alleles assayed in this study and this putative site would account for the reported association between fibrinogen levels and genotype at these three loci.

Given the design of these analyses, the over-representation of alleles in the Quechua may not represent selection for these alleles at altitude but rather selection against these alleles at low altitude. This issue was addressed by genotyping chimpanzee DNA at these loci on the assumption that the chimp genotype would represent the ancestral human genotype. The results of this were ambiguous. The alleles common in the Quechua (G\textsuperscript{455}, C\textsuperscript{148} and G\textsuperscript{448}) matched the chimp at base -455 but not at bases -148 or +448 and the three-loci haplotypes that were common in all three human populations were not represented in the chimp.

Despite the over-representation of the alleles associated with lower fibrinogen in the Quechua there was no evidence of lower plasma viscosity in that population compared with Caucasians (figure 3.4). However, there are some caveats to be considered before drawing any firm conclusions from these data. As it was not possible to separate the plasma immediately after extraction, sample quality was not optimal. Any subsequent remixing of the blood after extraction (inevitable under the conditions in which our samples were obtained) should be avoided as it will contribute to artificially increased viscosity (Harkness, 1971). As the Caucasian blood samples were taken in the field (thereby suffering the same abuse as the Quechua samples) comparisons between the Caucasian and the Quechua plasma should be independent of these potential sources of variation. Unfortunately, the Caucasian sample size is quite small (n = 3).

The data presented above are consistent with selection acting to favour the alleles associated with reduced fibrinogen in the Quechua. However, selection acts on phenotypes and not on genotypes. To determine if reduced plasma viscosity is an adaptive characteristic of this population, extensive analysis of the rheological properties of Quechua blood and the influence...
of fibrinogen genotype on these properties would be required. Such a study could provide insights into a potential adaptive mechanism that, despite having only an indirect effect on oxygen utilization, may confer a selective advantage in high altitude populations.
Chapter 4

Angiotensin converting enzyme (ACE) and angiotensin 2 receptor type 1 allele frequencies in the Quechua

Introduction

In 1997, Montgomery and co-workers reported a striking association between alleles in the gene encoding angiotensin converting enzyme (ACE) and high altitude performance in British mountaineers (Montgomery et al., 1997a). Allele distributions were compared between a control group of healthy, unrelated British males and 33 unrelated British climbers who had ascended beyond 7,000 m. An intronic insertion allele (ACE-I) in the ACE gene was significantly over-represented in the climbers and, in an elite subset of the climbers who had attained more than 8,000 m, there were no homozygotes for the deletion allele (ACE-D). The authors suggested that the ACE-I allele might contribute to performance at high altitude. In addition they reported that the training response for a simple repetitive muscle motion (elbow flexion) was significantly better in individuals carrying at least one copy of the insertion allele. The authors maintain that, as the time course for this improvement (10 weeks) is too short for muscle growth to account for the improvement, the increased performance is due to improved endurance characteristics. Other studies have shown an association between ACE I/D genotype and physical performance. The ACE-I allele was reported to be over-represented in Australian Olympic level rowers (Gayagay et al., 1998) and Hagberg et al., (1998) reported that among a cohort of physically matched postmenopausal women, those with an I/I genotype had a significantly higher $\dot{V}O_{2\max}$ (+ 6.3 ml kg\(^{-1}\) min\(^{-1}\)) than D/D homozygotes. Heterozygotes had an intermediate value (+ 3.3 ml kg\(^{-1}\) min\(^{-1}\)). This was attributed to an increased arteriovenous $O_2$ differential, 27% of which could be accounted for by ACE (I/D) genotype. Exercise-induced left ventricular hypertrophy is also influenced by ACE genotype with the ACE-I allele associated with reduced coronary enlargement (Montgomery et al., 1997b).

---


48
ACE is a component of the renin-angiotensin system (figure 4.1). It is a primarily a membrane bound metalloprotease, although some circulating activity can be detected. ACE removes two peptides from the decapeptide angiotensin I (a weak vasoconstrictor) generating angiotensin 2 (AT-2), an eight amino acid polypeptide with potent vasoconstricting properties. AT-2 has a life span of only a couple of minutes before it is degraded by angiotensinase thus the system allows rapid, reversible changes in vascular tone. The importance of the role of ACE and the angiotensin system in the regulation of cardiovascular function is reflected in the efficacy of various ACE inhibitors and AT-2 receptor blockers in treating heart failure and post-myocardial infarct (Butler et al., 1997). One such inhibitor (captopril) has also been shown to be efficacious in the treatment of high altitude pulmonary hypertension (Niazova et al., 1996). This raises the intriguing possibility that the overrepresentation of the ACE-I allele in climbers may not be due to performance enhancement per se, but rather due to a reduced susceptibility to high altitude pulmonary edema.

Figure 4.1. ACE mediated vasoconstriction. ACE inactivates the vasodilator bradykinin while activating the potent vasoconstrictor angiotensin 2. Adapted from Opie, 1999.
The insertion/deletion polymorphism in the ACE gene was first reported by Rigat et al. in 1990. In that study, D/D homozygotes had a 65% increase in ACE activity compared to I/I homozygotes, while I/D heterozygotes had an intermediate phenotype (+31%). The polymorphism was subsequently characterized as a 287 base Alu repeat insertion in intron 16 (see GenBank, Accession number X62855) and various studies have shown 14-28% of the variance in enzymatic activity to be associated with the genotype at this loci (Tiret et al., 1992; reviewed in Schunkert, 1997). This phenotype may be dependent on genetic background. Bloem et al. (1996) reported a correlation between ACE I/D phenotype and enzyme activity in Caucasian, but not Afro-Caribbean, children.

The ACE deletion allele has been proposed to be a risk factor for cardiovascular disease; however, the evidence for such a role is controversial. Cambien et al. (1992) and Evens et al. (1994) reported an over-representation of the D allele in patients with myocardial infarct. This observation was corroborated by studies in Europe, America and Japan and the phenotype expanded to include ischemic heart disease (IHD) and coronary artery disease (CAD) (summarized in Butler et al., 1997). However, studies in a number of other populations, including Caucasians (Agerholm-Larsen et al., 1997), Pima Indians (Nagi et al., 1998), Chinese (Chuang et al., 1997) (Saha et al., 1996), Japanese (Ishigami et al., 1995) and Southeast Asians (Saha et al., 1996), have failed to detect this association. It would appear that the cardiovascular phenotype of the ACE I/D is affected by genetic background and thus be dependent, to some extent, on race and gender (e.g. Sagnella et al., 1999).

AT-2 binds to two subtypes of cell surface receptors, the type 1 receptor (AT2-R1), which predominates in the vascular smooth muscle and cardiomyocytes; and the type 2 receptors found in the brain, uterus and adrenal medulla. Of these AT2-R1, a G protein-coupled, seven trans-membrane domain cell surface receptor, is the most physiologically active (Opie, 1999). Using single stranded conformational polymorphism analysis, Bonnardeaux et al., (1994) detected three polymorphisms in, or near, the single exon encoding this receptor. One of these, an A to C transversion at base 1166 in the 3' flanking region (A/C\textsuperscript{1166}), was associated with blood pressure with the C allele over-represented in hypertensive Caucasian subjects. As the
mutation itself does not appear to be functional, the mechanism by which it manifests the putative phenotype is unknown. The authors postulate that it may be in disequilibrium with a yet uncharacterized variant that is responsible for the phenotype. As is the case for ACE I/D polymorphism, the results of studies investigating this association are often inconsistent. Wang et al., (1997) reported a over-representation of the C allele in Caucasian hypertensives (0.40 vs. 0.29 in normotensives). Szombathy et al., 1998 reported no significant frequency difference between normotensives and hypertensives but did find an association between the C allele and higher systolic blood pressure in older or overweight hypertensives. Conversely, Castellano et al., (1996) reported lower blood pressures in C/C homozygotes. These inconsistencies could be accounted for by linkage disequilibrium between the A/C\textsuperscript{1166} polymorphism and a functional mutation. If the strength of the disequilibrium varied between the populations under examination, the association between the marker (the A/C genotype) and the phenotype would vary. All of the studies described above were in Caucasians, but this does not guarantee a common genetic background. As yet, no functional mutation in the AT2-R1 gene has been reported, and while a search around the AT2-R1 locus identified seven polymorphic sites, none of the alleles were in disequilibrium with A\textsuperscript{1166} or C\textsuperscript{1166} (Poirer et al., 1998). A second possibility is an interaction between the AT2-R1 mutation (or the putative linked functional polymorphism) and a variant at another gene. There is evidence for such an interaction between the A/C\textsuperscript{1166} and the ACE I/D polymorphisms. Having the C\textsuperscript{1166} allele enhances the effect of having the D allele at the ACE I/D locus on the development of myocardial infarct (Tiret et al., 1994).

The mechanism by which the ACE-I allele enhances physical performance is unclear, but as it appears to facilitate oxygen exchange (Hagberg et al., 1998), it may be particularly beneficial in high altitude populations. If that were the case, and assuming that the allele was present in the population that initially migrated to the Andes, selection may have resulted in an over-representation of the insertion allele in the Quechua. As the phenotype of this mutation may be influenced by the A/C\textsuperscript{1166} mutation in the AT2-R1 gene, there may be co-selection influencing allele frequencies at that locus as well. To address these issues, allele frequencies for
the ACE I/D and AT2-R1 A/C\textsuperscript{1166} polymorphisms were determined in a group of unrelated Quechua living at over 3500 m on the Peruvian \textit{altiplano} and compared with low altitude populations.

**Materials and Methods**

**Genotyping**

1) ACE polymorphism

DNA was prepared as described earlier and assayed using a three primer PCR based assay (Evans \textit{et al.}, 1994; see figure 4.2). Primer sequences are given in appendix ii. In the absence of the \textit{Alu} insertion, the primer pair ACE-1 : ACE-3 amplifies an 84 base product. In the presence of the insertion, the primer pair ACE-2 : ACE-3 amplifies a 65 base product. Both products are amplified in heterozygotes. Competition for primers limits amplification of the large (372 bp) ACE-1 : ACE-3 product. 100 ng DNA was PCR amplified in a reaction containing 25 \textmu M of ACE-1 and ACE-3, 7.5 \textmu M ACE-2, 0.2 mM dNTPs, 1.5 mM MgCl\textsubscript{2}, 20 mM Tris/Cl pH 8.4 and 50 mM KCl. Amplification conditions were 94°C, 1 min.; 55°C, 1 min.; 72°C, 2 min. for 40 cycles in a Perkin Elmer DNA Thermal Cycler. Products were electrophoresed on 8% polyacrylamide gels, stained with EtBr and recorded using a Polaroid photo-documentation system. Sample PCR products are shown in figure 4.3a.

2) AT2-R1 polymorphism

The A/C\textsuperscript{1166} polymorphism in AT2-RI was detected by \textit{Dde I} digestion of PCR amplified product. DNA was amplified with the primers ATR-3:ATR-4 (see appendix ii). DNA (50-200 ng) was amplified in a Perkin Elmer DNA Thermal Cycler using 0.033 nmoles of each primer and 0.625 units \textit{Taq} polymerase (GibcoBRL, Gaithersburg, MD). The final reaction mixture (25 \textmu l) was 0.2 mM dNTPs, 1.0 mM MgCl\textsubscript{2}, 20 mM Tris/Cl pH 8.4 and 50 mM KCl. Amplification conditions were 94°C, 1 min.; 55°C, 1 min.; 72°C, 2 min. for 40 cycles. Amplified product (10 \textmu l) was then digested using 0.2 -1.0 units \textit{Dde I} under the conditions prescribed by
the manufacturer (BRL). Digests were analysed as discussed above and examples are shown in figure 4.3b.

![Diagram](image)

**Figure 4.2:** PCR based assay for the insertion/deletion polymorphism in the angiotensin converting enzyme (ACE) gene. The alu insertion is shown in gray. ACE-1, ACE-2 and ACE-3 are primers. Diagnostic amplification products are shown in black. The large product potentially amplified by ACE-1: ACE-3 (striped) in the presence of the insertion is rarely seen.

**Results**

**ACE polymorphism**

The ACE I/D genotype was determined for 63 Quechua, 50 Na-Dene and 34 Caucasians. Allele and genotype frequencies for each are presented in Table 4.1a. Genotype frequencies are in Hardy-Weinberg equilibrium in all three groups. The frequencies of the insertion allele are significantly higher in the both the Quechua and the Na-Dene compared with Caucasians ($P<0.001$); however, there is no significant difference between the two Amerindian groups. Our allele frequencies for Caucasians fall within the range of values previously reported (Table 4.2).

**AT2-R1 polymorphism**

The A/C $^{1166}$ genotype was determined for 60 Quechua and 31 Caucasians. Allele and genotype frequencies for each are presented in Table 2.1b. Genotype frequencies for both groups are in Hardy-Weinberg equilibrium and allele frequencies do not significantly differ between the two groups. Our allele frequency for Caucasians does not differ from literature values (e.g. Hingorani and Brown, 1995). No association was found between alleles at the two loci.

53
Allele frequencies for both the ACE and the AT2-R1 polymorphisms are available through The Allele Frequency Database (Alfred), which is maintained by Dr. Ken Kidd’s lab at Yale University (New Haven, CT). The URL for this site is http://alfred.med.yale.edu/alfred/index.asp.

Table 4.2: ACE and AT2-R1 genotype and allele frequencies. a) Results for the 287 base insertion (I)/deletion (D) polymorphism in intron 16 of the angiotensin converting enzyme (ACE) gene for sample populations of Quechua, Na-Dene and Caucasians. b) Results for the A/C\textsuperscript{1166} polymorphism in the angiotensin 2 receptor I gene in Quechua and Caucasians.

<table>
<thead>
<tr>
<th>Population</th>
<th>ACE-I/D Genotype</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D/D</td>
<td>I/D</td>
<td>I/I</td>
<td></td>
</tr>
<tr>
<td>Quechua</td>
<td>2</td>
<td>24</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Na-Dene</td>
<td>5</td>
<td>16</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>9</td>
<td>21</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Allele frequencies of both native American groups differ significantly from Caucasians (at p<0.001).

<table>
<thead>
<tr>
<th>Population</th>
<th>A/C\textsuperscript{1166} genotype</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A/A</td>
<td>A/C</td>
<td>C/C</td>
<td></td>
</tr>
<tr>
<td>Quechua</td>
<td>34</td>
<td>20</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>14</td>
<td>16</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Allele frequencies do not significantly differ between the two samples.

**Discussion**

The mechanisms by which variants in ACE influence human performance are unknown however the association between ACE I/D genotype and \(\dot{V}O_2\text{max}\) (Hagberg et al., 1998) suggests that the ACE-I allele (or an allele to which it is linked) may facilitate oxygen uptake. \(\dot{V}O_2\text{max}\) is a measure of the maximal rate of oxygen consumption achievable by an organism and is a strong indicator of aerobic performance. It has been proposed that successful adaptation to hypoxic environments would be reflected in \(\dot{V}O_2\text{max}\) values close to sea level values (Baker, 1971).
Figure 4.3 a,b. Assay for polymorphisms in the ACE and AT2-R1 genes. Representative polyacrylamide gels demonstrating a) the insertion (I) and deletion (D) alleles of the angiotensin converting enzyme gene (ACE) and b) the A/C^{1166} Ddel RFLP in the angiotensin 2 receptor 1 gene. Genotypes are given across the top of each panel. Sizes of the bands are given in base pairs.
Consistent with this hypothesis, several studies have reported relatively high $\dot{V}O_{2\text{max}}$ values in both South American and Asian high altitude native populations (reviewed in Baker, 1976).

Table 4.2. Published allele frequencies for the ACE I/D polymorphism in various populations.

