

# Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits

Harry Campbell<sup>1,\*†</sup>, Andrew D. Carothers<sup>2,†</sup>, Igor Rudan<sup>1,3,†</sup>, Caroline Hayward<sup>2</sup>, Zrinka Biloglav<sup>3</sup>, Lovorka Barac<sup>4</sup>, Marijana Pericic<sup>4</sup>, Branka Janicijevic<sup>4</sup>, Nina Smolej-Narancic<sup>4</sup>, Ozren Polasek<sup>1,3</sup>, Ivana Kolcic<sup>3</sup>, James L. Weber<sup>5</sup>, Nicholas D. Hastie<sup>2</sup>, Pavao Rudan<sup>4</sup> and Alan F. Wright<sup>2</sup>

<sup>1</sup>Department of Public Health Sciences, University of Edinburgh, Edinburgh, UK, <sup>2</sup>MRC Human Genetics Unit, Western General Hospital, Edinburgh, UK, <sup>3</sup>School of Public Health, University Medical School, Zagreb, Croatia, <sup>4</sup>Institute of Anthropological Research, Zagreb, Croatia and <sup>5</sup>Molecular Diagnostic Genotyping Laboratory, Marshfield Clinic Research Foundation, Marshfield, USA

Received August 20, 2006; Revised and Accepted December 17, 2006

The dramatic changes in human population structure over the last 200 years have resulted in significant levels of outbreeding, which, in turn, is predicted to lead to increased levels of individual genetic diversity (genome-wide heterozygosity,  $h$ ). To investigate possible effects of these large demographic changes on global health, we studied the effect of  $h$ , measured as relative heterozygosity,  $h_R$ , on 15 disease-related traits in four groups of individuals with widely differing ancestral histories (ranging from outbred to inbred) from the Dalmatian islands in Croatia. Higher levels of  $h_R$ , estimated using 1184 STR/indel markers, were found in the outbred group ( $P < 0.0001$ ) and were associated with lower blood pressure (BP) and total/LDL cholesterol ( $P = 0.01$  and  $0.01$ , respectively) after controlling for other factors, with BP showing a strong sex effect (males  $P > 0.5$  and females  $P = 0.002$ ). These findings, if replicated, suggest that  $h_R$  be considered as a genetic risk factor in genetic epidemiological studies on common disease traits. They are consistent with the well-known effects of heterosis (hybrid vigour) described when outcrossing animals and plants. Outbreeding resulting from urbanization and migration from traditional population subgroups may be leading to increasing  $h_R$  and may have beneficial effects on a range of traits associated with human health and disease. Other traits, such as age at menarche, IQ and lifespan, which have been changing during the decades of urbanization, may also have been influenced by demographic factors.

## INTRODUCTION

The demographic structure of human populations changed dramatically in the last 200 years. Over this period, total human population rapidly expanded (from 1.5 to over 6.2 billion) and the percentage of the human population inhabiting cities has increased from 2 to nearly 50% (1). This process of urban–rural migration, whereby people have migrated from villages of origin which are more genetically uniform to form larger settlements, has resulted in the breakdown of population barriers, leading to marriage outside the traditional group and mixing of genetically different populations (2,3).

This has also been reflected in the decreasing extent of population substructure over the period 1900 to present day (4). The health effects of this change in genetic structure of human population have not been investigated to date.

An immediate predicted effect of urbanization and admixture is an increase in average individual genome-wide heterozygosity,  $h$ , and this has been reflected in reports of higher heterozygosity in younger populations (5). Admixture of individuals from different genetic populations can be expected to increase  $h$  in the offspring, just as inbreeding within isolated communities can be expected to decrease  $h$ . Quantitative genetics theory predicts that this should influence the variation in

\*To whom correspondence should be addressed at: Department of Public Health Sciences, University of Edinburgh, Teviot Place, Edinburgh, Scotland EH8 9AG, UK. Tel: +44 1316506984; Fax: +44 1316506909; Email: harry.campbell@ed.ac.uk

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

biological traits that show significant dominance variance (6,7). These include traits that represent risk factors underlying conditions accounting for a major proportion of human disease burden such as systolic blood pressure (SBP), total cholesterol and LDL cholesterol (8,9). These traits were responsible for ~20% of the total burden of disease in industrialized countries in 2000 (hypertension 11% and hypercholesterolaemia 8%) (10). In contrast, traits that do not show dominance variance such as height, body mass index and HDL cholesterol would not be expected to be affected by increasing  $h$  (8,9).

