AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 95:271–276 (1994)

Major Gene for Percent of Oxygen Saturation of Arterial Hemoglobin in Tibetan Highlanders

CYNTHIA M. BEALL, JOHN BLANGERO,
SARAH WILLIAMS-BLANGERO, AND MELVYN C. GOLDSTEIN
Department of Anthropology, Case Western Reserve University, Cleveland, Ohio 44106-7125 (C.M.B., M.C.G.); Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas 78228-0147 (J.B., S.W.-B.)

KEY WORDS Hypoxia, Adaptation, Statistical genetics, Tibet

ABSTRACT This report employs a statistical genetic approach to analyze quantitative oxygen transport variables in a high-altitude (4,850–5,450 m) native Tibetan population and demonstrates the presence of a major gene influencing % O_2 saturation of arterial hemoglobin. This result suggests the hypothesis that individuals with the dominant allele for higher % O_2 saturation have a selective advantage at high altitude. Studies of the biologically distinctive Himalayan and Andean populations have greatly influenced thinking about ongoing human evolution and adaptation; this is the first statistical evidence for a major gene enhancing oxygen transport in a high-altitude native population.

Hypoxic stress is a constant and unavoidable feature of life at high altitudes requiring adaptations in the oxygen transport system for survival. Evidence accumulated over the past century attests to the biological distinctiveness of the high-altitude human populations that have lived for millennia on the Tibetan and Andean plateaus. In particular, native high-altitude populations have more effective oxygen transport systems relative to newcomers (Ward et al., 1989). While these observations support the hypothesis that the populations have been subject to natural selection for enhanced oxygen transport (Baker, 1976), the search for genes associated with high-altitude adaptation has been unsuccessful (Chakraborty et al., 1983). Here, we employ a statistical genetic approach to show that genetic factors are important determinants of quantitative oxygen transport variables in a high-altitude Tibetan population and to demonstrate the presence of a major gene influencing % O_2 saturation of arterial hemoglobin.

MATERIALS AND METHODS

Population and traits

Reasoning that a population under severe hypoxic stress would likely provide the clearest evidence for genes enhancing oxygen transport, we studied a population of pastoral nomads living permanently at the extremely high altitudes of 4,850–5,450 m (16,000–18,000′) in Phala, Tibet Autonomous Region of China, where the inspired pO_2 is roughly 45% lower than sea level. Three quantitative measurements of the oxygen transport system (forced vital capacity (FVC), hemoglobin concentration (Hb), and % O_2 saturation of arterial hemoglobin [SaO_2]) were obtained from 201 healthy nonpregnant people using techniques described elsewhere (Beall and Goldstein, 1990).

Received October 20, 1992; accepted May 18, 1994.

Address reprint requests to Cynthia M. Beall, Department of Anthropology, Case Western Reserve University, Cleveland, OH 44106-7125.

TABLE 1. Distribution of subjects by pedigree

Pedigree size	Number of pedigrees	Number of subjects
1	9	9
2	1	2
3	1	3
4	2	8
5	1	5
7	2	14
8	1	8
10	1	10
67	1	67
71	1	71
Total	20	197

The baseline year of data collection is 1987 for SaO_2 and FVC and 1988 for Hb. Data from other years, for individuals either pregnant or not present, that were collected with the same calibrated equipment are included in the analysis. SaO_2 was collected from 190 nomads 5 years of age and older in 1987 and 7 in 1988 (altogether 86 males, 111 females). Hb measurements were obtained from 127 nomads 15 years of age and older in 1988, 8 in 1986, and 1 in 1990 (altogether 65 males, 71 females). FVC assessments were obtained for 142 nomads 10 years of age and older in 1987, 8 in 1986, and 9 in 1988 (altogether 73 males, 86 females).