<table>
<thead>
<tr>
<th>Population</th>
<th>Frequency</th>
<th>Reference</th>
<th>Population</th>
<th>Frequency</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afro-Caribbean</td>
<td>I: 0.38</td>
<td>(Barley et al., 1996)</td>
<td>Inuit</td>
<td>I: 0.60</td>
<td>(de Maat et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>D: 0.62</td>
<td></td>
<td></td>
<td>D: 0.40</td>
<td></td>
</tr>
<tr>
<td>African (descent)</td>
<td>I: 0.44</td>
<td>(Sagnella et al., 1999)</td>
<td>South Asian</td>
<td>I: 0.61</td>
<td>(Sagnella et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>D: 0.56</td>
<td></td>
<td></td>
<td>D: 0.39</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>I: 0.43</td>
<td>(Sagnella et al., 1999)</td>
<td>Nepalese</td>
<td>I: 0.66</td>
<td>(Umemura et al., 1998)</td>
</tr>
<tr>
<td></td>
<td>D: 0.57</td>
<td></td>
<td></td>
<td>D: 0.34</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>I: 0.44</td>
<td>(Beige et al., 1998)</td>
<td>Chinese</td>
<td>I: 0.70</td>
<td>(Lee, 1994)</td>
</tr>
<tr>
<td></td>
<td>D: 0.56</td>
<td></td>
<td></td>
<td>D: 0.30</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>I: 0.48</td>
<td>(Barley et al., 1996)</td>
<td>Pima</td>
<td>I: 0.71</td>
<td>(Foy et al., 1996)</td>
</tr>
<tr>
<td></td>
<td>D: 0.52</td>
<td></td>
<td></td>
<td>D: 0.29</td>
<td></td>
</tr>
<tr>
<td>Asian-Indian</td>
<td>I: 0.55</td>
<td>(Saha et al., 1996)</td>
<td>Yanomama</td>
<td>I: 0.85</td>
<td>(Barley et al., 1994)</td>
</tr>
<tr>
<td></td>
<td>D: 0.45</td>
<td></td>
<td></td>
<td>D: 0.15</td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>I: 0.59</td>
<td>(Ishigami et al., 1995)</td>
<td>Australian</td>
<td>I: 0.97</td>
<td>(Lester et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>D: 0.41</td>
<td></td>
<td>Aborigines</td>
<td>D: 0.03</td>
<td></td>
</tr>
</tbody>
</table>

Immediately following travel to high altitude, lowlanders experience a reduction in $\dot{V}O_{2\text{max}}$ at a rate of about 10% per 1000 m ascended beyond 1500 m (Dua and Sen Gupta, 1980), and although there may be some improvement with time, $\dot{V}O_{2\text{max}}$ remains depressed even after prolonged residence. For example, an initial drop in $\dot{V}O_{2\text{max}}$ of 26% in Indian soldiers transferred to 4100 m improved to about a 21% reduction after two years (Dua and Sen Gupta, 1980). Altitude-induced $\dot{V}O_{2\text{max}}$ depression is not as pronounced in high altitude natives, and values in these populations have been reported to equal those found in sea level populations (see Baker, 1976). There may be a genetic component to this improved response to hypoxic exposure. While $\dot{V}O_{2\text{max}}$ is influenced by gender, age, BMI and level of activity, there is clearly also a heritable genetic component. In family aggregation studies, the heritability (the percent of variance that is due to genetic background) of increased $\dot{V}O_{2\text{max}}$ in response to training was estimated at 0.47 (Bouchard et al., 1998). In twin studies heritability of $\dot{V}O_{2\text{max}}$ ranged from 56
0.25 - 0.79 depending on method of measurement and analysis (reviewed in Klissouras, 1997). The diminishment of $\dot{V}O_{2\max}$ following translocation from sea-level to 4000 m was less in Quechua born or raised at sea level than in Caucasian lowlanders (Buskirk, 1976), and the author suggests that this may represent the influence of some genetic factor on the impact of hypoxia on the work capacity of people of Quechua heritage regardless of their altitude of residence.

Given the association between the ACE-I allele and $\dot{V}O_{2\max}$ (Hagberg et al., 1998), and the observations that high altitude populations have relatively high $\dot{V}O_{2\max}$ values (Baker, 1976), it was hypothesized that the ACE-I allele should be over-represented in the Quechua compared to sea level populations. This was tested by comparing the frequency of this allele in Quechua, Na-Dene and Caucasians; however, our data do not support the hypothesis. While the frequency of the ACE-I allele in the Quechua (0.78) was higher than in our Caucasian sample (0.43), it did not differ significantly from that in our Na-Dene sample (0.74), a lowland native American population and furthermore, was less common in the Quechua than in the Yanomama (0.85; Barley et al., 1994), a lowland South American population with whom the Quechua presumably share a relatively recent common ancestry.

There are a number of reasons why the over-representation of the ACE insertion allele seen in the British climbers might not be seen in the Quechua. There is considerable difference between the conditions found at 7000 m and those at 3000 m; and the benefits conferred by the insertion allele may be negligible at the lower altitude. It is also possible that the allele enhances individual, but not population fitness. Given the association between the ACE-D allele and heart disease, the fact that the ACE-I allele is not more highly represented globally suggests that it may have some negative effects (or conversely the D allele may have unknown beneficial effects). Recently it was reported that the ACE-I allele might be a risk factor for Alzheimer disease (Kehoe et al., 1999). Genetic background of the populations could be a factor as well. The intronic ACE I/D polymorphism itself appears not to be responsible for the phenotype (Rosatto et al., 1999). This suggests that the I/D polymorphism is acting as a marker for another

---

1 ACE is alternatively known as dipeptidyl carboxypeptidase 1, encoded by the gene DCP 1
mutation elsewhere in, or proximal to, the gene and the extent of association between the I/D genotype and the phenotype would depend on the degree of linkage disequilibrium between the two loci. The extent of linkage disequilibrium between any two loci can vary between populations and, if linkage disequilibrium between the I/D marker and the performance phenotype was absent in Native Americans, selection for the phenotype would not be reflected in the frequency of the ACE-I/D alleles. Recently, (Rieder et al., 1999) scanned 24 kilobases of sequence around the ACE gene in 11 individuals (African and European American) and found 78 polymorphic sites. Of these sites, which segregated into 13 haplotypes of varying commonality, 17 were in absolute disequilibrium with the ACE I/D polymorphism in intron 16. These haplotypes can be grouped into three clades, two of which include the ACE Alu deletion allele. Plasma ACE levels are 25% higher in these two clades than in the clade marked by the deletion allele (Farrall et al., 1999).

There are few studies of ACE allele frequencies in native Americans. Both our data for the Na-Dene and that of Foy et al., (1996) in the Pima indicate that the frequency of the ACE insertion allele is higher in Native American populations than in Caucasians. As the frequency of the ACE-I allele also tends to be higher in Asian populations (Table 3.1) than in Caucasians, this may reflect the ancestral allele frequencies of the migrants who crossed from Asia to North America.

The interaction between ACE-I/D and AT2-R1 A/C\textsuperscript{1166} genotypes reported in studies on cardiopathology (Tiret et al., 1994) raises the possibility that the ACE I/D phenotype may be influenced by A/C\textsuperscript{1166} genotype. The frequency of the A/C\textsuperscript{1166} was determined in the Quechua and it was found that neither allele was uncommon. It is unlikely, therefore, that the apparent lack of selection for the ACE-I allele in the Quechua could be attributed to the absence of an allele at the AT2-R1 locus that is involved in generating the beneficial phenotype. As the two genes are on different chromosomes (ACE: 17q33; AR2-R1: 3q22), linkage is not at issue however segregation distortion is formally possible. Independent assortment was tested for and no association was found between genotypes at the two loci.
Conclusions

The results of this study do not support the hypothesis that there is significant selective pressure acting in favour of transgenerational transmission of the ACE insertion allele in the Quechua. The frequency of the ACE-I allele is higher in the Quechua than in Caucasians but this is true for other Native American populations and therefore may represent an ancestral trait brought with these populations during the initial colonization of the Americas. It is possible that the relatively high frequency of this allele in the antecedents of the Quechua may have facilitated colonization of the highlands, but there is no evidence for subsequent selection favouring this allele once they were separated from their lowland relatives.
Chapter 5

Beta_2-adrenergic receptor allele frequencies in the Quechua

Introduction

The beta-adrenergic receptors are a family of G-protein coupled cell surface catecholamine receptors (Johnson, 1998; see figure 5.1). They are found in virtually all tissues and are involved in the sympathetic mediation of diverse physiological responses (reviewed in Bilezikian, 1987). One member of the family, the beta_2-adrenergic receptor (β_2-AR), is particularly important in regulating lung function. In response to catecholamine stimulation these receptors affect dilatation of both the pulmonary and bronchial vasculature as well as relaxation of the bronchial smooth muscle. The β_2-AR is also involved in regulating smooth muscle tone in a number of tissues, cardiac function (where it is secondary to the β_1-AR) and fat metabolism (Arner and Hoffstedt, 1999).

A number of polymorphisms have been characterized in the human β_2-AR gene (Reihsaus et al., 1993) (figure 5.1). These include: an A to G transition at base 46 (A/G_46) which substitutes a glycine (gly) for an arginine (arg) at residue 16, a G to C transversion at base 79 (G/C_79) which substitutes a glutamine (gln) for a glutamic acid (glu) at residue 27 and a C to T transition at base 491 (C/T_491) which substitutes an isoleucine (ile) for a threonine (thr) at residue 164. In addition, there are several silent mutations, which do not alter protein sequence.

The missense mutations at bases 46 and 79 are both in the NH_2 terminal extracellular domain but are phenotypically distinct. Although neither appear to affect agonist binding or G_s coupling, both alter agonist promoted regulation of the receptor (see review by Liggett and Raymond, 1995). The gly16 variant increases down regulation of the receptor subsequent to agonist exposure, whereas the gln27 variant reduces down regulation. When combined, the gly16 phenotype appears to dominate the gln27 phenotype. In vivo effects of these mutations have primarily been studied with respect to their effects on airway responsiveness and in asthma.

---

(e.g. Hall et al., 1995; Turki et al., 1995; Martinez et al., 1997; D'amato et al., 1998); however, linkage disequilibrium between the alleles complicates analysis of the individual actions of the mutations. The rare T allele at base 491 lowers affinity for a variety of agonists (Green et al., 1993) and has been associated with the progression of congestive heart disease (Liggett et al., 1998).

High altitude natives are characterized by increased lung capacity, for which a genetic component has been demonstrated (Greksa, 1996), and by increased thoracic blood volume (Velásquez, 1976). In addition, increased tidal volumes and $\dot{V}O_2$ max have been reported in high altitude natives compared to acclimated lowlanders of similar fitness level (Sun et al., 1990). Presumably these adaptations facilitate blood oxygenation in an environment where available oxygen is reduced by as much as a third. Given the role that the $\beta_2$-AR plays in regulating the flow of air and blood in the lungs, it is possible that the variants of these receptors that increase airway responsiveness to agonists ($A^{46}$, $G^{79}$, $C^{491}$), and therefore may contribute to increased air and blood flow in the lungs, would be favoured in these populations. Receptor desensitization, and therefore reduced responsiveness to agonists, has been shown to be associated with the development of high altitude pulmonary hypertension in highlanders (Aldashev et al., 1989).

In addition to pulmonary function correlates, $\beta_2$-AR genotype has been associated with body composition. Sedentary males (but not physically active males) who were homozygous for C at base 79 were significantly heavier and had a greater body mass index (BMI) than either heterozygotes or G/G homozygotes (Meirhaeghe et al., 1999) suggesting reduced fat metabolism in C/C homozygotes. This effect appears to be gender influenced as Large et al., (1997) found that the $G^{79}$ allele was associated with increased body mass and fat deposition in females and Ishiyama-Shigemoto et al., (1999) showed a correlation between the $G^{46}$ and $G^{79}$ genotypes and obesity, with the former association unique to females. It may be advantageous at high altitude to minimize lipid metabolism as the ATP yield per $O_2$ consumed is lower than in carbohydrate utilization (Holden et al., 1995), and if the correlation between $\beta_2$-AR genotype and body fat was due to the receptor variant effecting fat metabolism, then some genotypes may be over-represented in high altitude populations.
Because of the roles of the $\beta_2$-AR in pulmonary function and fat metabolism, allele frequencies for this gene in the Quechua were determined so as to ascertain whether any of the receptor variants were over-represented in this population compared with lowland populations. While such a finding could be explained by stochastic changes in populations due to genetic
drift, it could also represent selection favouring the allele, due either to its phenotype or to linkage with a beneficial variant at another locus.

The frequencies of the three missense mutations discussed above and two common silent mutations (G/A\textsuperscript{252} and C/A\textsuperscript{523}) were compared in the Quechua, Na-Dene and Caucasian samples. It was hypothesized that if trans-generational adaptations to high altitude populations involved \(\beta_2\)-AR function, then the frequencies of alleles in the gene encoding the receptor might differ in these populations when compared to lowlanders. In addition, both Caucasian and Quechua DNAs were assayed for the potential \(Rsa\) I RFLP reported in (Emorine \textit{et al}., 1987), and sequenced the coding sequence of the gene in three Quechua to determine if there were any as yet unreported polymorphisms common in this population.

\textit{Methods and Materials}

\textit{Genotyping}

Five polymorphisms in the \(\beta_2\)-AR gene were assayed by restriction endonuclease digestion of PCR amplified DNA: G/A\textsuperscript{46}, C/G\textsuperscript{79}, G/A\textsuperscript{252}, C/T\textsuperscript{491} and C/A\textsuperscript{523}. PCR product size and diagnostic restriction endonucleases are shown in Table 5.1 and primer sequences are given in appendix ii. The G/A transition at base 46 did not alter any known restriction endonuclease recognition site. A primer (8\_2AR-46Nco) was designed with a C replacing an A at the third to last position such that amplification of DNA with the G allele at base 46 generated a Nco I recognition site. A potential polymorphism, CA/AC\textsuperscript{422,3} (Emorine \textit{et al}., 1987), which could be detected by \(Rsa\) I digestion of \(\beta_2\)AR-P1 : \(\beta_2\)AR-P2 product, was also assayed in a subset of the Caucasian sample.

DNA (100-200 ng) was amplified using 0.033 nmoles of each primer and 0.625 units \textit{Taq} polymerase (GibcoBRL, Gaithersburg, MD). The final reaction mixture (25 \(\mu\)l) was 0.2 mM dNTPs, 1.5 mM MgCl\(_2\), 20 mM Tris/Cl pH 8.4 and 50 mM KCl. Reactions were overlaid with 2 drops of light mineral oil (Fisher Scientific) and PCR was performed in a Perkin Elmer DNA Thermal Cycler. Amplification conditions were 94°C, 1 min; 58°C, 1 min; 72°C, 2 min for
40 cycles. Following amplification the oil overlay was extracted by the addition of 20 µl chloroform and 20 µl of H₂O followed by a 10 s centrifugation at 12,000 x g. Amplification was confirmed by running 8 µl of the reaction on a 2% agarose gel and EtBr staining. PCR product (10 µl) was then digested in a 20 µl reaction using 0.2 - 1.0 units of the appropriate restriction endonuclease and the conditions prescribed by the manufacturers. Digests were then electrophoresed on 8% polyacrylamide gels, stained with EtBr and recorded using a Polaroid photo-documentation system (see figure 5.2).

Table 5.1: Location and product size of β₂-AR PCR primers.

<table>
<thead>
<tr>
<th>Primer pair</th>
<th>Product size</th>
<th>Polymorphism</th>
</tr>
</thead>
<tbody>
<tr>
<td>β₂AR-46Nco (28) : β₂AR-P2a (143)</td>
<td>135 bp</td>
<td>G/A⁴⁶ (NcoI)</td>
</tr>
<tr>
<td>β₂AR-P1 (-59) : β₂AR-P1A (142)</td>
<td>221 bp</td>
<td>C/G⁷⁹ (BbvI)</td>
</tr>
<tr>
<td>β₂AR-P1 (-59) : β₂AR-P2 (355)</td>
<td>433 bp</td>
<td>C/T²⁵² (MvaI)</td>
</tr>
<tr>
<td>β₂AR-P3 (275) : β₂AR-P4 (695)</td>
<td>440 bp</td>
<td>C/T⁴⁹¹ (MnlI)</td>
</tr>
<tr>
<td>β₂AR-P5 (641) : β₂AR-P6 (1054)</td>
<td>432 bp</td>
<td>*</td>
</tr>
<tr>
<td>β₂AR-P7 (898) : β₂AR-P8 (1280)</td>
<td>402 bp</td>
<td>*</td>
</tr>
</tbody>
</table>

Location of the 5' end of the primer is given in parentheses in the first column and is relative to the translation start site. Primer sequences are given in appendix ii. Diagnostic restriction endonuclease used for RFLP analysis is given in parentheses in the last column. The cytosine shown in bold face in β₂AR-46Nco replaces an adenosine such that a NcoI site is generated in the presence of the G allele at the polymorphic site. * primers used for sequencing only.
Figure 5.2. Assay for RFLPs in the β₂-AR gene. Representative polyacrylamide gels showing separation of digestion products for the RFLPs used to identify alleles of the β₂-adrenergic receptor gene. Small digestion products not shown. *Mnl*I digestion also generates constant 50 and 49 bp products. Genotypes are given across the top of each panel. The polymorphism and the diagnostic restriction endonuclease are given across the bottom. Fragment sizes are in base pairs. The *Nco*I polymorphism is generated using a primer with a single mismatch base. A T/T homozygote at base 491 was never observed.
Allele frequencies were established by gene counting and compared by Chi-square ($\chi^2$) analysis using StatView SE software (Abacus Concepts Inc., Berkley, CA). To correct for the number of tests (20) significance was accepted at a $P$ value of $< 0.0025$ for 1 degree of freedom. Linkage disequilibrium analysis (likelihood-ratio test for phase unknown genotypic data) and tests for Hardy-Weinberg equilibrium (modified Markov-chain random walk algorithm) were performed using Arlequin software (Schneider et al., 1997).

**DNA sequencing**

The $\beta_2$-AR gene was sequenced in three Quechua. The entire coding region of the gene was amplified from total genomic DNA in four PCR reactions using the following primer pairs: $\beta_2$AR-P1:P2, $\beta_2$AR-P3:P4, $\beta_2$AR-P5:P6 and $\beta_2$AR-P7:P8 (see appendix ii). Amplified DNA was separated on 2% agarose gels, purified with QIAquick columns (Qiagen, Mississauga, ON) and sequenced using primers $\beta_2$AR-P1 through $\beta_2$AR-P8. Primer synthesis and dideoxy DNA sequencing were performed by NAPS at U.B.C.

**Results**

Genotype and allele frequencies for the Quechua, Na-Dene and Caucasian samples are given in figure 5.2. For the G/A transition at base 46, allele frequencies in the Na-Dene differed from both the Quechua and the Caucasians. Allele frequencies for C/G$^{79}$ differed among all three samples with the C allele being significantly more common in the Native Americans. The Quechua were monomorphic for this allele. Allele frequencies of both silent mutations (G/A$^{252}$, C/A$^{523}$) were different in the Quechua compared with both the Na-Dene and the Caucasians. The frequency of the relatively rare T allele at base 491 did not differ between the three samples. There was no significant difference in allele frequencies between the group that donated blood and the group that donated buccal cells or between groups from the different villages.