Three of our recent findings form a scientific basis for this study. First, we have reported important inbreeding effects on a range of physiological traits and diseases in humans, which are consistent with this hypothesis (11,12). Secondly, we have proposed a model of highly polygenic genetic architecture of complex traits and diseases of post-reproductive age, which could not have been mainly shaped through pre-reproductive selective forces (13). Thirdly, we have shown that at a population-based level, high-density SNP marker scans can be useful for investigating the effects of homozygosity levels on quantitative traits (QTs) that are sensitive to inbreeding (14). In this study, we extend our observations of inbreeding effects by estimating the effect of relative heterozygosity,  $h_R$ , on several biomedically relevant human QTs: SBP, diastolic blood pressure (DBP), total cholesterol, HDL and LDL cholesterol, triglycerides, glucose, height, weight and body mass index.

## RESULTS

### Genotyping results

A total of 385 samples and controls and 1240 polymorphic markers were analysed, giving a total of about 477 000 genotypes. We typed control DNA samples 'blind' as a means of estimating typing error and found the averaged genotype error rate to be 0.69%. The distribution of individuals by percentage of markers successfully completed throughout the sample was 97–100% completed –27.5% of sample; 94–96% completed –21.5% sample; 90–93% completed –21.0% sample and <90% typed –29.7% of the sample. In total, 1188 markers were assessed for Hardy–Weinberg (HW) equilibrium with 198 showing some evidence of departure from HW (at  $P < 0.05$  level).

### Variation in $h_R$ across sample groups

The mean value of  $h_R$  ranged from  $-0.013 \pm 0.004$  (SEM) in the inbred group to  $+0.013 \pm 0.004$  (SEM) in the outbred group; the difference between these groups being large and highly significant ( $P < 0.0001$ ) (Tables 1 and 2). Thus, recent inbreeding and outbreeding were associated with a change in  $h_R$  in opposite directions, as predicted, and of approximately equal magnitude from the value of  $h_R$  in the groups of individuals with no recent history of inbreeding or outbreeding. The mean level of relative heterozygosity varied substantially across the groups, the mean value being almost 3% higher in the 'inbred' than 'outbred' groups

**Table 1.** Comparison of relative heterozygosity,  $h_R$ , across sample groups

Description	<i>N</i>	Mean	SE (mean)
Inbred	73	-0.0126	0.0039
Endogamous	48	-0.0002	0.0048
Exogamous	48	+0.0008	0.0048
Outbred	64	+0.0132	0.0042

**Table 2.** Differences in relative heterozygosity across sample groups

Comparison	Difference	SE (difference)	<i>T</i> -value	<i>P</i> -value
Inbred–outbred	-0.0258	0.0057	-4.52	<0.0001
Inbred–endogamous	-0.0124	0.0062	-2.01	0.046
Endogamous–outbred	-0.0134	0.0064	-2.10	0.036

Investigation of the range of heterozygosity,  $h$ , across groups by estimating relative heterozygosity ( $h_R = \text{excess heterozygosity/expected heterozygosity}$ ) for each group as a means of controlling for differing allele frequencies across different communities.

Group definitions:

- Inbred: evidence of recent inbreeding from genealogical data.
- Endogamous: all four of participant's grandparents born in village of residence; no evidence of recent inbreeding from genealogy.
- Exogamous: two to three of participant's grandparents born in village of residence, no evidence of recent inbreeding from genealogy.
- Outbred: zero to one of participant's grandparents born in village of residence and no two born in same settlement; includes some from outside Dalmatian island populations.

(Table 2: significance of difference in  $h_R$  between inbred and outbred groups  $P < 0.0001$ ).

### Effect of level of heterozygosity on QT values

The relationship between  $h_R$  and QT values was explored by multiple regression analyses, with each QT as the dependent variable. A total of 15 QT variables were analysed as outcome variables. Table 3 summarizes the estimated coefficients and their standard errors in these models, for the 231 individuals on whom full information was available. Five of the 15 QTs were significant at the  $P < 0.05$  level. For SBP, the main predictors were age ( $P < 0.0001$ ), sex ( $P = 0.092$ ), ln weight ( $P = 0.022$ ) and  $h$  ( $P = 0.009$ ). The estimated effect of  $h_R$  was an increase of 1.09 mm (SE 0.42) in SBP per 0.01 increase in  $h_R$ . For DBP, the predictors were age ( $P < 0.0001$ ), sex ( $P = 0.52$ ), log weight ( $P = 0.003$ ), socio-economic status ( $P = 0.035$ ) and  $h$  ( $P = 0.011$ ). The estimated effect of  $h_R$  was an increase of 0.52 mm (SE 0.20) in DBP per 0.01 increase in  $h_R$ .