Pedigree structure

Kinship relationships among nomads were determined from extensive interviews. Pedigree accuracy was checked by follow-up questioning and by computer evaluation using the program PEDSYS (Dyke, 1989). Based on this genealogical information, the sampled individuals were placed into 20 distinct pedigrees. Table 1 provides a breakdown of sample size by pedigree for the 197 individuals with SaO_2 level data. Table 1 shows that 188 individuals were assigned to 11 pedigrees with an additional nine subjects representing independent individuals. Two large complex pedigrees (of sizes 71 and 67) provide the bulk of sampled individuals. These complex pedigrees provided information on many classes of relatives and were particularly informative for statistical genetic analysis. Of the 11 pedigrees with more than one member, six contained members spanning three generations while five spanned two generations. There were 93 father-offspring pairs, 104 mother-offspring pairs, 75 father-mother-offspring triplets, and 168 sib pairs. The 32 complete nuclear families ranged from 3–11 members.

Analytical Methods

Our general strategy for statistical genetic analysis is to proceed from simple models (classical quantitative genetic models) which assume polygenic inheritance to more complex models (mixed major-locus models) that allow for the effects of specific major loci in addition to polygenic effects.

Quantitative genetic analyses

We tested a series of hypotheses regarding sources of variation in SaO_2, Hb, and FVC measures using maximum likelihood variance decomposition methods (Hopper and Mathews, 1982; Blangero, 1993) available in the computer program PAP (Hasstedt, 1989) as modified by Blangero (1993). These analyses provide information regarding the relative importance of genetic and random environmental effects.

The model that we considered is based on the following linear function for the vector of phenotypes in a pedigree of size n:

$$y = \mu 1_n + (X - 1_n s')\beta + g + e \quad (1)$$

where y is the $n \times 1$ vector of phenotypes, μ is the grand mean of the trait in the population, X is an $n \times k$ matrix containing k covariates such as age, 1_n is a vector of n ones, s represents a vector of baseline covariates (e.g., 0 for qualitative covariates and x for continuous covariates), β is a $k \times 1$ vector of regression coefficients, g is the vector of additive genetic values, and e is a vector of random environmental deviations. Equation 1 shows that covariate effects can be simultaneously evaluated using this approach.

Given this model, the expected variance/covariance matrix for y is written

$$\begin{aligned} \mathrm{Var}(y) &= \Omega \\ &= 2\,\Phi\,\sigma_g^2 + I_n\,\sigma_e^2 \end{aligned} \quad (2)$$

where Φ is the $n \times n$ matrix of kinship coefficients and I_n is an identity matrix of order n. The variance terms in equation 2 include

the additive genetic variance (σ_g^2) and the random environmental variance (σ_e^2). For the current analyses, we have chosen to reparameterize the problem by estimating the phenotypic standard deviation [$\sigma_p = \sqrt{(\sigma_g^2 + \sigma_e^2)}$] and the additive genetic heritability ($h^2 = \sigma_g^2/\sigma_p^2$) which gives the proportion of interindividual phenotypic variation due to additive genetic effects.

Assuming multivariate normality of y, the likelihood of the pedigree is easily calculated and optimization methods can be used for parameter estimation. Subsequent hypothesis testing is performed using likelihood ratio tests. For example, to test the hypothesis that there is no additive genetic variance for a given trait, we compared the likelihood of the model in which h^2 is estimated against the likelihood of the nested submodel in which h^2 is forced to be 0. Twice the difference between these two likelihoods yields a likelihood ratio test statistic that is distributed similarly to a χ^2 variate except that the P value is halved (Hopper and Mathews, 1982).

Complex segregation analysis of SaO_2 levels

Complex segregation analysis (Elston and Stewart, 1971) was performed on the SaO_2 level data using the computer program PAP (Hasstedt, 1989) as modified by Blangero and Konigsberg (1991). Due to the more restricted samples sizes, we chose not to perform segregation analyses of the FVC and Hb data. For SaO_2, we compared the likelihoods of a set of restricted models representing various transmission hypotheses with an unrestricted general model allowing a mixture of up to three normal phenotypic distributions. These three distributions can be related to unobservable genotypes or, more generally, ousiotypes (Cannings et al., 1978) with or without genetic inheritance. Ousiotypes (genotypes) are the product of two discrete factors (alleles), A or a. The three resulting ousiotypes can be denoted as AA, Aa, and aa. To simplify computations, the frequencies of the ousiotypes are usually assumed to follow Hardy-Weinberg proportions (p^2:$2p(1-p)$:$(1-p)^2$).