In both Native American samples there was strong linkage disequilibrium (table 5.3) between A/G$^{46}$, G/A$^{252}$ and C/A$^{523}$ ($P < 0.01$) consistent with the haplotypes A$^{46}$ G$^{252}$ C$^{523}$/ G$^{46}$ A$^{252}$ A$^{523}$. In the Caucasians there was significant linkage disequilibrium between A/G$^{46}$ and G/C$^{79}$ ($P < 0.01$) with the following haplotype frequencies in informative genotypes: 0.28 A$^{46}$C$^{79}$,
0.00 A\textsuperscript{46}G\textsuperscript{79}, 0.25 G\textsuperscript{46}C\textsuperscript{79} and 0.47 G\textsuperscript{46}G\textsuperscript{79}. There was also strong linkage disequilibrium between G/A\textsuperscript{252} and C/A\textsuperscript{523} (P < 0.01) consistent with the haplotypes G\textsuperscript{252}C\textsuperscript{523}/A\textsuperscript{252}A\textsuperscript{523}. All three samples were either monomorphic or in Hardy-Weinberg equilibrium at all five loci. Allele frequencies in the Caucasian sample did not significantly differ from previously published Caucasian values (Reihsaus \textit{et al.}, 1993).

Allele frequencies for the β\textsubscript{2}-AR polymorphisms are available through The Allele Frequency Database (Alfred), which is maintained by Dr. Ken Kidd’s lab at Yale University (New Haven, CT). The URL for this site is http://alfred.med.yale.edu/alfred/index.asp.

In addition to the aforementioned polymorphisms, a subset of twenty Caucasian DNAs was analyzed for a potential Rs\textsuperscript{a} I polymorphism at bases 422,423. All samples were AC/AC suggesting that the CA sequence reported in (Emorine \textit{et al.}, 1987) was either a sequencing artifact or a rare variant.

The coding region of the gene encoding the β\textsubscript{2}-AR was sequenced in three Quechua. No new polymorphisms were detected and sequence data were consistent with RFLP analysis. Sequence data have been submitted to GenBank (accession numbers AF169225, AF202305 and AF203386).

\textit{Discussion}

Allele frequencies of four of the five β\textsubscript{2}-adrenergic receptor polymorphisms assayed in the Quechua, the Na-Dene and the Caucasian samples varied between the three samples, whereas the T\textsuperscript{491} allele was uniformly rare. The G/A\textsuperscript{46} missense mutation and the two silent mutations G/A\textsuperscript{252} and C/A\textsuperscript{523} were in strong linkage disequilibrium in both native American populations suggesting that the G\textsuperscript{46}G\textsuperscript{252}C\textsuperscript{523}/A\textsuperscript{46}A\textsuperscript{252}A\textsuperscript{523} haplotype predominated in the Asiatic antecedents of native Americans; however haplotype frequencies differed significantly between the Quechua and the Na-Dene. The Quechua and the Na-Dene are thought to have descended from different waves of migration into the New World (Cavalli-Sforza \textit{et al.}, 1994) and have been diverging for at least 14,000 years. The difference in haplotype frequencies between the two New World groups may have resulted from genetic drift occurring subsequent
Table 5. β₂-adrenergic receptor genotype and allele frequencies for the Quechua, Na-Dene and Caucasian samples

<table>
<thead>
<tr>
<th>G/A&lt;sup&gt;46&lt;/sup&gt; genotype</th>
<th>Population</th>
<th>G/G</th>
<th>G/A</th>
<th>A/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quechua</td>
<td>24</td>
<td>28</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Na-Dene</td>
<td>4</td>
<td>14</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>11</td>
<td>16</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>G/C&lt;sup&gt;79&lt;/sup&gt; genotype</th>
<th>Population</th>
<th>G/G</th>
<th>G/C</th>
<th>C/C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quechua</td>
<td>0</td>
<td>0</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>Na-Dene</td>
<td>0</td>
<td>9</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>7</td>
<td>15</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>G/A&lt;sup&gt;252&lt;/sup&gt; genotype</th>
<th>Population</th>
<th>G/G</th>
<th>G/A</th>
<th>A/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quechua</td>
<td>11</td>
<td>28</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Na-Dene</td>
<td>35</td>
<td>12</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>23</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C/T&lt;sup&gt;491&lt;/sup&gt; genotype</th>
<th>Population</th>
<th>C/C</th>
<th>C/T</th>
<th>T/T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quechua</td>
<td>63</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Na-Dene</td>
<td>49</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Caucasian</td>
<td>29</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C/A&lt;sup&gt;523&lt;/sup&gt; genotype</th>
<th>Population</th>
<th>C/C</th>
<th>C/A</th>
<th>A/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quechua</td>
<td>11</td>
<td>28</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Na-Dene</td>
<td>35</td>
<td>13</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>23</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

1 indicates significant difference from the Na-Dene; 2 indicates significant difference from the Caucasians (P<0.0025). Caucasian frequencies did not differ from those previously published (see text for references). G/A<sup>252</sup> and C/A<sup>523</sup> are silent polymorphisms and thus do not alter protein structure.
Table 5.3: Linkage disequilibrium between the β-AR allele. Allele association in the Quechua, Na-Dene and Caucasians.

### Quechua

<table>
<thead>
<tr>
<th>Polymorphic locus</th>
<th>G/A 46</th>
<th>C/G 79</th>
<th>G/A 252</th>
<th>C/T 491</th>
<th>C/A 523</th>
</tr>
</thead>
<tbody>
<tr>
<td>G/A 46</td>
<td>*</td>
<td>nt</td>
<td>+</td>
<td>nt</td>
<td>+</td>
</tr>
<tr>
<td>C/G 79</td>
<td></td>
<td>nt</td>
<td></td>
<td>nt</td>
<td></td>
</tr>
<tr>
<td>G/A 252</td>
<td>*</td>
<td>nt</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>C/T 491</td>
<td>*</td>
<td></td>
<td></td>
<td>nt</td>
<td></td>
</tr>
<tr>
<td>C/A 523</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

### Na-Dene

<table>
<thead>
<tr>
<th>Polymorphic locus</th>
<th>G/A 46</th>
<th>C/G 79</th>
<th>G/A 252</th>
<th>C/T 491</th>
<th>C/A 523</th>
</tr>
</thead>
<tbody>
<tr>
<td>G/A 46</td>
<td>*</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>C/G 79</td>
<td></td>
<td>*</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>G/A 252</td>
<td>*</td>
<td></td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>C/T 491</td>
<td></td>
<td>*</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>C/A 523</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

### Caucasian

<table>
<thead>
<tr>
<th>Polymorphic locus</th>
<th>G/A 46</th>
<th>C/G 79</th>
<th>G/A 252</th>
<th>C/T 491</th>
<th>C/A 523</th>
</tr>
</thead>
<tbody>
<tr>
<td>G/A 46</td>
<td>*</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C/G 79</td>
<td></td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>G/A 252</td>
<td>*</td>
<td></td>
<td>-</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>C/T 491</td>
<td></td>
<td>*</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>C/A 523</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

All pairs denoted by + are in significant linkage disequilibrium \( P<0.01 \). The polymorphisms at bases 79 and 491 are monomorphic in the Quechua and were not tested (nt).
to the ancestral separation or possibly from selection favouring one haplotype. Two of the
linked mutations (G/A\textsuperscript{252} and C/A\textsuperscript{523}) are silent and presumably have no functional phenotype on
which selection could act; however the A/G\textsuperscript{46} polymorphism is reported to be associated with
several phenotypes. As the G\textsuperscript{46} allele is the less common allele in the Na-Dene and other Asian
populations (Weir \textit{et al.}, 1998; Ishiyama-Shigemoto \textit{et al.}, 1999), the relatively high frequency of
the G\textsuperscript{46} allele in the Quechua may have arisen after their ancestors separated from those of the
Na-Dene.

The Quechua were monomorphic for the C\textsuperscript{79} allele; however, the difference in frequency
between the Quechua and the Na-Dene, while significant, is not large and could be accounted for
by genetic drift. Furthermore, the frequency of C\textsuperscript{79} is high in Chinese (0.80; Weir \textit{et al.}, 1998)
and in Japanese (0.93; Ishiyama-Shigemoto \textit{et al.}, 1999), suggesting that a relatively high
frequency of the C\textsuperscript{79} allele may be characteristic of people of Asian origin. If the increased
frequency of the C\textsuperscript{79} allele that was observed in the Quechua and the Na-Dene was a
characteristic of Amerindian populations, it may account for the reduced venodilatory response to
isoproterenol in Native Americans compared to Caucasians (Vajo \textit{et al.}, 1998). The complete
absence of the G\textsuperscript{79} allele in our Quechua sample was somewhat surprising. Assuming an
average racial admixture of 0.247 between Quechua and Caucasians (Salzano and Callegari-
Jacques, 1988), and an G\textsuperscript{79} frequency of 0.27 in Hispanic Caucasians (Martinez \textit{et al.}, 1997),
one would expect eight G\textsuperscript{79} alleles in the Quechua sample due to admixture alone. However,
given our relatively small sample size, we lacked the statistical power to determine whether this
was significant. The absence of the G\textsuperscript{79} allele may reflect selection favouring the C\textsuperscript{79} allele in the
Quechua; however more data are needed to address this issue, especially pertaining to admixture
in the communities in which our Quechua volunteers lived and the frequency of this allele in
other South American populations.

Numerous studies have shown that catecholamine levels increase with altitude (see Ward
\textit{et al.}, 1995). Wolfel \textit{et al.}, (1994) reported a significant increase in norepinephrine, and a
concomitant increase in systemic hypertension, at 4300 m. Chronic agonist exposure can reduce
the receptor response by both reducing receptor density and by desensitizing the receptor itself.
Aldashev et al., (1989) found that this desensitization was associated with pulmonary hypertension in Kyrgyz highlanders living at between 3600 - 4200 m in Eastern Pamir (Central Asia). Both the A\textsuperscript{46} and C\textsuperscript{79} alleles are associated with increased receptor down-regulation and thereby increased receptor sensitivity to catecholamines (Green et al., 1994). It was anticipated that selection might favour this phenotype at high altitude as it could facilitate air and blood flow in the lungs. Our results were opposite to our expectations and showed reduced frequencies of these alleles in the Quechua.

Catecholamines effect diverse systems and increased adrenergic response, while expected to be advantageous at altitude in pulmonary tissue may be detrimental elsewhere. The decrease in \(\beta\)-adrenergic receptor density that has been reported in high altitude populations has been postulated to reduce myocardial oxygen consumption by limiting heart rate (Antezana et al., 1992). Even if cardiac adrenergic response is primarily mediated by \(\beta_1\) receptors, some role for \(\beta_2\)-AR has been reported (Bilezikian, 1987).

Several studies have reported associations between G/A\textsuperscript{46} and G/C\textsuperscript{79} genotype and body fat. If the Quechua preferentially metabolize carbohydrates as a metabolic strategy to improve the amount of ATP obtainable per mole of oxygen, alleles associated with increased fat deposition may be beneficial. The absence of the G\textsuperscript{79} encoded gly27 supports this hypothesis given the data of Meirhaeghe et al., (1999) which show this allele to be associated with increased body fat in Caucasian males. However others report that it is the alternate allele (C\textsuperscript{79}) that is associated with increased fat deposition (Large et al., 1997; Ishiyama-Shigemoto et al., 1999). This apparent discrepancy may be due to different genetic backgrounds, including gender, as the Large et al., (1997) study focuses on in females and the (Ishiyama-Shigemoto et al., 1999) study is with Asians. The idea that the absence of the G\textsuperscript{79} allele may reflect fuel use preference in high altitude natives is intriguing and bears further investigation.

An alternate explanation for the variation in haplotype frequencies between the Quechua and the Na-Dene is selection having acted on an uncharacterized variant linked to the loci that was examined. The coding region of the \(\beta_2\)-AR gene in three Quechua was sequenced and no new mutations were detected. The probability of missing a polymorphism present at the
frequency of the less common haplotype (40%) in six chromosomes is less than 5% however, it is not necessary for such a mutation to be in the coding sequence or even in the same gene.

There is evidence that some of the phenotypic effects of $\beta_2$-AR polymorphisms are haplotypic rather than allelic. In a study of bronchial hyper-responsiveness in normal subjects (D'amato et al., 1998) showed an association with the G$^{46}$C$^{79}$ (gly16, gln27) haplotype but not with individual alleles. Our data suggest that these alleles are in linkage disequilibrium in Caucasians but not in the Na-Dene. The Quechua were not tested as the sample was monomorphic for the C$^{79}$ allele. As the strong linkage disequilibrium between two loci can confound phenotypic analysis of single alleles, the Quechua could prove to be a valuable population in which to study the clinical effects of the A/G$^{46}$ polymorphism in airway disorders.

There are significant differences in both allele frequencies and in haplotype structure in the populations that we examined. It is possible that selection has acted to favour variants in the $\beta_2$-AR in the Quechua as an adaptive response to hypoxic environment in which their ancestors chose to live. This issue however cannot be addressed without comparing the Quechua to closely related, low altitude control populations. As the ancestors of the Na-Dene and Quechua may have arrived in North America as part of different migration waves (Cavalli-Sforza et al. 1994), the two groups may be genetically distant. However, recent HLA analysis (Monsalve et al. 1999) supports a single wave of migration suggesting that these two populations may be more similar than that previously suspected. The divergence between two populations as temporally and geographically separate as European Caucasians and the Quechua would be even more pronounced than that between the native American groups. While comparisons between these populations may be of limited value in determining the basis for the observed differences in allele frequency and haplotype structure, the data do demonstrate that some alleles in the $\beta_2$-adrenergic gene are greatly over-represented in the high altitude natives and raises the possibility that these alleles, or alleles to which they are linked, may play some role in the adaptive response to long term hypoxia.
Introduction

Exposure to hypoxia triggers a rapid hematological response characterized by an increase in hematocrit that compensates for the reduction in oxygen availability by increasing the oxygen carrying capacity of the blood. Elevated hematocrits become apparent within a week of exposure to high altitude and values of 60% are considered normal for sea level sojourners (Ward et al., 1995). Although these values tend to decline somewhat from an initial peak, some degree of secondary polycythemia can be maintained for years in long term migrants (Heath and Williams, 1995).

Although there are some reports of very high values in high altitude natives (e.g. 71.1%; Hurtado, 1932), not all studies agree that secondary polycythemia is a universal characteristic of these populations. Andean values tend to be somewhat higher than those reported for Himalayans living at the same altitudes, and some authors (e.g. Beall et al., 1987) postulate that this may be due to greater hypoxaemic stimulus of erythropoiesis, possibly reflecting less effective adaptation in the Andeans. In 1983, Garrutto and Dutt, recognizing that diverse socio-environmental factors can contribute to polycythemia, examined a number of hematological parameters in traditional, agro-pastoral Quechua communities located at 4200 m near Puno, Peru. They found that the mean hematocrit in males 6-21 years old was only 10% higher than age matched sea level values, that reticulocyte counts (a measure of erythropoietic activity) were not significantly increased and that the average hematocrit in adult males was 51.4%, well within the normal range for men at sea level (e.g. 40%-54% Thomas, 1993). The authors concluded that excessive secondary polycythemia is not necessarily a characteristic of Andean native populations and suggested that the high values reported in some studies were due to environmental and demographic factors introduced by sampling in mining communities (Garrutto and Dutt, 1983; see also Ballew et al., 1989).

Chapter 6

Sequence of the erythropoietin and the HIF-1α genes in the Quechua
Studies of Quechua or Aymara residing at lower altitudes are uncommon; however, some hematological data has been published. Arnaud et al., 1985 measured hematocrits of both Quechua and Aymara males living in Santa Cruz, Bolivia at 450 m. Average hematocrits for both groups (42.1%, 41.6%) were either near, or below, the low end of normal sea level range (40% in Thomas, 1993; 42% in Spraycar, 1995). Another study of Aymara living at around 500 m in coastal Chile also reported a relatively low average hematocrit (39.6%; Clench et al., 1982). Although this study combined male and female data, the average hematocrit was still in the low end of the normal female range (37% - 47%; Spraycar, 1995). On the whole, the data suggest that despite chronic exposure to hypoxic conditions, Andean highlanders tend not to have excessive compensatory polycythemia when resident at altitude and have relatively low hematocrits when living at lower altitudes.

The early steps of erythropoiesis, the process by which hematopoietic stem cells differentiate into mature RBCs, is mediated by the glycoprotein hormone erythropoietin (Epo). Epo, which is synthesized in the adult kidney, interacts with a cell surface receptor on mature erythroid progenitor cells (CFU-E cells) and stimulates their maturation into pro-erythroblasts that subsequently divide into 16 RBCs (figure 6.1). In the absence of Epo, the CFU-E cells eventually die without differentiating further (Erslev and Caro, 1988).