Under a mixed model (which includes both a major factor and a residual polygenic component), the phenotype of the j-th individual with ousiotype i is

$$(y_j \mid o_j = i) = \mu_i + g_j + e_j \qquad (3)$$

with

$$\mathrm{Var}(y_j \mid o_j = i) = \sigma_g^2 + \sigma_e^2 \qquad (4)$$

where o is the ousiotype, and μ_i ($i = AA, Aa, aa$) is the mean associated with the i-th ousiotype. The genetic component (σ_g^2) of the conditional variance given in equation 4 represents the residual additive genetic variance. The unconditional variance of y has an additional variance component attributable to the effect of the major factor:

$$\mathrm{Var}(y) = \sigma_o^2 + \sigma_g^2 + \sigma_e^2 \qquad (5)$$

where σ_o^2 is the variance due to the major factor (locus).

Three means (μ_{AA}, μ_{Aa}, μ_{aa}) can be estimated corresponding to each ousiotype. A common standard deviation [$\sigma = \sqrt{(\sigma_g^2 + \sigma_e^2)}$] for the residual phenotypic distributions is assumed. Using the mixed model as implemented in PAP, residual nonindependence among relatives due to biological kinship is allowed for by including a polygenic heritability parameter (h^2) which refers to the proportion of phenotypic variance due to additive genetic variance within each ousiotype's phenotypic distribution. The proportion of phenotypic variance attributable to random environmental variation within a phenotypic distribution is given by $1 - h^2$. Three arbitrary transmission parameters (τ_{AA}, τ_{Aa}, τ_{aa}), denoting the probability that an individual of a given ousiotype transmits factor A to an offspring, can be estimated in the most general model.

Four classes of restricted models were tested against the most general model using the unified approach of Lalouel et al. (1983). The simplest model (the sporadic model) included only random environmental effects. In this type of model, all individual trait values are independent of one another. When multiple distributions are considered, a closely related class of model (the finite mixture model) is obtained by forcing the admixture parameter (p_A) to equal the transmis-

TABLE 2. The proportion of phenotypic variation among Phala nomads due to genetic effects on oxygen transport measures: Maximum likelihood estimates of additive genetic heritability, h^2, their standard errors, likelihood ratio test statistics (Λ), and associated P values

Trait	h^2	Λ	P
SaO_2	0.41 ± 0.14	16.56	< 0.001
Hb	0.66 ± 0.16	18.30	< 0.001
FVC	0.27 ± 0.15	5.38	0.011

sion parameters ($p_A = \tau_{AA} = \tau_{Aa} = \tau_{aa}$). The finite mixture model assumes random environmental effects for major factors but does permit residual polygenic inheritance. The classical polygenic model is identical to the one described by equations 1 and 2 and allows only for polygenic inheritance. The mixed Mendelian model incorporates transmission probabilities fixed at their Mendelian expectations ($\tau_{AA} = 1$, $\tau_{Aa} = 1/2$, $\tau_{aa} = 0$) and additionally allows for a residual polygenic background. For each model, the required parameters were estimated by numerical maximization of the likelihood of the data given the assumed transmission model.

Each restricted model was compared with the unrestricted general model using likelihood ratio statistics obtained as twice the difference between the $\log_e$ likelihoods of the unrestricted and restricted models. These tests statistics are asymptotically distributed as chi-square variates with degrees of freedom equal to the difference in the number of parameters between the two competing models. The best model (of those considered) is the one requiring the fewest estimated parameters without being significantly worse than the most general model.

RESULTS AND DISCUSSION

Table 2 shows the results of the quantitative genetic analyses for all three measures of oxygen transport. All analyses allowed for simultaneous estimation of covariate effects including sex sex-specific age and age^2 effects (parameter estimates not shown). All three traits exhibited significant additive genetic components as assessed by the likelihood ratio test. Hemoglobin level has the highest heritability ($h^2 = 0.66$), followed by SaO_2 level ($h^2 = 0.41$) and FVC ($h^2 = 0.27$). These results indicate that there are substantial genetic determinants influencing oxygen transport. Additionally, these findings show that there is considerable power in our pedigree structure to detect genetic effects even with our relatively modest sample sizes.

Once the presence of a genetic component to SaO_2 was established, complex segregation analysis was used to determine if a segregating autosomal locus could explain all or part of the genetic variation in this trait. A major gene influencing SaO_2 was detected. Competing transmission hypotheses were evaluated using the unified model of Lalouel et al. (1983) after correcting for age and sex effects. All resulting means were standardized to reflect those expected in 23-year-old males. The model comparisons revealed that only two component distributions were required to adequately account for the observed variation in SaO_2, resulting in consideration of genetic models with two phenotypic distributions attributable to one locus with two alleles. In these models the heterozygote mean μ_{Aa} was constrained to equal that of the high homozygote, μ_{aa} (i.e., a model in which allele *a* is dominant over allele *A*).

The results of the complex segregation analysis are summarized in Table 3. The Mendelian model provided the best fit to the general model with all other models rejected as being significantly different from the general model. Therefore, the presence of a major gene is the best fitting, most parsimonious explanation for the familial patterning of quantitative SaO_2 levels. The estimated allele frequency for the dominant allele *a* for higher SaO_2 is 0.446 ± 0.095. As noted, the Mendelian model allows for two phenotypic distributions. Individuals in the lower one are comprised of *AA* homozygotes and have a mean SaO_2 of 78.1 ± 1.0%, while the upper distribution of *Aa* heterozygotes and *aa* homozygotes contains individuals with a mean SaO_2 of 84.0 ± 0.5%. There was also evidence for a significant residual polygenic effect ($h^2 = 0.253$, $P = 0.045$). The common within-genotype phenotypic standard deviation was estimated at 2.9 ± 0.2%.

Figure 1 presents a histogram of the observed data and the theoretical genotypic distributions obtained from the parameters

TABLE 3. Evaluation of hypotheses about the mode of transmission of the SaO_2 gene[1]

Parameter	General	Finite mixture	Mendelian	Polygenic	Sporadic
p_A	0.301	0.418	0.446	(1)[2]	(1)
τ_{AA}	0.842	0.418	(1)	(1)	(1)
τ_{Aa}	0.801	0.418	(1/2)	(1/2)	(1/2)
τ_{aa}	0.000	0.418	(0)	(0)	(0)
μ_{AA}	78.07	78.64	78.07	82.76	82.93
$\mu_{Aa} = \mu_{aa}$	83.98	83.62	83.96	82.76	82.93
σ	2.939	3.241	2.955	3.746	3.780
h^2	0.209	0.545	0.253	0.409	(0)
χ^2[3]	—	9.08	3.92	11.44	28.00
df[4]	—	~2	~2	~4	~5
P	—	0.011	0.141	0.022	< 0.001

[1] Results of complex segregation analysis of SaO_2 in 197 Tibetan nomads: maximum likelihood estimates and χ^2 statistics.
[2] Values in parentheses represent fixed parameters.
[3] χ^2 compares a given model with the general model.
[4] The number of degrees of freedom is approximate because τ_{aa} was estimated on the lower boundary (0) of the parameter space.