Synthesis of Epo is stimulated by reduced PaO$_2$ (Klausen, 1998). This can result from a number of factors including: impaired hemoglobin function, reduced hemoglobin concentration and, as in the case of hypoxic stimulation of erythropoiesis, low SaO$_2$ secondary to reduced oxygen availability. Within 1 - 2 days following exposure to 3500 -4500 m, there is a substantial increase in immunoreactive Epo levels (Milledge and Cotes, 1985). In subjects who remained at these altitudes, Epo levels declined to sea level values after approximately three weeks, whereas they continued to rise in subjects who went on to higher altitudes. Epo levels remain elevated until PaO$_2$ returns to normal, at which point they begin to decline (Erslev and Caro, 1988). Failure of this negative feedback control is suspected to play a role in the development of chronic mountain sickness, or Monge’s disease, a condition that is characterized by extreme secondary polycythemia.
Epo production is regulated at the RNA level by hypoxia inducible factor-1 (HIF-1), an oxygen sensitive transcription complex that binds to the 3’ enhancer region of the Epo gene (figure 6.2) and triggers expression of the gene (see review by Bunn et al., 1998). HIF-1 is comprised of an alpha (HIF-1α) and a beta subunit (HIF-1-β, also known as arylhydrocarbon receptor nuclear translocator (ARNT)) with the alpha subunit responsible for the oxygen mediated activity of the factor. In addition to Epo, HIF-1 regulates a number of hypoxia induced genes including vascular endothelial growth factor, tyrosine kinase and some glycolytic enzymes.

The critical roles of erythropoietin and HIF-1α in the response to hypoxia make the genes encoding them prime candidates for this study. A survey of the literature (PubMed, Nov. 1999) failed locate any references to functional polymorphisms in either of these gene. While this unfortunately precludes the type of analysis employed previously, it does not exclude the possibility that there are as yet unidentified variants in these genes that are unique to high altitude populations. This possibility was addressed by sequencing the coding region of both genes and the downstream regulatory region of the Epo gene in three unrelated Quechua. In addition, allele frequencies of a polymorphism in the 3’ region of the EPO gene, proximal to, but not in, the HIF-1 binding sequence (Percy et al., 1997), were determined in the Quechua and the Caucasians.
Materials and methods

DNA sequencing

1) Erythropoietin

Genomic DNA was prepared as described in chapter 2. The complete coding region of the erythropoietin gene was amplified from two Quechua DNA samples and one Quechua lymphoblast line (GM11197) by PCR using both exonic and intronic primers (see figure 6.3 and appendix ii). PCR (25 µl) were 0.2 mM dNTPs, 1.5 mM MgCl₂, 20 mM Tris/Cl pH 8.4 and 50 mM KCl and conditions were 94°C, 1 min.; 58°C, 1 min.; 72°C, 2 min. for 40 cycles. DNA sequencing was performed by NAPS.

Using BLAST, the final contiguous sequences were aligned to a published human Epo sequence (Kobilka et al., 1987; GenBank accession number M15169) and to Epo sequences in the non-redundant GenBank database. The resulting alignments were checked for mismatches that could be indicative of polymorphic loci.
Figure 6.3. Location of erythropoietin gene PCR primers used for amplification and sequencing of the coding regions (black) of the erythropoietin (Epo) gene. The 3' enhancer region is shaded

2) HIF-1α

Total RNA was prepared from three Quechua lymphoblast cell lines (GM11197, GM11198, GM11201) and cDNA prepared by reverse transcription as described in chapter 2. The cDNA was used as a template for PCR amplification of the coding sequence of the HIF-1α gene. PCRs (25 μl) were 0.2 mM dNTPs, 1.5 mM MgCl₂, 20 mM Tris/Cl pH 8.4 and 50 mM KCl and conditions were 94°C, 1 min.; 58°C, 1 min.; 72°C, 2 min. for 40 cycles using the following primer pairs: HIF-1α-S1 : HIF-1α-S2, HIF-1α-S3 : HIF-1α-S4, HIF-1α-S5 : HIF-1α-S6, HIF-1α-S7 : HIF-1α-S8 (see appendix ii). The same primers were used individually to sequence these amplification products. DNA sequencing was performed by NAPS at U.B.C.

The final contiguous sequences were aligned and checked for mismatches as described above. The reference sequence was from Wang et al., 1995 (GenBank accession number U22431).

Genotyping

The frequencies of both the C/T+662 and the newly identified T/G+769 polymorphisms were determined in Quechua and Caucasians. The region encompassing the C/T+662 was amplified using the primers EpoMae : Epo-4 (conditions described above). EpoMae is a degenerate primer designed such that a Mae III restriction endonuclease recognition site is generated in the presence of the C allele. Amplification product was digested with Mae III under the conditions prescribed by the manufacturer (GibcoBRL, Gaithersburg, MD) and visualized as described in chapter 2. Sample digest products are shown in figure 6.4a.

The T/G+769 polymorphism was assayed using a degenerate primer (EpoBbv) that generated a diagnostic Bbv I site when the G allele was present. This assay was chosen over
using the naturally variant Nla IV site because Nla IV does not function efficiently in the presence of KCl and therefore requires extra purification step prior to digestion to remove KCl carried over from the PCR reaction. DNA was PCR amplified using the primers Epo-7 : EpoBbv (reactions as described above). Amplification conditions were 94°C, 1 min.; 60°C, 1 min.; 72°C, 2 min. for 30 cycles. PCR products were then digested with the restriction endonuclease Bbv I under the conditions prescribed by the manufacturer (New England Biolabs, Beverly, MA) and assayed as described in chapter 2. Sample digestion products are shown in figure 6.4b.

Allele frequency data were analyzed as described in chapter 2. Significance was accepted at \( p < 0.025 \) to correct for the number of tests.

**Results**

1) Sequence and polymorphisms in the erythropoietin gene

The genomic sequence of the erythropoietin gene was determined in three Quechua (see appendix iii). Sequence data are available through GenBank (GenBank, accession numbers AF202306, AF202307, AF202308 (Quechua 1); AF202309, AF202310, AF202311 (Quechua 2); AF202312, AF202313 and AF202314 (Quechua 3)). One of the Quechua sequenced was heterozygous for an previously unreported polymorphism, a T to G transversion 769 bases downstream of the termination codon (T/G\(^{+769}\)) (figure 6.5). All other sequence was identical in the three Quechua. Alignments with various Epo sequences in the non-redundant DNA data base detected a variant base proximal to T/G\(^{+769}\), a C to T transition at base 662 (C/T\(^{+662}\)) (figure 6.5). This polymorphism had previously been reported in the literature (Percy et al., 1997). All three Quechua sequenced were homozygous for the C allele.
Figure 6.4. Assay for RFLPS in the erythropoietin gene. Representative polyacrylamide gels showing separation of digestion products for the RFLPs used to identify alleles of the 3' region of the erythropoietin gene. In addition to the products marked by the arrows, both reactions also generate a small (> 30 base) product when there is digestion at the polymorphic site. Genotypes are given across the top of each panel. The polymorphism and the diagnostic restriction endonuclease are given across the bottom. Fragment sizes are in base pairs. A T/T homozygote at C/T+662 was never observed. The apparent doublet bands in the right-hand image are gel artefacts.
Genotypes and allele frequencies for both polymorphisms in Quechua and Caucasians are given in tables 6.6 a,b. Allele frequencies of the T/G\textsuperscript{+769} polymorphism did not significantly differ between the Quechua and the Caucasians. The Quechua were monomorphic for the C allele at C/T\textsuperscript{+662} and allele frequencies differed significantly (p=0.007) from the Caucasians at this site. There was no significant linkage disequilibrium between alleles at the two loci in the Caucasian sample and genotype frequencies at both loci were in Hardy-Weinberg equilibrium.

2) Sequence of the hypoxia inducible factor-1\(\alpha\) gene.

The coding sequence of the HIF-1\(\alpha\) gene was determined in three Quechua (see appendix iii). No variants were detected within the Quechua sequences or between the Quechua and sequences in the GenBank database. A single representative sequence is given in appendix iii. All Quechua HIF-1\(\alpha\) sequences are available through GenBank, accession numbers AF207601, AF207602 and AF208487.

Discussion

1) Erythropoietin

No variation was detected between the coding sequences of the Epo gene in the three Quechua that were sequenced or between the Quechua sequences and the Epo coding
Table 6.1: Erythropoietin allele genotype and frequencies. 

a) Genotypes and allele frequencies for the C/T$^{+662}$ polymorphism in the erythropoietin gene in Quechua and Caucasian samples.  
b) Genotypes and allele frequencies for the T/G$^{+769}$ polymorphism in the erythropoietin gene in Quechua and Caucasian samples.

<table>
<thead>
<tr>
<th>Population</th>
<th>C/T$^{+662}$ genotypes</th>
<th>C/T$^{+769}$ genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C/C C/T T/T</td>
<td>T/T T/G G/G</td>
</tr>
<tr>
<td>Quechua</td>
<td>52 0 0</td>
<td>29 13 4</td>
</tr>
<tr>
<td>Caucasian</td>
<td>27 5 0</td>
<td>14 9 6</td>
</tr>
</tbody>
</table>

*Allele frequencies differ significantly (P<0.025).

**Allele frequencies do not differ significantly.
sequences available through GenBank. This does not preclude the existence of such variants; however, in a screen of six chromosomes, the probability of missing an allele present at a frequency of 50% is less than five percent.

One variant in the 3' untranslated region (C/T*662) was detected during alignment of Epo sequences aligned to Epo sequences in the GenBank database. This polymorphism had previously been described (Percy et al., 1997). It is presumed not to have any effect in the synthesis or function of Epo and was not over-represented in patients with polycythemia (Percy et al., 1997). The T allele at this loci was absent in the Quechua that were assayed, and at 9%, was significantly more frequent in the Caucasians.

A second polymorphism, T/G*769, was detected in one of the three Quechua sequenced. This polymorphic site is downstream of the HIF-1 binding site and does not alter the sequence of any of the regions known to be involved in hypoxic induction of the gene. Allele frequencies for this locus did not differ significantly between the Quechua and Caucasians.

Given their locations, it is unlikely that either of these polymorphisms affect the synthesis of Epo. The difference in allele frequencies between the Quechua and the Caucasians at C/T*662 could be accounted for by genetic drift. A second possible explanation is that alleles at this loci are linked to alleles at another loci that are under selection. Restriction fragment length polymorphisms detectable with Xba I (Semenza et al., 1987) and BglII (Beru and Payton, 1991) have been reported in the Epo gene. No phenotype has been described for either of these and the actual polymorphic base has not been identified.

Chronic mountain sickness is a condition that primarily occurs later in life in some long term high altitude residents. It is more common in Andean populations than in Himalayans, possibly due to the lower ventilatory response to hypoxia in the former population. The disease is characterized by extreme polycythemia secondary to chronic hypoxemia and may involve a loss of homeostatic regulation of erythropoiesis. The underlying mechanism is unknown but an over-reaction of erythropoietic tissue to hypoxemic stimulus has been proposed as a possible culprit (Monge-C et al., 1992).
Epo levels in Peruvian natives living in Lima (sea level) or Cerro de Pasco (4300 m) were measured by Léon-Velarde et al., (1991). They reported significantly higher levels of immunoreactive Epo in the Cerro de Pasco residents. Curiously, when they compared high altitude natives with and without excessive erythrocytosis, they found no difference in Epo levels, despite the polycythemic natives having a 20% greater blood hemoglobin concentration. A similar phenomenon was reported by Dainiak et al., (1988) in patients native to La Paz, Bolivia (3600 m) who had been diagnosed with chronic mountain sickness. Of the eight patients tested, only two had significantly elevated levels of Epo (~250 mU/ml). The remaining six, despite having hematocrits ranging from 64 - 79%, had Epo levels similar to controls (~25 mU/ml) who had been recruited at the same altitude. One possible explanation for this is that there are variants in Epo that alter its hematopoietic activity. It would be interesting to sequence the Epo genes of patients with Monge’s disease, especially comparing those with and without elevated Epo. It is intriguing to speculate that there might be a genetic contribution to the etiology of Monge’s disease; however, I am not aware of any evidence for a familial pattern of transmission having been reported for this condition.

2) HIF-1α

HIF-1 consists of two subunits, HIF-1α and HIF-1β. The β subunit is present in the cell under normoxic conditions and activation of the factor is determined by the availability of the α subunit, which appears to be oxygen dependent. The steady state levels of transcripts for both subunits are unaffected by oxygen tension, therefore regulation of the α subunit is presumed to be at the protein level. This suggests that mutations affecting HIF-1α activity would more likely be located in the coding sequence of the gene than in the transcriptional regulatory regions. No potential variants were identified in any of the three Quechua HIF-1α cDNA sequences. As discussed earlier, this does not preclude the existence of such a polymorphism, but the probably of missing a common (i.e. >50%) variant in such a screen is less than 5%.
Conclusions

A sequencing screen of six chromosomes failed to locate any variants in the coding regions of either the Epo or HIF-1α genes in the Quechua. One previously unreported polymorphism (T/G\textsuperscript{+769}) was found in the 3’ region of the Epo gene. This transversion did not alter any known regulatory sequences and there was no significant difference in allele frequencies between the Quechua and the Caucasians. A second downstream polymorphism (C/T\textsuperscript{+662}) that was proximal to, but not in, the HIF-1 binding site was also assayed in the Quechua and the Caucasians. Allele frequencies at this loci did differ significantly between the two populations. Whether this is due to non-directional population divergence or due to selection favouring an allele at a linked locus cannot be determined from these data.
In order to identify potential genetic contributions to high altitude adaptation in humans, the frequencies of allelic variants of a number of candidate genes, chosen because of their potential to influence oxygen delivery, were determined in Quechua-speaking Amerindians who were living at over 3200 m on the Andean altiplano in Peru. Based on the phenotypes associated with these alleles and the theoretical requirements of high altitude populations, hypotheses were made about the relative frequencies of these variants in the Quechua compared with two low altitude populations: Na-Dene speaking Amerindians from coastal B.C. and Caucasians of Western European descent. The four preceding chapters describe the results of these analyses, and although the general methodology and the samples analyzed are in common, each chapter represents an independent study and therefore incorporates its own discussion. The following section presents a brief summary of the results followed by a general discussion of the issues that should be considered in their interpretation.

The genes examined included the β-fibrinogen gene, the β2-adrenergic receptor gene and two genes involved in angiotensin-mediated vasomodulation. Variants in the β-fibrinogen gene that have been previously associated with lower levels of fibrinogen were significantly over-represented in the Quechua compared to both lowland samples. These data support the hypothesis that, because of their potential to reduce plasma viscosity, these alleles may have been selected for in the Quechua. Preliminary data did not demonstrate lower plasma viscosity in the Quechua than in Caucasians measured under the same conditions; however, the Caucasian sample size was quite small. Plasma viscosity is quite variable and is influenced by a large number of environmental and behavioral conditions. A large-scale analysis of plasma viscosity and fibrinogen concentration in the Quechua, as well as in other high altitude populations, would be an interesting follow-up to the genetic data presented in this thesis.

The insertion allele in the ACE gene, which had been previously associated with high-altitude endurance (Montgomery et al., 1997), was not over-represented in the Quechua
compared to lowlanders. As was discussed in chapter 4, while the genotypic data suggest that changes in the ACE gene are not associated with high altitude adaptation in the Quechua, a role for the enzyme in high altitude adaptation can not be excluded by these data. Given the efficacy of ACE blockers in the treatment of high altitude pulmonary hypertension (Niazova et al., 1996), it would be interesting to determine the ACE I/D genotype in individuals who have suffered from HAPE. An association between the ACE-deletion allele and susceptibility to HAPE could account for the under-representation of this variant in elite climbers reported by Montgomery and co-workers. Furthermore, identification of a potential genetic predisposition to HAPE could be of clinical importance in the prevention of this potentially fatal condition.

Significant differences in allele frequencies were observed for several polymorphic sites in the β2-AR gene. The most substantial changes were in silent mutations. These could be acting as linked markers for as yet uncharacterized functional mutations, although sequence data presented in this thesis suggests that there are no common mutations in the coding sequence in the Quechua (as represented by our sample). Future studies could undertake a large-scale polymorphism screen in, and around, this gene, as there may be functional changes in the regulatory regions flanking the coding sequence, or undetected variants in the body of the gene.

As of this writing, no polymorphisms in the coding region of the gene encoding HIF-1α have been reported in the literature. HIF-1α cDNA from three Quechua was sequenced and no variants were detected. Again, while this does not preclude the existence of such variants, it does suggest that they are not common in the Quechua communities that were sampled in this study. Lastly, the Epo gene was examined. Once again, no functional polymorphisms had been reported in the literature. Sequencing the gene in three Quechua detected a previously unreported polymorphism in the 3’ untranslated region of the gene, near the regulatory regions that control hypoxia-mediated expression. The allele frequencies of this polymorphism did not differ between Quechua and Caucasians, although allele frequencies for a previously reported polymorphism in the 3’ untranslated region did differ between the two populations. Neither of these polymorphisms alters the coding sequence or any known regulatory sites and therefore is unlikely to generate a phenotype. However, as discussed below, they could be linked to other
variants that do alter activity. Because of the role of Epo in haematopoiesis, it would be interesting to extend the search for mutations in the Epo gene to patients diagnosed with chronic mountain disease, particularly in cases where Epo concentrations are normal despite pathologically high hematocrits.

One advantage to examining the genetic characteristics of a population (as opposed to physiological characteristics) is that, within a generation, genotype is independent of environment. Measuring the actual concentration or activity of the products of the genes that were examined in this project would be substantially more difficult as most of them vary with the physical condition, age, gender, health and lifestyle of the subjects. The disadvantage of working exclusively at the DNA level is that selection acts on phenotype, not genotype, and the relationship between genotype and phenotype may vary for a number of reasons including environment, genetic background and haplotype structure (linkage disequilibrium).