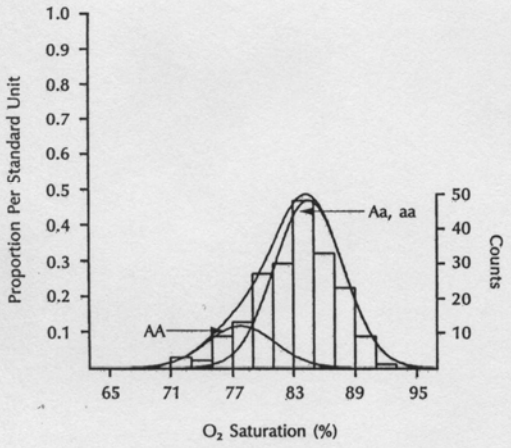

Fig. 1. Observed SaO_2 distribution and theoretical genotypic distributions.

of the best fitting model. Using variance decomposition formulas (Blangero and Konigsberg, 1991), we calculated that the major locus accounts for 39% of the total phenotypic variation in SaO_2 in this population, while polygenes account for an additional 15% of the variation.

To our knowledge, this analysis provides the first statistical evidence for a major gene influencing normal quantitative variation in an oxygen transport variable, as well as the first evidence for a gene enhancing oxygen transport in a high-altitude native population. These results suggest the hypothesis that individuals with the dominant allele *a* are at a selective advantage in the hypoxic high-altitude environment. Investigation of other Himalayan and Andean populations using this approach will enable direct comparison of allele frequencies and mean effects and test the hypothesis that they are genetically adapted to their stressful high-altitude environments via the same mechanism.

ACKNOWLEDGMENTS

We are grateful to the Phala nomads for their hospitality and cooperation, to our Tibetan assistants for their excellent fieldwork, and to the Tibetan Academy of Social Sciences for permission to conduct this research. Kingman P. Strohl, M.D., and Michael J. Decker of the Case Western Reserve University School of Medicine helped to initiate the field study of % O_2 saturation of arterial hemoglobin, and Criticare Systems Inc. (Milwaukee, WI) loaned a backup pulse oximeter. The research was supported by grants from the National Academy of Sciences, National Program for Advanced Study and Research in China, National Geographic Society, and the National Science Foundation.

LITERATURE CITED

Baker PT (1976) Evolution of a project: Theory, method and sampling. In PT Baker and MA Little (eds.): Man in the Andes. A Multidisciplinary Study of High-Altitude Quechua. Stroudsburg, PA: Dowden, Hutchinson, and Ross, Inc., pp. 1–20.

Beall CM, and Goldstein MC (1990) Hemoglobin concentration, percent oxygen saturation, and arterial oxygen content of Tibetan Nomads at 4,850–5,450 m. In

JR Sutton, G Coates, and JE Remmers (eds.): Hypoxia: The Adaptations. Toronto: B.C. Decker, Inc., pp. 59–65.

Blangero J (1993) Statistical genetic approaches to human adaptability. Hum. Biol. *65:*941–966.

Blangero J, and Konigsberg LW (1991) Multivariate segregation analysis using the mixed model. Genet. Epidemiol. *8:*299–316.

Cannings C, Thompson EA, and Skolnick MH (1978) Probability functions on complex pedigrees. Adv. Appl. Prob. *10:*26–61.

Chakraborty R, Clench J, Ferrell RE, Barton SA, and Schull WJ (1983) Genetic components of variation of red cell glycolytic intermediates at two altitudes among the South American Aymara. Ann. Hum. Biol. *10:*174–184.

Dyke B (1989) PEDSYS. A Pedigree Data Management System. Users Manual. PGL Technical Report No. 2. San Antonio, TX: Southwest Foundation for Biomedical Research.

Elston RC, and Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum. Hered. *21:* 523–542.

Hasstedt SJ (1989) Pedigree Analysis Package, V3.0. Salt Lake City, UT: Department of Human Genetics.

Hopper JL, and Mathews JD (1982) Extensions to multivariate normal models for pedigree analysis. Ann. Hum. Genet. *46:*373–383.

Lalouel JM, Rao DC, Morton NE, and Elston RC (1983) A unified model for complex segregation analysis. Am. J. Hum. Genet. *35:*816–826.

Ward MP, Milledge JS, and West JB (1989) High Altitude Medicine and Physiology. Philadelphia: University of Pennsylvania Press.