An example of the complex interactions between genotype and environment is the relationship between the frequency of the hemoglobin allele that causes sickle cell anemia and the prevalence of malaria. The abnormal red blood cells arising from the hemoglobin variant are less efficient oxygen carriers, and therefore deleterious, but they also confer resistance to malaria, a beneficial trait in regions where the disease is common. Predictions about the prevalence of the allele that causes sickling would have to take into consideration where the subject population lived. Another example is the common C/T<sup>677</sup> polymorphism in methylenetetrahydrofolate reductase (MTHFR). This allele is associated with vascular deficiencies, neural tube defects and cancer. However, as these phenotypes only manifest in individuals whose diets are lacking in folic acid (see review by Bailey and Gregory 3<sup>rd</sup>, 1999), the selective pressure to remove this obviously deleterious allele would be considerably blunted in populations with a diet that included sufficient amounts of this vitamin.

This project focused only on a single environmental condition: hypobaric hypoxia. While this is a pervasive and immutable stress, it is unlikely to be the only environmental variable differentiating the three study populations, and variables other than altitude of residence were not controlled for in these experiments (although both native American populations were sampled
from non-urban areas). It is possible that additional environmental variables could blunt, or exacerbate, the selective pressures that were predicted to result from hypoxia. ACE and β-fibrinogen genotypes have been associated with cardiovascular disease, and any consideration of selection acting upon alleles in these genes should take into account the myriad other environmental factors, such as diet and lifestyle, that influence cardiovascular health and therefore could affect selective pressures. A similar argument could be made for the alleles of the B_2AR gene that influence (variably) fat deposition.

Phenotype may also vary between populations due to differences in genetic background. In the case of the ACE I/D polymorphism, correlations were reported between the deletion allele and cardiovascular disease in European, American and Japanese studies (reviewed in Butler et al., 1997) but not Pima Indians (Nagi et al., 1998), Chinese (Chuang et al., 1997; Saha et al., 1996) or Southeast Asians (Saha et al., 1996). Again, this could be due to environmental differences but could also reflect intrinsic differences in the populations. The interaction between ACE and the ACE receptor demonstrates that ACE phenotype may not be predictable without knowing receptor genotype (Tiret et al., 1994). A population monomorphic for the latter may be invariant at the former regardless of genotype. It was only assumed that there was an association between ACE I/D genotype and ACE levels this population. If there was not, then there would not be selective pressure favouring the ACE allele even if reduced ACE levels were advantageous.

Given the complexity of human physiology, changes in the function of a single gene may impact upon more than one characteristic. Predictions of positive selective value, as were made in this project, can be easily confounded if the allele has an over-riding deleterious effect elsewhere. Again, the ACE insertion/deletion polymorphism serves as an example. While the insertion allele appears to confer some cardiovascular advantage, it is also associated with the development of neurological pathology (Kehoe et al., 1999), and selection will favour the allele that is of most overall benefit to the population. This can be difficult to predict. Late onset diseases such as Alzheimer's tend to be post-reproductive, and a predisposition to these conditions will not be selected against unless there is some social role for aged individuals that enhances the reproductive fitness of the population. This raises an important point: to be selected for an allele
must confer a reproductive advantage. If the elite climbers in whom Montgomery and co-
workers (1997) reported a high frequency of the ACE insertion allele spend all their time
climbing mountains and never settle down to have families, the allele, regardless of how
beneficial it might be, could likely become less common in the population.

For several of the polymorphisms examined in this project the phenotypes are based on
association rather than demonstrated functional changes, and it must be considered that the alleles
may not be directly responsible for the phenotype. In fact, two of the five \( B_2 \)-AR alleles
examined did not alter the protein coding sequence, both polymorphisms in Epo were
downstream of the coding region and not in defined regulatory regions, and the polymorphism in
ACE was within an intron. As these changes would not be anticipated to have a phenotypic
effect, the association with a phenotype may well be due to linkage disequilibrium with as yet
unreported alleles that do cause the phenotypic effects. Linkage disequilibrium is seen when
alleles at adjacent polymorphic loci do not assort independently during meiosis and are
transmitted between generations as haplotypes (sets of co-segregating alleles at linked loci) rather
than as individual alleles. For example, the data presented in chapter 4 show that the alleles at the
three \( \beta \)-fibrinogen loci appear to be inherited as the same two haplotypes in all three populations
examined. The three sites are quite close together (within 1000 bases) and presumably,
insufficient time has passed since these mutations arose for recombination to “shuffle” the
alleles such that they are transmitted independently of one another.

The extent of linkage disequilibrium between loci can vary between populations. This was
evident in the putative \( B_2 \)-AR haplotypes that varied between the two Amerindian populations and
the Caucasians. Once a mutation occurs, its separation from syntenic (\textit{i.e.} on the same
chromosome) alleles requires a recombinant event between the two loci, and the probability of
this occurring depends on the distance separating them. Populations can differ due to
recombination events that occurred exclusively in their ancestors, because of unequal
distributions of haplotypes during subsequent population divisions, or due to genetic drift once
populations were separated. Superimposed on these random sources of variation is the
possibility that any of the associated alleles confers a selective advantage, in which case the
frequency of all of the linked alleles will increase. Potential differences in populations can complicate association studies because if the allele that is being assayed is only a marker for a functional allele, then the association will only be observed in populations in which the two loci are in disequilibrium.

The advantage to linkage disequilibrium in genetic analysis is that it increases the probability of detecting functional variants, because changes in the frequencies of these (which tend to far less common than inert mutations) will be mirrored in any allele to which they are tightly linked. For example, two silent mutations were assayed in the B2-AR gene, neither of which would be expected to have any effect on receptor function. Of the five alleles examined, these two (actually one haplotype) varied the most between the Quechua and the two lowland populations. This could be due to them being in disequilibrium with alleles at locus that has been selected for in the Quechua. Sequencing the coding region of three Quechua failed to find this potentially interesting site, but that does not rule out selectable variations in regulatory regions of the gene (or even in proximal genes).

Genetic variation between populations is not solely the result of selection. While differing allele frequencies may be consistent with selection favouring one allele over the alternative, they may also have arisen from the non-directional genetic divergence that inevitably occurs when populations are reproductively separated for long periods of time. Some of this variation will occur due to random changes in frequency commonly referred to as genetic drift. This usually occurs when a breeding population is sufficiently small that significant changes in allele frequency may occur by chance alone. A simple example involves a population of two individuals, both of whom are heterozygous at a locus for which there are two alleles (i.e. Aa and Aa). If they have two offspring, there is a one in eight (12.5%) chance that these offspring will be homozygous for the same allele (i.e. AA, AA or aa, aa) and therefore one allele will have been eliminated from the population. Obviously, this is an extreme case, but it serves to illustrate that significant change can occur by chance alone, especially if the population is small (if there were five couples, the chances of losing an allele in one generation would be less than 0.005%). Even in large populations, genetic drift can significantly effect the distribution of alleles. This is
especially true if the alleles are uncommon to begin because, once lost within a population, an allele is unlikely to be replaced. The absence of the T allele of the Epo C/T\textsuperscript{659} polymorphism in the Quechua could well be due to drift as the allele may not have been that common to begin with (it was found to be at 9% in the Caucasian sample). The same is true for the G\textsuperscript{79} allele in the β\textsubscript{2}AR gene, which is present at 9% in the Na-Dene but absent in the Quechua. Causes of monomorphism are difficult to assess because there is no way of estimating the time scale over which the loss occurred.

Founder effect, where a population is descended from a small progenitor group is a special case of genetic drift. Analysis of mitochondrial genotypes in South American populations (Monsalve et al., 1994) suggests that the there was not a significant founder effect in the colonizing of South America, and the Quechua sampled for this project were heterogeneous at most of the loci examined, consistent with substantial variation their ancestral population.

There will always be some variation between populations due to chance. If large numbers of randomly selected polymorphisms are examined, the probability of detecting significant differences in allele frequencies by chance alone can become quite high (the number of tests can be corrected for but may require prohibitively large sample sizes). Focusing on candidate genes is one way of addressing this issue. The chance of spurious associations can be reduced, while the chance of detecting a valid association is increased (to the extent that the "candidacy" is legitimate). Another way to address the possibility of observed differences being due to chance is to examine other populations that face similar selective pressures. Random variation would be expected to differ between the populations, whereas similar trends should be seen if the changes are directional and, as most researchers consider consistent replication to be the best evidence for true association (Xu et al., 1998), the most obvious follow-up to this project would be a similar analysis in Asian high altitude populations, such as Sherpas or Tibetans.

The aforementioned caveats notwithstanding, the analysis of association between genotype and phenotype is a potent tool in separating the effects of nature and nurture. Heritability studies and family analysis are useful techniques with which to establish whether there is a genetic component contributing to the development of complex traits and adaptive
phenotypes; however, ultimate elucidation may require teasing out genetic factors on a gene by gene basis (Risch and Merikangas, 1996). As there are an estimated 50,000-100,000 genes in the human genome, some a priori reason to suspect the gene product of playing a role in the phenotype would be needed to pre-screen genes for analysis. Even with gene candidacy, such an analysis would be a long and expensive process, given the immense complexity of many physiological processes (e.g. the clotting reaction, the ostensibly straightforward haemostatic response, involves more than 500 gene products). A further confounding factor is that a genetic variant may need to be superior in only one tissue, for one moment, prior to reproduction in order to confer a selective advantage. That the benefit conferred by an allele may not be apparent in the extant organism greatly complicates the task of predicting what would, or would not be, under selective pressure.

While this study was never intended to be an extensive analysis of gene frequencies in the Quechua or the Na-Dene, the data that it generated should be of interest to researchers who study human origins, distribution and diversity. There may be a finite window of opportunity to collect data of this sort as advances in communication and transportation, as well as changing attitudes toward cultural boundaries, tend to lead to a homogenization of the race. As eminent population geneticist Luca Cavalli-Sforza expressed it in 1991, the founding year of The Human Genome Diversity Project: "The genetic diversity of people now living harbours clues to the evolution of our species, but the gate to preserve these clues is closing rapidly."

Eventually, the full extent of human diversity may be known. The emerging technology of microarrays, combined with the soon to be completed human genome project (and the single nucleotide polymorphism project) offers the promise of determining the entire molecular phenotype of an individual in a single experiment. This powerful methodology may eventually allow researchers to characterise the genetic composition of entire populations, and perhaps finally resolve the issue of whether, over the course of the past hundred centuries, the sturdy and vigorous people who live high in the thin air of the Andes mountains have evolved in response to the demands of life on the altiplano.

*consumatum est*  
*Jim Rupert 21/4/00*
## Abbreviations

The following is a list of abbreviations that appear more than once in the text. Each abbreviation is also defined after its initial usage. Standard chemical and metric abbreviations are not defined.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE</td>
<td>angiotensin converting enzyme</td>
</tr>
<tr>
<td>AT-2</td>
<td>angiotensin 2</td>
</tr>
<tr>
<td>AT2-R1</td>
<td>angiotensin 2, receptor type 1</td>
</tr>
<tr>
<td>AT2-R2</td>
<td>angiotensin 2, receptor type 2</td>
</tr>
<tr>
<td>BMI</td>
<td>body mass index</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>β2-AR</td>
<td>beta 2 adrenergic receptor</td>
</tr>
<tr>
<td>dNTPs</td>
<td>deoxynucleotides</td>
</tr>
<tr>
<td>Epo</td>
<td>erythropoietin</td>
</tr>
<tr>
<td>Epo-R</td>
<td>erythropoietin receptor</td>
</tr>
<tr>
<td>EtBr</td>
<td>ethidium bromide</td>
</tr>
<tr>
<td>FVC</td>
<td>forced vital capacity</td>
</tr>
<tr>
<td>GRE</td>
<td>glucocorticoid response element</td>
</tr>
<tr>
<td>HAPE</td>
<td>high altitude pulmonary edema</td>
</tr>
<tr>
<td>HIF</td>
<td>hypoxia inducible factor</td>
</tr>
<tr>
<td>HRE</td>
<td>hormone response element</td>
</tr>
<tr>
<td>HVR</td>
<td>hypoxic ventilatory response</td>
</tr>
<tr>
<td>MCH</td>
<td>mean corpuscular hemoglobin content</td>
</tr>
<tr>
<td>MCHC</td>
<td>mean corpuscular hemoglobin conc.</td>
</tr>
<tr>
<td>mtDNA</td>
<td>mitochondrial DNA</td>
</tr>
<tr>
<td>NAPS</td>
<td>nucleic acid and protein services</td>
</tr>
<tr>
<td>PaO₂</td>
<td>partial pressure of oxygen, arterial</td>
</tr>
<tr>
<td>PAO₂</td>
<td>partial pressure of oxygen, alveolar</td>
</tr>
<tr>
<td>PO₂</td>
<td>partial pressure of oxygen</td>
</tr>
<tr>
<td>PiO₂</td>
<td>partial pressure of oxygen, inspired</td>
</tr>
<tr>
<td>RBC</td>
<td>red blood cell</td>
</tr>
<tr>
<td>RFLP</td>
<td>restriction fragment length polymorphism</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>SaO₂</td>
<td>arterial O₂ saturation</td>
</tr>
<tr>
<td>U.B.C.</td>
<td>University of British Columbia</td>
</tr>
<tr>
<td>VO₂max</td>
<td>O₂ uptake, maximal rate</td>
</tr>
<tr>
<td>ybp</td>
<td>years before present</td>
</tr>
<tr>
<td>WBC</td>
<td>white blood cell</td>
</tr>
</tbody>
</table>
References


adrenoceptor polymorphism with lower airway reactivity in asthmatic subjects. Lancet 345, 1213-
1214.


Harkness, J. (1971). The viscosity of human blood plasma; its measurement in health and

University Press).

Hebbel, R. P., Eaton, J. W., Kronenberg, R. S., Zanjani, E. D., Moore, L. G. and Berger, E. M.

2, 197-205.

Impact of polymorphisms in the alpha- and beta- fibrinogen gene on plasma fibrinogen


Hochachka, P. W., Stanley, C., Matheson, G. O., McKenzie, D. C., Allen, P. S. and Parkhouse,
70(4), 1720-1730.

Hochachka, P. W., Clark, C. M., Brown, W. D., Stanley, C., Stone, C. K., Nickles, J., Allen, P. S.
and Holden, J. E. (1994). The brain at high altitude: Hypometabolism as a defense against

physiological systems in the evolution of human hypoxia tolerance. Comp. Biochem. Physiol
124A (1), 1-17.


Thomas, A., Lamlum, H., Humphries, S. and Green, F. (1994). Linkage disequilibrium across the fibrinogen locus as shown by five genetic polymorphisms, G/A-455 (HaeIII), C/T-148 (HindIII/AluI), T/G (AvaII) and BclI (β-fibrinogen) and TaqI (α-fibrinogen) and their detection by PCR. Hum. Mut. 3, 79-81.


Umemura, S., Kawasaki, T., Ishigami, T., Fujita, T., Hibi, K., Kawasaki, M., Itoh, K., Yoshimizu,

venodilatation in Native Americans. Int. Angiol. 17 (4), 276-281.

Baker and M. A. Little, eds. (Stroudsburg, Pennsylvania: Dowden, Hutchinson and Ross, Inc.),

Pulmonary gas exchange, diffusing capacity in natives and newcomers at high altitude. Resp.
Physiol. 34, 219-231.

Acad. Sci. (USA) 92 (12), 5510-5514.


M. A. Little, eds. (Stroudsburg, Pennsylvania: Dowden, Hutchinson and Ross, Inc.).

James, A., Liggett, S. B. and Paré, P. D. (1998). $\beta_2$-adrenergic receptor haplotypes in mild, 

110


Appendix i. Information form

### Questionnaire for High Altitude Native Study

#### Donor Information

<table>
<thead>
<tr>
<th>Name:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Last</td>
<td></td>
</tr>
<tr>
<td>First</td>
<td></td>
</tr>
<tr>
<td>M.I.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sex:</th>
<th>M</th>
<th>F</th>
<th>Age:</th>
<th>Years</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Is the Donor's General Health Good?:</th>
<th>Yes</th>
<th>No</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Does the donor have a history of heart disease?:</th>
<th>Yes</th>
<th>No</th>
<th></th>
</tr>
</thead>
</table>

### Smoking History:

<table>
<thead>
<tr>
<th>Non-smoker:</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Smokes:</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>1-5 cigarettes per day:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5-15 cigarettes per day:</td>
<td>(check one box)</td>
</tr>
<tr>
<td>more than 15 cigarettes per day:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Former smoker - years since stopped smoking:</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Place of residence:</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Place of birth:</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Occupation:</th>
<th></th>
</tr>
</thead>
</table>

Has the donor spent more than 24 hours at an altitude greater than 1000M **above** their home in the last year?:  
<table>
<thead>
<tr>
<th>yes</th>
<th>no</th>
<th>How much time?:</th>
<th></th>
</tr>
</thead>
</table>

Has the donor spent more than 24 hours at an altitude greater than 1000M **below** their home in the last year?:  
<table>
<thead>
<tr>
<th>yes</th>
<th>no</th>
<th>How much time?:</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Sample volume:</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Date taken:</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Time taken:</th>
<th></th>
</tr>
</thead>
</table>

| To be filled out in lab: |  |

<table>
<thead>
<tr>
<th>Lab identification number:</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Comments:</th>
<th></th>
</tr>
</thead>
</table>

J.L.R. 30/4/98
**Appendix ii.** PCR primer sequences. Alternate bases in degenerate primers are underlined and in bold face. All primers were prepared by NAPS at U.B.C.

### Angiotensin 2 receptor type 1 gene

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT2R1-3</td>
<td>5' ATA ATG TAA GCT CAT CCA CC 3'</td>
<td>AT2R1-4</td>
<td>5' GAG ATT GCA TTT CTG TCA GT 3'</td>
</tr>
</tbody>
</table>

### Angiotensin converting enzyme (ACE) gene

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE-1</td>
<td>5' CAT CCT TTC TCC CAT TTC TC 3'</td>
<td>ACE-2</td>
<td>5' TGG GAT TAC AGG CTT GAT ACA G 3'</td>
</tr>
<tr>
<td>ACE-3</td>
<td>5' ATT TCA GAG CTG GAA TAA AAT T 3'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Erythropoietin gene

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epo-1</td>
<td>5' GCG CCC CAG GTC GCT GAG 3'</td>
<td>Epo-2</td>
<td>5' CAC CAA CAT TGC TTG TGC CAC 3'</td>
</tr>
<tr>
<td>Epo-3</td>
<td>5' ACC AGG TGT GTC CAC CTG G 3'</td>
<td>Epo-4</td>
<td>5' CCC AGG GCA AGG CTT GAG 3'</td>
</tr>
<tr>
<td>Epo-5</td>
<td>5' CCT GGC CCT GCT GTC GGA A 3'</td>
<td>Epo-6</td>
<td>5' AAC TCT TCC CAG CCG TGG G 3'</td>
</tr>
<tr>
<td>Epo-7</td>
<td>5' TAC CTG TTT TCG CAC CTA CC 3'</td>
<td>Epo-8</td>
<td>5' GGA GAA CTT AGG TGG CAA GC 3'</td>
</tr>
<tr>
<td>Epo-9</td>
<td>5' AAG CTG AAG CTG TAC ACA GG 3'</td>
<td>Epo-10</td>
<td>5' AGG AGG GAG AGG GTG ACA T 3'</td>
</tr>
<tr>
<td>Epo-11</td>
<td>5' TCC CTC CCC GCC TGA CTC T 3'</td>
<td>Epo-12</td>
<td>5' ATC TCA TTT GCG AGC CTG AT 3'</td>
</tr>
<tr>
<td>Epo-13</td>
<td>5' CAG GGC CCG GGA GCA GCC 3'</td>
<td>Epo-14</td>
<td>5' GGT CTC TGT TTG AGC GGG 3'</td>
</tr>
<tr>
<td>Epo-15</td>
<td>5' GAT TTA GCG CCC CGG CTA TT 3'</td>
<td>Epo-16</td>
<td>5' AGT TCC TTT TTT TTT TTG CC 3'</td>
</tr>
<tr>
<td>Epo-17</td>
<td>5' GAT GGA GGT GAG TTC CTT TT 3'</td>
<td>Epo-Nla</td>
<td>5' TCC AAA TCC CCT GGC TCT G 3'</td>
</tr>
<tr>
<td>Epo-Bbv</td>
<td>5' CTG CAT TCA AGG CCT CAC CC 3'</td>
<td>Epo-Mae</td>
<td>5' CAG GTC CAG GTC CCG GTA A 3'</td>
</tr>
</tbody>
</table>
### Appendix ii (continued). PCR primer sequences

#### $\beta_2$-adrenergic receptor gene

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_2$AR-46Neo</td>
<td>5' CCT TCT TGC TGG CAC CCC AT 3'</td>
<td>$\beta_2$AR-P2a</td>
<td>5' CCA GCA CAT TGC CAA ACA CG 3'</td>
</tr>
<tr>
<td>$\beta_2$AR-P1</td>
<td>5' CGT GGG 'TCC GCC TGC TGA 3'</td>
<td>$\beta_2$AR-P1A</td>
<td>5' ACC AGA ACA TTG CCA AAC AC 3'</td>
</tr>
<tr>
<td>$\beta_2$AR-P2</td>
<td>5' CAC AGG GTC TCA ATG CTG GC 3'</td>
<td>$\beta_2$AR-P3</td>
<td>5' CCC ATA TTC TTA TGA AAA TG 3'</td>
</tr>
<tr>
<td>$\beta_2$AR-P4</td>
<td>5' GCC CTG AGA TTG TGT AAT CT 3'</td>
<td>$\beta_2$AR-P5</td>
<td>5' CAT GGT CTT CGT CTA CTC CA 3'</td>
</tr>
<tr>
<td>$\beta_2$AR-P6</td>
<td>5' CGG TTG CTG GAG TAG CCA TT 3'</td>
<td>$\beta_2$AR-P7</td>
<td>5' GAT AAC CTC ATC CGT AAG GA 3'</td>
</tr>
<tr>
<td>$\beta_2$AR-P8</td>
<td>5' TTA GTC GTC TGT TTA GIG TTC TG 3'</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### $\beta$-fibrinogen gene

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fib-1A</td>
<td>5' ATT GAG TCA CGA TTT TAG TG 3'</td>
<td>Fib-1B</td>
<td>5' GAA ATG TTT ACC TTT CTG TA 3'</td>
</tr>
<tr>
<td>Fib-1Arev</td>
<td>5' CAC TAA AAT CTG GAC TCA TT 3'</td>
<td>Fib-1C</td>
<td>5' AAG GGT CTT TCT GAT GTG TA 3'</td>
</tr>
<tr>
<td>Fib-B6</td>
<td>5' GGG ACA TGG CAA AGC ATG G 3'</td>
<td>Fib-B7</td>
<td>5' GGT GAG CAA GAG AAA TGA AG 3'</td>
</tr>
</tbody>
</table>

#### Hypoxia inducible factor-1 alpha (HIF-1$\alpha$) gene

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIF-1$\alpha$-S1</td>
<td>5' GTG AAG ACA TCG CGG GGA 3'</td>
<td>HIF-1$\alpha$-S2</td>
<td>5' CAG GAT CCT TGC CAA AAG AG 3'</td>
</tr>
<tr>
<td>HIF-1$\alpha$-S3</td>
<td>5' TCA TGC TTT GGA CTC TCA TT 3'</td>
<td>HIF-1$\alpha$-S4</td>
<td>5' TTC CAG TTA CGT TCC TTC GA 3'</td>
</tr>
<tr>
<td>HIF-1$\alpha$-S5</td>
<td>5' CTC AGG ACA CAG ATT TAG AC 3'</td>
<td>HIF-1$\alpha$-S6</td>
<td>5' CTC ACT ACC TAA AGC AGT CT 3'</td>
</tr>
<tr>
<td>HIF-1$\alpha$-S7</td>
<td>5' CAA ACT GAG TTA ATC CCA TGT 3'</td>
<td>HIF-1$\alpha$-S7rev</td>
<td>5' CAT GGG ATT AAC TCA GTT TG 3'</td>
</tr>
<tr>
<td>HIF-1$\alpha$-S8</td>
<td>5' GCA ACT TGA GGA AGT ACC AT 3'</td>
<td>HIF-1$\alpha$-S8rev</td>
<td>5' GCA ACT TGA GGA AGT ACC AT 3'</td>
</tr>
<tr>
<td>HIF-1$\alpha$-S9</td>
<td>5' ACT ACA GTT CCT GAG GAA GA 3'</td>
<td>HIF-1$\alpha$-S9rev</td>
<td>5' ACT ACA GTT CCT GAG GAA GA 3'</td>
</tr>
<tr>
<td>HIF-1$\alpha$-S10</td>
<td>5' CTT ACC ATC AGC TAT TGG CG 3'</td>
<td>HIF-1$\alpha$-S11</td>
<td>5' GAC AAA CCT AAG AAG GAA CC 3'</td>
</tr>
</tbody>
</table>
Appendix iii. DNA sequences

a) Alignment of the DNA sequence of three regions of the \( \beta \)-fibrinogen gene sequenced from a chimpanzee (Pan troglodytes) cell line and previously published human sequences. Vertical bars indicate consensus between both sequences. Small case denotes untranslated sequence (5' region) whereas large case is coding sequence. The sites of the G/C\(^{455}\), C/T\(^{148}\) and G/A\(^{448}\) polymorphisms are shown. Human sequences are 1) human fibrinogen beta gene 5' region and exon 1 (GenBank accession number X05018) and 2) human fibrinogen beta-chain mRNA, (GenBank accession number J100129), exon 8. Chimpanzee sequences are available through GenBank: accession numbers AF200354, AF200355 and AF200356.

Chimp: 1  ttcatagaataggtatgaatttgttattttgttattttgattaatgtctaaaacaaaag 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Human1: 946 ttcatagaataggtatgaatttgttattttgttattttgattaatgtctaaaacaaaag 1005

Chimp: 61  ataaacacattatgataatacattactattgattttataatggcccctttgaaatagaatt 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Human1: 1006 ataaacacattatgataatacattactattgattttataatggcccctttgaaatagaatt 1065

G/C\(^{455}\)

Chimp: 121 atgtcattgtcagaaaacataagcatttatggtatatcatt 161
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Human1: 1066 atgtcattgtcagaaaacataagcatttatggtatatcatt 1106

Chimp: 1  ttgcccttgtgagtaggtcaaatttactaagcttagatttgttttctcacatattctttcg 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Human1: 1128 ttgcccttgtgagtaggtcaaatttactaagcttagatttgttttctcacatattctttcg 1187

Chimp: 61  gaggctgtgtagtttccacattaatttaccagaaacaagatacacatctctctttgagga 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Human1: 1188 gaggctgtgtagtttccacattaatttaccagaaacaagatacacatctctctttgagga 1247

Human1: 1248 gtgccctaacttcccatcattttgtccaattaaatgaattgaagaaatttaatgtttcta 1307

Chimp: 181  aactagaccaacaaagaataaatgtttatgtatgacgaattaaatagttttgggaagatg 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Human1: 1308 aactagaccaacaaagaataaatgtttatgacgaattaaatagttttgggaagatg 1367

C/T\(^{148}\)

Chimp: 241 ttgcttaaatgataaaatggttcagccaacaagtgaaccaaaaattaaatattaac 296
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Human1: 1368 ttgcttaaatgataaaatggttcagccaacaagtgaaccaaaaattaaatattaac 1423

G/A\(^{448}\)

Chimp: 1  GATGATGGTGTAGTATGGATGAATTGGAAGGGGTCATGGTACTCAATGAAGAAGATGAGT 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Human2: 1360 GATGATGGTGTAGTATGGATGAATTGGAAGGGGTCATGGTACTCAATGAAGAAGATGAGT 1419
Appendix iii a continued.

Chimp: 61  ATGAAGATCAGGCCCTTCTTCCCACAGCAATAGTCCCCAATA7AGTAGATTTTTGCTCTTC 120

Human2: 1420 ATGAAGATCAGGCCCTTCTTCCCACAGCAATAGTCCCCAATA7AGTAGATTTTTGCTCTTC 1479

Chimp: 121  TGTATGTGACAACATTTTTGTACATTATGTTATTGGAATTTTCTTTCATACATTATATTC 180

Human2: 1480 TGTATGTGACAACATTTTTGTACATTATGTTATTC 1539

Chimp: 181  CTCTAAAACTCTCAAGCAGACGTGAGTGTGACTTTTTGAAAAAAGTATAGGATAAAATTAC 240

Human2: 1540 CTCTAAAACTCTCAAGCAGACGTGAGTGTGACTTTTTGAAAAAAGTATAGGATAAAATTAC 1599

Chimp: 241  ATTAAAATA 249

Human2: 1600 ATTAAAATA 1608
Appendix iii. DNA sequences (continued)

b) Sequence of the β2-adrenergic receptor gene in three Quechua speaking natives. DNA sequence of three Quechua are aligned with the sequenced published in (Kobilka et al., 1987) (GenBank accession number M15169). Numbering starts at the translation start site, the coding sequence is in upper case letters. The initiation codon and termination codons are in boldface. Known polymorphic sites are underlined and those described in the text are indicated by an arrow. All discrepancies between sequences occur at known polymorphic sites. Heterozygosity is indicated using DNA degenerate code (R= A/G, M = A/C and Y = T/C) Polymorphisms not described in the text are described in (Reihsaus et al., 1993). Quechua sequences are available through GenBank, accession numbers AF169225, AF202305 and AF203386.

<table>
<thead>
<tr>
<th>Kobilka, 87:</th>
<th>Quechua 1:</th>
<th>Quechua 2:</th>
<th>Quechua 3:</th>
</tr>
</thead>
<tbody>
<tr>
<td>-30 ccagtgccttaacctgccagactgcgccATGGGGCAACCCGGGAACGGCAGCGCCTTC</td>
<td>ccagtgccttaacctgccagactgcgccATGGGGCAACCCGGGAACGGCAGCGCCTTC</td>
<td>tgcagactgcgccATGGGGCAACCCGGGAACGGCAGCGCCTTC</td>
<td>gccttaacctgccagactgcgccATGGGGCAACCCGGGAACGGCAGCGCCTTC</td>
</tr>
<tr>
<td>31 TTGCTGGCACCCAATAGAAGCCATGCGCCGGACCACGACGTCAGGCAGCAAAGGGACGAG</td>
<td>TTGCTGGCACCCAATGGAAGCCATGCGCCGGACCACGACGTCAGGCAGCAAAGGGACGAG</td>
<td>TTGCTGGCACCCAATGGAAGCCATGCGCCGGACCACGACGTCAGGCAGCAAAGGGACGAG</td>
<td>TTGCTGGCACCCAATAGAAGCCATGCGCCGGACCACGACGTCAGGCAGCAAAGGGACGAG</td>
</tr>
<tr>
<td>91 GTGCTGGGTGGTCATGGGCATCGTCATGTCTCTCATC</td>
<td>GTGCTGGGTGGTCATGGGCATCGTCATGTCTCTCATC</td>
<td>GTGCTGGGTGGTCATGGGCATCGTCATGTCTCTCATC</td>
<td>GTGCTGGGTGGTCATGGGCATCGTCATGTCTCTCATC</td>
</tr>
<tr>
<td>210 AATGTGCTGGATCAGAATGGGACAGGGATCTATGGATGCTGTCACGAGCTCAGAACTAC</td>
<td>AATGTGCTGGATCAGAATGGGACAGGGATCTATGGATGCTGTCACGAGCTCAGAACTAC</td>
<td>AATGTGCTGGATCAGAATGGGACAGGGATCTATGGATGCTGTCACGAGCTCAGAACTAC</td>
<td>AATGTGCTGGATCAGAATGGGACAGGGATCTATGGATGCTGTCACGAGCTCAGAACTAC</td>
</tr>
<tr>
<td>270 TTCATCACTTCACTGGGCCTGTGCTGATCTGGTCATGGGCCTAGCAGTGGTGCC</td>
<td>TTCATCACTTCACTGGGCCTGTGCTGATCTGGTCATGGGCCTAGCAGTGGTGCC</td>
<td>TTCATCACTTCACTGGGCCTGTGCTGATCTGGTCATGGGCCTAGCAGTGGTGCC</td>
<td>TTCATCACTTCACTGGGCCTGTGCTGATCTGGTCATGGGCCTAGCAGTGGTGCC</td>
</tr>
</tbody>
</table>

Kobilka, 87: 30-30 ccagtgccttaacctgccagactgcgccATGGGGCAACCCGGGAACGGCAGCGCCTTC 30

Kobilka, 87: 31 TTGCTGGCACCCAATAGAAGCCATGCGCCGGACCACGACGTCAGGCAGCAAAGGGACGAG 90

Kobilka, 87: 91 GTGCTGGGTGGTCATGGGCATCGTCATGTCTCTCATC 150

Kobilka, 87: 151 AATGTGCTGGATCAGAATGGGACAGGGATCTATGGATGCTGTCACGAGCTCAGAACTAC 210

Kobilka, 87: 211 TTCATCACTTCACTGGGCCTGTGCTGATCTGGTCATGGGCCTAGCAGTGGTGCC 270

G/A↑

C/G↑
Appendix iii b continued

| Kobilka, 87: | 271 GCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGA |
| Quechua    | GCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGA |
| Quechua    | GCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGA |
| Quechua    | GCCGCCCATATTCTTATGAAAATGTGGACTTTTGGCAACTTCTGGTGCGA |

| Kobilka, 87: | 331 TCCATTGATGTGCTGTGCGTCACGGCCAGCATTGAGACCCTGTGCGTGATCGCAGTGGAT |
| Quechua    | TCCATTGATGTGCTGTGCGTCACGGCCAGCATTGAGACCCTGTGCGTGATCGCAGTGGAT |
| Quechua    | TCCATTGATGTGCTGTGCGTCACGGCCAGCATTGAGACCCTGTGCGTGATCGCAGTGGAT |
| Quechua    | TCCATTGATGTGCTGTGCGTCACGGCCAGCATTGAGACCCTGTGCGTGATCGCAGTGGAT |

| Kobilka, 87: | 391 CGCTACTTTGCCATTACTTCACCTTTCAAGT^CAGAGCCTGCTGACCAAGAATAAGGCC |
| Quechua    | CGCTACTTTGCCATTACTTCACCTTTCAAGT^CAGAGCCTGCTGACCAAGAATAAGGCC |
| Quechua    | CGCTACTTTGCCATTACTTCACCTTTCAAGT^CAGAGCCTGCTGACCAAGAATAAGGCC |
| Quechua    | CGCTACTTTGCCATTACTTCACCTTTCAAGT^CAGAGCCTGCTGACCAAGAATAAGGCC |

| Kobilka, 87: | 451 CGGGTGATCATTCTGATGGTGTGGATTGTGTCAG<X:CTTACCTCCTTCTTGCCCATTCAG |
| Quechua    | CGGGTGATCATTCTGATGGTGTGGATTGTGTCAGGrc |
| Quechua    | CGGGTGATCATTCTGATGGTGTGGATTGTGTCAGGCCTTACCTCCTTCTTGCCCATTCAG |
| Quechua    | CGGGTGATCATTCTGATGGTGTGGATTGTGTCAGGCCTTACCTCCTTCTTGCCCATTCAG |

| Kobilka, 87: | 511 ATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGC |
| Quechua    | ATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGC |
| Quechua    | ATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGC |
| Quechua    | ATGCACTGGTACCGGGCCACCCACCAGGAAGCCATCAACTGCTATGCCAATGAGACCTGC |

| Kobilka, 87: | 571 TGSTGACTCTTCTTCACTCAAGACCACCTATGCGGATTTCTGCATGTTTCTCCAAGTGCTT |
| Quechua    | TGSTGACTCTTCTTCACTCAAGACCACCTATGCGGATTTCTGCATGTTTCTCCAAGTGCTT |
| Quechua    | TGSTGACTCTTCTTCACTCAAGACCACCTATGCGGATTTCTGCATGTTTCTCCAAGTGCTT |
| Quechua    | TGSTGACTCTTCTTCACTCAAGACCACCTATGCGGATTTCTGCATGTTTCTCCAAGTGCTT |

| Kobilka, 87: | 631 CCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCCAGCTC |
| Quechua    | CCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCCAGCTC |
| Quechua    | CCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCCAGCTC |
| Quechua    | CCCCTGGTGATCATGGTCTTCGTCTACTCCAGGGTCTTTCAGGAGGCCAAAAGGCCAGCTC |

| Kobilka, 87: | 691 CAGAAGGTTGGACAATACTGAGGGCCGCCCTCTCCATGCGGAGCCCCCAAAGGCCAGCTC |
| Quechua    | CAGAAGGTTGGACAATACTGAGGGCCGCCCTCTCCATGCGGAGCCCCCAAAGGCCAGCTC |
| Quechua    | CAGAAGGTTGGACAATACTGAGGGCCGCCCTCTCCATGCGGAGCCCCCAAAGGCCAGCTC |
| Quechua    | CAGAAGGTTGGACAATACTGAGGGCCGCCCTCTCCATGCGGAGCCCCCAAAGGCCAGCTC |
Appendix iii b continued.

| Kobilka, 87: | 751 GATGGGCGGACGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAGGACACAAA | 810 |
| Quechua 1: | GATGGGCGGACGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAGGACACAAA |
| Quechua 2: | GATGGGCGGACGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAGGACACAAA |
| Quechua 3: | GATGGGCGGACGGGCATGGACTCCGCAGATCTTCCAAGTTCTGCTTGAGGACACAAA |

| Kobilka, 87: | 811 GCCCTCAAGACGGCATCATCATGCGAGCTTCTGCCTGCCCTCTTC | 870 |
| Quechua 1: | GCCCTCAAGACGGCATCATCATGCGAGCTTCTGCCTGCCCTCTTC |
| Quechua 2: | GCCCTCAAGACGGCATCATCATGCGAGCTTCTGCCTGCCCTCTTC |
| Quechua 3: | GCCCTCAAGACGGCATCATCATGCGAGCTTCTGCCTGCCCTCTTC |

| Kobilka, 87: | 871 ATCGTTAACATTGTGCTGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTC | 930 |
| Quechua 1: | ATCGTTAACATTGTGCTGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTC |
| Quechua 2: | ATCGTTAACATTGTGCTGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTC |
| Quechua 3: | ATCGTTAACATTGTGCTGTGATCCAGGATAACCTCATCCGTAAGGAAGTTTACATCCTC |

| Kobilka, 87: | 931 CTAAATTGGATAGGCTATGTCAATTCTGGTTTCAATCCCCTTATCTACTGCCGGAGCCCA | 990 |
| Quechua 1: | CTAAATTGGATAGGCTATGTCAATTCTGGTTTCAATCCCCTTATCTACTGCCGGAGCCCA |
| Quechua 2: | CTAAATTGGATAGGCTATGTCAATTCTGGTTTCAATCCCCTTATCTACTGCCGGAGCCCA |
| Quechua 3: | CTAAATTGGATAGGCTATGTCAATTCTGGTTTCAATCCCCTTATCTACTGCCGGAGCCCA |

| Kobilka, 87: | 991 GATTTGAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTAT | 1050 |
| Quechua 1: | GATTTGAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTAT |
| Quechua 2: | GATTTGAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTAT |
| Quechua 3: | GATTTGAGGATTGCCTTCCAGGAGCTTCTGTGCCTGCGCAGGTCTTCTTTGAAGGCCTAT |

| Kobilka, 87: | 1051 GGCAATGGCTACTCCAGCAACGGCAACACAGGGGAGCAGAGTGGATATCACGTGGAACAG | 1110 |
| Quechua 1: | GGCAATGGCTACTCCAGCAACGGCAACACAGGGGAGCAGAGTGGATATCACGTGGAACAG |
| Quechua 2: | GGCAATGGCTACTCCAGCAACGGCAACACAGGGGAGCAGAGTGGATATCACGTGGAACAG |
| Quechua 3: | GGCAATGGCTACTCCAGCAACGGCAACACAGGGGAGCAGAGTGGATATCACGTGGAACAG |

| Kobilka, 87: | 1111 GAGAAAGAAAATAAAGACCTGGCTGGAAGACCTCCCCAGGACGAGATTCTTGTTGGCCCAT | 1170 |
| Quechua 1: | GAGAAAGAAAATAAAGACCTGGCTGGAAGACCTCCCCAGGACGAGATTCTTGTTGGCCCAT |
| Quechua 2: | GAGAAAGAAAATAAAGACCTGGCTGGAAGACCTCCCCAGGACGAGATTCTTGTTGGCCCAT |
| Quechua 3: | GAGAAAGAAAATAAAGACCTGGCTGGAAGACCTCCCCAGGACGAGATTCTTGTTGGCCCAT |

| Kobilka, 87: | 1171 CAACGGTACTGGCTTGCTGAGGACACCTCCCCAGGACGAGATTCTTGTTGGCCCAT | 1230 |
| Quechua 1: | CAACGGTACTGGCTTGCTGAGGACACCTCCCCAGGACGAGATTCTTGTTGGCCCAT |
| Quechua 2: | CAACGGTACTGGCTTGCTGAGGACACCTCCCCAGGACGAGATTCTTGTTGGCCCAT |
| Quechua 3: | CAACGGTACTGGCTTGCTGAGGACACCTCCCCAGGACGAGATTCTTGTTGGCCCAT |

| Kobilka, 87: | 1231 TCACTGCTAAGgacagttttttctacttttaaga 1265 |
| Quechua 1: | TCACTGCTAAGgacagttttttctacttttaaga |
| Quechua 2: | TCACTGCTAAGgacagttttttctacttttaaga |
| Quechua 3: | TCACTGCTAAGgacagttttttctacttttaaga |
Appendix iii. DNA sequences (continued)

c). Erythropoietin gene sequenced from three Quechua. Reference sequence (upper) is a region of chromosome 7q22 sequence (GenBank accession number Number AF053356; see Glockner and Tsui, 1998). The T/A<sup>+662</sup> and the T/G<sup>+769</sup> polymorphisms in the three prime untranslated region are underlined. Vertical bars indicate consensus between all four sequences. The coding sequence is numbered and the initiation and termination codons are boldfaced. Quechua sequences are available through GenBank with the following accession numbers AF202306, AF202307, AF202308 (Quechua 1); AF202309, AF202310, AF202311 (Quechua 2); AF202312, AF202313 and AF202314 (Quechua 3).

Exon 1

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>7q22 seq.:</td>
<td>cccggccaggcggcagATGGGGGTGCACGgtgagtactcgcgggctgggcgctcccgcc</td>
<td></td>
</tr>
<tr>
<td>Quechua 1:</td>
<td>cccggccaggcggcagATGGGGGTGCACGgtgagtactcgcgggctgggcgctcccgcc</td>
<td></td>
</tr>
<tr>
<td>Quechua 2:</td>
<td>cccggccaggcggcagATGGGGGTGCACGgtgagtactcgcgggctgggcgctcccgcc</td>
<td></td>
</tr>
<tr>
<td>Quechua 3:</td>
<td>cccggccaggcggcagATGGGGGTGCACGgtgagtactcgcgggctgggcgctcccgcc</td>
<td></td>
</tr>
</tbody>
</table>

Exons 2 and 3

<table>
<thead>
<tr>
<th></th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>7q22 seq.:</td>
<td>cctggctatctgttctagAATGTCCTGCCTGGCTGTGGCTTCTCCTGTCCCTGCTGTCGC</td>
</tr>
<tr>
<td>Quechua 1:</td>
<td>cctggctatctgttctagAATGTCCTGCCTGGCTGTGGCTTCTCCTGTCCCTGCTGTCGC</td>
</tr>
<tr>
<td>Quechua 2:</td>
<td>cctggctatctgttctagAATGTCCTGCCTGGCTGTGGCTTCTCCTGTCCCTGCTGTCGC</td>
</tr>
<tr>
<td>Quechua 3:</td>
<td>cctggctatctgttctagAATGTCCTGCCTGGCTGTGGCTTCTCCTGTCCCTGCTGTCGC</td>
</tr>
</tbody>
</table>

Exon 1

<table>
<thead>
<tr>
<th></th>
<th>159</th>
</tr>
</thead>
<tbody>
<tr>
<td>7q22 seq.:</td>
<td>TGGAGAGGTACCTCTTGGAGGCCAAGGAGGCCGAGAATATCACGgtgagaccccttcccc</td>
</tr>
<tr>
<td>Quechua 1:</td>
<td>TGGAGAGGTACCTCTTGGAGGCCAAGGAGGCCGAGAATATCACGgtgagaccccttcccc</td>
</tr>
<tr>
<td>Quechua 2:</td>
<td>TGGAGAGGTACCTCTTGGAGGCCAAGGAGGCCGAGAATATCACGgtgagaccccttcccc</td>
</tr>
<tr>
<td>Quechua 3:</td>
<td>TGGAGAGGTACCTCTTGGAGGCCAAGGAGGCCGAGAATATCACGgtgagaccccttcccc</td>
</tr>
</tbody>
</table>
Appendix iii c continued.

7q22 seq.: agcacattccacagaactcagctacegggttcagggaactctccccagatccaggaacc
Quechua 1: agcacattccacagaactcagctacegggttcagggaactctccccagatccaggaacc
Quechua 2: agcacattccacagaactcagctacegggttcagggaactctccccagatccaggaacc
Quechua 3: agcacattccacagaactcagctacegggttcagggaactctccccagatccaggaacc

7q22 seq.: tgccacattgggtgggtggaaggctagaatccactgtgcctacatagaaag
Quechua 1: tgccacattgggtgggtggaaggctagaatccactgtgcctacatagaaag
Quechua 2: tgccacattgggtgggtggaaggctagaatccactgtgcctacatagaaag
Quechua 3: tgccacattgggtgggtggaaggctagaatccactgtgcctacatagaaag

7q22 seq.: tctggtggccccaaaccatacctggaaactaggcaaggagcaaagccagcagatcctacg
Quechua 1: tctggtggccccaaaccatacctggaaactaggcaaggagcaaagccagcagatcctacg
Quechua 2: tctggtggccccaaaccatacctggaaactaggcaaggagcaaagccagcagatcctacg
Quechua 3: tctggtggccccaaaccatacctggaaactaggcaaggagcaaagccagcagatcctacg

7q22 seq.: gcctgtgggccagggccagagccttcagggacccttgactccccgggctgtgtgcatttc
Quechua 1: gcctgtgggccagggccagagccttcagggacccttgactccccgggctgtgtgcatttc
Quechua 2: gcctgtgggccagggccagagccttcagggacccttgactccccgggctgtgtgcatttc
Quechua 3: gcctgtgggccagggccagagccttcagggacccttgactccccgggctgtgtgcatttc

7q22 seq.: agACGGGCTGTGCTGAACACTGCAGC
Quechua 1: agACGGGCTGTGCTGAACACTGCAGC
Quechua 2: agACGGGCTGTGCTGAACACTGCAGC
Quechua 3: agACGGGCTGTGCTGAACACTGCAGC

7q22 seq.: TTAATTTCTATGCCTGGAAGAGGATGGAGgtgagttccttt
Quechua 1: TTAATTTCTATGCCTGGAAGAGGATGGAGgtgagttccttt
Quechua 2: TTAATTTCTATGCCTGGAAGAGGATGGAG
Quechua 3: TTAATTTCTATGCCTGGAAGAGGATGGAG

Exons 4 and 5 and 3' untranslated region

7q22 seq.: ggtcagctgactcccagagtccactccctgtagGTCGGGCAGCAGGCCGTAGAAGTCTGG
Quechua 1: ggtcagctgactcccagagtccactccctgtagGTCGGGCAGCAGGCCGTAGAAGTCTGG
Quechua 2: ggtcagctgactcccagagtccactccctgtagGTCGGGCAGCAGGCCGTAGAAGTCTGG
Quechua 3: ggtcagctgactcccagagtccactccctgtagGTCGGGCAGCAGGCCGTAGAAGTCTGG

7q22 seq.: CAGGGCCTGGCCCTGCTGTCGGAAGCTGTCCTGCGGGGCCAGGCCCTG
Quechua 1: CAGGGCCTGGCCCTGCTGTCGGAAGCTGTCCTGCGGGGCCAGGCCCTG
Quechua 2: CAGGGCCTGGCCCTGCTGTCGGAAGCTGTCCTGCGGGGCCAGGCCCTG
Quechua 3: CAGGGCCTGGCCCTGCTGTCGGAAGCTGTCCTGCGGGGCCAGGCCCTG

Exons 4 and 5 and 3' untranslated region
Appendix iii c continued.

7q22 seq.: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 1: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 2: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 3: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC

7q22 seq.: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 1: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 2: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 3: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC

7q22 seq.: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG
Quechua 1: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG
Quechua 2: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG
Quechua 3: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG

426

7q22 seq.: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 1: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 2: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 3: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC

7q22 seq.: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG
Quechua 1: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG
Quechua 2: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG
Quechua 3: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG

427

7q22 seq.: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 1: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 2: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC
Quechua 3: TCCCAGCGGGAGCCCGCTCTGCTCTGCTCTGCCAGC

7q22 seq.: CCCCTCCAGATGCGGCCTCAGCTGCTCCACTCCGAACAATCACTGCTGACACTTTCCGCA
Quechua 1: CCCCTCCAGATGCGGCCTCAGCTGCTCCACTCCGAACAATCACTGCTGACACTTTCCGCA
Quechua 2: CCCCTCCAGATGCGGCCTCAGCTGCTCCACTCCGAACAATCACTGCTGACACTTTCCGCA
Quechua 3: CCCCTCCAGATGCGGCCTCAGCTGCTCCACTCCGAACAATCACTGCTGACACTTTCCGCA

582

7q22 seq.: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG
Quechua 1: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG
Quechua 2: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG
Quechua 3: CTCACCACTCTCTCTGGGCTCTGGGAGCCCAGGTGATGAGAGCAGCCACTTCTGCTTG

7q22 seq.: AACTCTTCCGAGTCTACTCCAATTTCCTCCGGGGAAAGCTGAAGCTGTACACAGGGGAGG
Quechua 1: AACTCTTCCGAGTCTACTCCAATTTCCTCCGGGGAAAGCTGAAGCTGTACACAGGGGAGG
Quechua 2: AACTCTTCCGAGTCTACTCCAATTTCCTCCGGGGAAAGCTGAAGCTGTACACAGGGGAGG
Quechua 3: AACTCTTCCGAGTCTACTCCAATTTCCTCCGGGGAAAGCTGAAGCTGTACACAGGGGAGG

7q22 seq.: CCTGCAGGACAGGGGACAGATGACAGGTGTGTCCACCTGGCATATCCACCACTCCT
Quechua 1: CCTGCAGGACAGGGGACAGATGACAGGTGTGTCCACCTGGCATATCCACCACTCCT
Quechua 2: CCTGCAGGACAGGGGACAGATGACAGGTGTGTCCACCTGGCATATCCACCACTCCT
Quechua 3: CCTGCAGGACAGGGGACAGATGACAGGTGTGTCCACCTGGCATATCCACCACTCCT

582

7q22 seq.: CCTGCAGGACAGGGGACAGATGACAGGTGTGTCCACCTGGCATATCCACCACTCCT
Quechua 1: CCTGCAGGACAGGGGACAGATGACAGGTGTGTCCACCTGGCATATCCACCACTCCT
Quechua 2: CCTGCAGGACAGGGGACAGATGACAGGTGTGTCCACCTGGCATATCCACCACTCCT
Quechua 3: CCTGCAGGACAGGGGACAGATGACAGGTGTGTCCACCTGGCATATCCACCACTCCT

7q22 seq.: caccaacattgttgtgccacacctcccccccggccacactccgtgaacccegtcagagggtcct
Quechua 1: caccaacattgttgtgccacacctcccccccggccacactccgtgaacccegtcagagggtcct
Quechua 2: caccaacattgttgtgccacacctcccccccggccacactccgtgaacccegtcagagggtcct
Quechua 3: caccaacattgttgtgccacacctcccccccggccacactccgtgaacccegtcagagggtcct
Appendix iii c continued.

7q22 seq.: cagctcagcgcagcctgtcccacatggacactccagtgccagcaatgacatctcaggggcc
Quechua 1: cagctcagcgcagcctgtcccacatggacactccagtgccagcaatgacatctcaggggcc
Quechua 2: cagctcagcgcagcctgtcccacatggacactccagtgccagcaatgacatctcaggggcc
Quechua 3: cagctcagcgcagcctgtcccacatggacactccagtgccagcaatgacatctcaggggcc

7q22 seq.: agaggaactgtccagagagcaactctgagatctaaggatgtcacagggccaacttgaggg
Quechua 1: agaggaactgtccagagagcaactctgagatctaaggatgtcacagggccaacttgaggg
Quechua 2: agaggaactgtccagagagcaactctgagatctaaggatgtcacagggccaacttgaggg
Quechua 3: agaggaactgtccagagagcaactctgagatctaaggatgtcacagggccaacttgaggg

7q22 seq.: cccagagcaggaagcattcagagagcagctttaaactcagggacagagccatgctgggaa
Quechua 1: cccagagcaggaagcattcagagagcagctttaaactcagggacagagccatgctgggaa
Quechua 2: cccagagcaggaagcattcagagagcagctttaaactcagggacagagccatgctgggaa
Quechua 3: cccagagcaggaagcattcagagagcagctttaaactcagggacagagccatgctgggaa

7q22 seq.: gacgcctgagctcactcggcaccctgcaaaatttgatgccaggacacgctttggaggcga
Quechua 1: gacgcctgagctcactcggcaccctgcaaaatttgatgccaggacacgctttggaggcga
Quechua 2: gacgcctgagctcactcggcaccctgcaaaatttgatgccaggacacgctttggaggcga
Quechua 3: gacgcctgagctcactcggcaccctgcaaaatttgatgccaggacacgctttggaggcga

7q22 seq.: tttacctgttttgcacctaccatcagggacaggatgacctggagaacttaggtggcaag
Quechua 1: tttacctgttttgcacctaccatcagggacaggatgacctggagaacttaggtggcaag
Quechua 2: tttacctgttttgcacctaccatcagggacaggatgacctggagaacttaggtggcaag
Quechua 3: tttacctgttttgcacctaccatcagggacaggatgacctggagaacttaggtggcaag

7q22 seq.: ctgtgacttctccaggtctcacgggcatgggcactcccttggtggcaagagcccccttga
Quechua 1: ctgtgacttctccaggtctcacgggcatgggcactcccttggtggcaagagcccccttga
Quechua 2: ctgtgacttctccaggtctcacgggcatgggcactcccttggtggcaagagcccccttga
Quechua 3: ctgtgacttctccaggtctcacgggcatgggcactcccttggtggcaagagcccccttga

7q22 seq.: cacacagcctgtctgacctctcgaccctaccgggcctgaggccacaagctctgcctacgc
Quechua 1: cacacagcctgtctgacctctcgaccctaccgggcctgaggccacaagctctgcctacgc
Quechua 2: cacacagcctgtctgacctctcgaccctaccgggcctgaggccacaagctctgcctacgc
Quechua 3: cacacagcctgtctgacctctcgaccctaccgggcctgaggccacaagctctgcctacgc

7q22 seq.: ccaagttttgtgtattcttcaacctcattgacaagaactgaaaccaccaatatgactctt
Quechua 1: ccaagttttgtgtattcttcaacctcattgacaagaactgaaaccaccaatatgactctt
Quechua 2: ccaagttttgtgtattcttcaacctcattgacaagaactgaaaccaccaatatgactctt
Quechua 3: ccaagttttgtgtattcttcaacctcattgacaagaactgaaaccaccaatatgactctt

7q22 seq.: ccagttttttgttgtatcttcaaacctcattgacaagaacttagtttgaacatgagcagacag
Quechua 1: ccagttttttgttgtatcttcaaacctcattgacaagaacttagtttgaacatgagcagacag
Quechua 2: ccagttttttgttgtatcttcaaacctcattgacaagaacttagtttgaacatgagcagacag
Quechua 3: ccagttttttgttgtatcttcaaacctcattgacaagaacttagtttgaacatgagcagacag

123
Appendix iii c continued.

7q22 seq.: tgcagcaggtccaggtccgggaaacgaggggtggagggggctgggccctacgtgctgtct
Quechua 1: tgcagcaggtccaggtccgggaaacgaggggtggagggggctgggccctacgtgctgtct
Quechua 2: tgcagcaggtccaggtccgggaaacgaggggtggagggggctgggccctacgtgctgtct
Quechua 3: tgcagcaggtccaggtccgggaaacgaggggtggagggggctgggccctacgtgctgtct

7q22 seq.: cacacagcctgtctgacctctcgaccctaccgggcctgaggccacaagctctgcctacgc
Quechua 1: cacacagcctgtctgacctctcgaccctaccgggcctgaggccacaagctctgcctacgc
Quechua 2: cacacagcctgtctgacctctcgaccctaccgggcctgaggccacaagctctgcctacgc
Quechua 3: cacacagcctgtctgacctctcgaccctaccgggcctgaggccacaagctctgcctacgc

7q22 seq.: tggtcaataaggtgtctccattcaaggcctcaccgcagtaaggcagtgccaaacc
Quechua 1: tggtcaataaggtgtctccattcaaggcctcaccgcagtaaggcagtgccaaacc
Quechua 2: tggtcaataaggtgtctccattcaaggcctcaccgcagtaaggcagtgccaaacc
Quechua 3: tggtcaataaggtgtctccattcaaggcctcaccgcagtaaggcagtgccaaacc

124
Appendix iii. DNA sequences (continued)
d). Representative Quechua hypoxia inducible factor 1a (HIF-1α) cDNA sequence. Sequences were obtained for three Quechua lymphoblast cell lines (GM11197 - GM11201) and no variants were detected. Reference sequence (upper) is from (Wang et al., 1995) (GenBank accession number U22431). Vertical bars indicate consensus. Coding sequence is in upper case. The initiation codon and termination codons are boldfaced. Quechua sequences are available through GenBank (accession numbers AF207601, AF207602 and AF208487).

Wang, 95: gattcaccATGGAGGGCGCGGCAGCGACAGAAAAGATTACTCTGTGAACTGTC
Quechua: gattcaccATGGAGGGCGCGGCAGCGACAGAAAAGATTACTCTGTGAACTGTC

Wang, 95: GAAAAAGAAAGCTCTCGAGATGCAGCCAGATCTCGGCGAAGTAAAGAATCTGAAGTTTTTT
Quechua: GAAAAAGAAAGCTCTCGAGATGCAGCCAGATCTCGGCGAAGTAAAGAATCTGAAGTTTTTT

Wang, 95: ATGAGCTTGTGCTACATCGAGTGAGTGTGTCGACGTAAGTTTTTT
Quechua: ATGAGCTTGTGCTACATCGAGTGAGTGTGTCGACGTAAGTTTTTT

Wang, 95: GAAAAGAAAAGTCTCGAGATGCAGCCAGATCTCGGCGAAGTAAAGAATCTGAAGTTTTTT
Quechua: GAAAAGAAAAGTCTCGAGATGCAGCCAGATCTCGGCGAAGTAAAGAATCTGAAGTTTTTT

Wang, 95: CTGTGATGAGGCTTACGATCAGCTATTTG
Quechua: CTGTGATGAGGCTTACGATCAGCTATTTG

Wang, 95: TGGATATTGAAGATGACATGAAAGCACAGATGAATTGCT
Quechua: TGGATATTGAAGATGACATGAAAGCACAGATGAATTGCT

Wang, 95: GTTTTGTTATGGGTTCTCACAGATGATGGTGACATGATTTACATTTCTC
Quechua: GTTTTGTTATGGGTTCTCACAGATGATGGTGACATGATTTACATTTCTC

Wang, 95: GTTTTGTTATGGGTTCTCACAGATGATGGTGACATGATTTACATTTCTC
Quechua: GTTTTGTTATGGGTTCTCACAGATGATGGTGACATGATTTACATTTCTC

Wang, 95: CATGTGACCATGAGGAAATGAGAGAAATGCTTACAC
Quechua: CATGTGACCATGAGGAAATGAGAGAAATGCTTACAC

Wang, 95: GTAAAGAACAAAACACACAGCGAAGCTTTTTTCTCAGAATGTAACCTACTGAC
Quechua: GTAAAGAACAAAACACACAGCGAAGCTTTTTTCTCAGAATGTAACCTACTGAC

Wang, 95: TTCACGTATATGATACCAACAGTAACCAACCTCAGTGTGGGTATAAGAAACCACCTATGA
Quechua: TTCACGTATATGATACCAACAGTAACCAACCTCAGTGTGGGTATAAGAAACCACCTATGA

125
Appendix iii d continued.

Wang, 95: CCTGCTTGCGTCTGAGTTTGAGCAACCCATCTTCAACCCCATCAAATATTAATTTCTCTT

Quechua: CCTGCTTGCGTCTGAGTTTGAGCAACCCATCTTCAACCCCATCAAATATTAATTTCTCTT

Wang, 95: ATAGCAAGACTTCTCCAGTTCTGAGCTGGAGATATGAAAATTTCTCTTGTATGAGAA

Quechua: ATAGCAAGACTTCTCCAGTTCTGAGCTGGAGATATGAAAATTTCTCTTGTATGAGAA

Wang, 95: GAATTACCGAATTGATGGGATATGAGCCAGAAGAACTTTTAGGCCGCTCAATT

Quechua: GAATTACCGAATTGATGGGATATGAGCCAGAAGAACTTTTAGGCCGCTCAATT

Wang, 95: ATTATCATGCTTTGGACTCTGATCATCTGACCAAAACTCATCATGATATGTTTACTAAAG

Quechua: ATTATCATGCTTTGGACTCTGATCATCTGACCAAAACTCATCATGATATGTTTACTAAAG

Wang, 95: GACAAGTCACCACAGGACAGTACAGGATGCTTGCCAAAAGAGGTGGATATGTCTGGGTTC

Quechua: GACAAGTCACCACAGGACAGTACAGGATGCTTGCCAAAAGAGGTGGATATGTCTGGGTTC

Wang, 95: AAACTCAAGCAACTGTCATATATAACACCAAGAATTCTCAACCACAGTGCATTGTATGTG

Quechua: AAACTCAAGCAACTGTCATATATAACACCAAGAATTCTCAACCACAGTGCATTGTATGTG

Wang, 95: TGAATTACGTTGTGAGTGGTATTATTCAGC

Quechua: TGAATTACGTTGTGAGTGGTATTATTCAGC

Wang, 95: AATGTGTCCTTAAACCGGTTGAATCTTCAGATATGAAAATGACTCAGCTATTCACCAAAG

Quechua: AATGTGTCCTTAAACCGGTTGAATCTTCAGATATGAAAATGACTCAGCTATTCACCAAAG

Wang, 95: CTTTGCTGGCCCCAGCCGCTGGAGACACAATCATATCTTT

Quechua: CTTTGCTGGCCCCAGCCGCTGGAGACACAATCATATCTTT

Wang, 95: C^GAAACTGATGACCAGCAACTTGAGGAAGTACCATTATATAATGATGTAATGCTCCCCT

Quechua: CAGAAACTGATGACCAGCAACTTGAGGAAGTACCATTATATAATGATGTAATGCTCCCCT

Wang, 95: CACCCAACGAAAAATTACAGAATATATATATTTGGCAATGCTCCTTTATATGCCTTCC

Quechua: CACCCAACGAAAAATTACAGAATATATATTTGGCAATGCTCCTTTATATGCCTTCC

Wang, 95: CGCCAAAGCCACTTTCGGAAGTAGTGCTGACCCTGCACTCAATCAAGAAGTTGCATTAAAAT

Quechua: CGCCAAAGCCACTTTCGGAAGTAGTGCTGACCCTGCACTCAATCAAGAAGTTGCATTAAAAT

Wang, 95: TAGAACCAAATCCAGAAGCACCTCAGGGATTTCTCTTATACCATGCCCCAGATCCGATAG

Quechua: TAGAACCAAATCCAGAAGCACCTCAGGGATTTCTCTTATACCATGCCCCAGATCCGATAG
Appendix iii d continued.

Wang, 95: CACCTAGTCCCTCCGAGAGCACTAGAAGAGTTCACTTCCAGCTTCCAGAGTG
Quechua: CACCTAGTCCCTCCGAGAGCACTAGAAGAGTTCACTTCCAGCTTCCAGAGTG

Wang, 95: AATATTGTTTTTATGTGGATAGTGATATGGTCAATGAATTCAAGT
Quechua: AATATTGTTTTTATGTGGATAGTGATATGGTCAATGAATTCAAGT

Wang, 95: AACTTTTTGCTGAAGACACAGAAGC AAGAACCCATTTTCTACTCAGGACACAGATT
Quechua: AACTTTTTGCTGAAGACACAGAAGC AAGAACCCATTTTCTACTCAGGACACAGATT

Wang, 95: ACTTGGAGATGTTAGCTCCCTATATCCCAATGGATGATGACTTCCAGTTACGTTCCTTCG
Quechua: ACTTGGAGATGTTAGCTCCCTATATCCCAATGGATGATGACTTCCAGTTACGTTCCTTCG

Wang, 95: ATCAGTTGTCACCATTAGAAAGCAGTTCCGCAAGC
Quechua: ATCAGTTGTCACCATTAGAAAGCAGTTCCGCAAGC

Wang, 95: CAGTTACAGTATTCCAGCAGACTCAAATACAAGAACCTACTGCTAATGCCACCACTACCA
Quechua: CAGTTACAGTATTCCAGCAGACTCAAATACAAGAACCTACTGCTAATGCCACCACTACCA

Wang, 95: CTGCCACCACTGATGAATTAAAAAC^GTGACAAAAGACCGTATGGAAGACATTAAAATAT
Quechua: CTGCCACCACTGATGAATTAAAAAC^GTGACAAAAGACCGTATGGAAGACATTAAAATAT

Wang, 95: TGATTGCATCTCCATCTCCTACCCACATACATAAAGAAACTACTAGTGCCACATCATCAC
Quechua: TGATTGCATCTCCATCTCCTACCCACATACATAAAGAAACTACTAGTGCCACATCATCAC

Wang, 95: CATATAGAGATACTCAAAGTCGGACAGCCTCACCAAACAGAGCAGGAAAAGGAGTCATAG
Quechua: CATATAGAGATACTCAAAGTCGGACAGCCTCACCAAACAGAGCAGGAAAAGGAGTCATAG

Wang, 95: AACAGACAGAAAAATCTCATCCAAGAAGCCCTAACGTGTTATCTGTCGCTTTGAGTCAAA
Quechua: AACAGACAGAAAAATCTCATCCAAGAAGCCCTAACGTGTTATCTGTCGCTTTGAGTCAAA

Wang, 95: GAACTACAGTTCCTGAGGAAGAACTAAATCCAAAGATACTAGCTTTGCAGAATGCTCAGA
Quechua: GAACTACAGTTCCTGAGGAAGAACTAAATCCAAAGATACTAGCTTTGCAGAATGCTCAGA

Wang, 95: GAAAGCGAAAAATGGAACATGATGGTTCACTTTTTCAAGCAGTAGGAATTG
Quechua: GAAAGCGAAAAATGGAACATGATGGTTCACTTTTTCAAGCAGTAGGAATTG

Wang, 95: TACAGCAGCCAGACGATCATGCAGCTACTACATCACTTTCTTGGAAACGTGTAA
Quechua: TACAGCAGCCAGACGATCATGCAGCTACTACATCACTTTCTTGGAAACGTGTAA

Wang, 95: GCAAATCTAGTGAACAGAATGGAATGGAGCAAAAGACAATTATTTTAATACCCTC
Quechua: GCAAATCTAGTGAACAGAATGGAATGGAGCAAAAGACAATTATTTTAATACCCTC

Wang, 95: GAAAGCGAAAAATGGAACATGATGGTTCACTTTTTCAAGCAGTAGGAATTG
Quechua: GAAAGCGAAAAATGGAACATGATGGTTCACTTTTTCAAGCAGTAGGAATTG

Wang, 95: TACAGCAGCCAGACGATCATGCAGCTACTACATCACTTTCTTGGAAACGTGTAA
Quechua: TACAGCAGCCAGACGATCATGCAGCTACTACATCACTTTCTTGGAAACGTGTAA

Wang, 95: GCAAATCTAGTGAACAGAATGGAATGGAGCAAAAGACAATTATTTTAATACCCTC
Quechua: GCAAATCTAGTGAACAGAATGGAATGGAGCAAAAGACAATTATTTTAATACCCTC

127
Appendix iii d continued.

Wang, 95:  TAGCATGTAGACTGCTGGGGCAATCAATGGATGAAAGTGGATTACCACA
Quechua: TAGCATGTAGACTGCTGGGGCAATCAATGGATGAAAGTGGATTACCAC

Quechua: TAGCATGTAGACTGCTGGGGCAATCAATGGATGAAAGTGGATTACC^CAGCTGACCAGTT

Wang, 95:  ATGATTGTGAAGTTAATGCTCCTATAGAAGGC^  
Quechua: ATGATTGTGAAGTTAATGCTCCTATACAAGGCAGCAGAAACCTACTCAAGGOTGAAGAAT

Wang, 95:  TACTCAGAGCTTTGGATCAAGTTAACTGAgctttttcttaatttcatcctcc
Quechua: TACTCAGAGCTTTGGATCAAGTTAACTGAgctttttcttaatttcatcctcc