After adjusting for treatments effects (see Materials and Methods), the estimated effect on mean SBP of reducing  $h_R$  by 0.01 increased from 1.09 to 1.27 mm. A similar result held for mean DBP, suggesting that the effect of ignoring medication is to underestimate the effect of  $h_R$ . Regression dilution resulting from imprecision in  $h_R$  estimates is predicted to result in an underestimation of the regression coefficients for BP of ~47% (see Materials and Methods).

Three other QTs—(log)total cholesterol, (log)LDL cholesterol and forced expiratory flow (FEF<sub>25</sub>)—showed a significant relationship with  $h_R$  (Table 3). The estimated effects of

**Table 3.** Effect of relative heterozygosity ( $h_R$ ) on QT values<sup>a</sup>

QT	N	Multiple regression model		
		Coefficient	Standard error	P-value
SBP	223	-102.8	42.0	<b>0.015</b>
DBP	223	-47.7	20.5	<b>0.021</b>
Log(total cholesterol)	200	-1.083	0.439	<b>0.014</b>
Log(LDL cholesterol)	201	-1.539	0.597	<b>0.011</b>
Log(HDL cholesterol)	200	+0.172	0.368	0.640
Log(triglyceride)	201	-0.461	0.914	0.615
Log{log(glucose)}	201	+0.241	0.216	0.266
Log(creatinine) <sup>b</sup>	200	-0.084	0.328	0.799
Urate	201	-147.9	136.7	0.281
Forced vital capacity	201	-2.381	1.250	0.058
Forced expiratory flow (FEF <sub>25</sub> ) <sup>b</sup>	200	-3.174	1.366	<b>0.021</b>
Forced expiratory flow (FEF <sub>50</sub> )	201	-1.433	2.582	0.580
Forced expiratory flow (FEV <sub>1</sub> )	201	-1.458	1.065	0.173
Peak expiratory flow	201	+4.257	3.489	0.224
Log(body mass index) <sup>c</sup>	231	+0.068	0.279	0.807

<sup>a</sup>Predictor variables in the model were age, sex, ln(height), ln(weight), education level, socio-economic status, smoking and alcohol indexes and individual  $h_R$  as defined in the text. Variables not significant at  $P = 0.05$  were individually dropped until no further simplification was possible.

<sup>b</sup>One outlier removed.

<sup>c</sup>ln(height) and ln(weight) not included as predictors for this variable.

$h_R$  on the QTs of greatest public health importance were large. Extrapolating directly from study data, the effect is equivalent to a rise of 6.8 mmHg (SBP), 3.3 mmHg (DBP), 6.8% (total cholesterol) and 9.6% (LDL cholesterol), with the mean fall in heterozygosity expected in the offspring of a first cousin consanguineous marriage. These effects of  $h_R$  are underestimates because of regression dilution bias and, for BP, treatment effects.

**Other outcome variables.** Models were explored with height (ln height) as the outcome variable (because height has not shown dominance variance effects in previous studies) and age, sex, years of schooling, socio-economic status score and  $h_R$  as predictors. The only significant predictors were age ( $P < 0.0001$ ) and sex ( $P < 0.0001$ ). In particular,  $h_R$  was not significant ( $P = 0.50$ ). Similarly, log(BMI) as outcome variable showed no significant effect of  $h_R$ .

### Investigation of gender effects

The relationship between  $h_R$  and SBP and DBP (but none of the other traits) showed a strong interaction with sex (Table 4), with the effect being essentially confined to females (sex  $\times h_R$  highly significant at  $P < 0.001$ ). The effects of increasing  $h_R$  on males (for both SBP and DBP) were small, positive and non-significant, whereas those on females were large, negative and highly significant. None of the remaining QTs showed any effect of sex  $\times h_R$  with the exception of FVC, where it was of marginal significance ( $P = 0.05$ ). Supplementary Material, Figure S3(A–D) shows plots of residuals of SBP and DBP for males and females, after fitting age against  $h_R$ .

**Table 4.** Sex-specific effects of  $h_R$  on BP

	Coefficient	SE	P-value
SBP, males	+2.5	56.3	0.96
SBP, females	-188.4	60.4	0.002
DBP, males	+8.4	29.8	0.78
DBP, females	-89.2	27.7	0.002

Multiple regression model as in Table 3, but with sexes analysed separately. None of the remaining QTs showed significant interactions with sex. The unit of the coefficients is in mmHg per 0.01 increase in  $h_R$ .

## DISCUSSION

### Study hypothesis

Impairment of function due to homozygosity of recessive alleles has been reported through inbreeding effects on a wide range of traits, suggesting a large number of deleterious alleles in the genome. As most identified genetic variants causing complex disease in humans are partially recessive (15,16), we predicted that individual genome-wide heterozygosity ( $h$ ) might influence a wide range of complex disease traits. Although there is an extensive literature on adverse effects of inbreeding on reproduction, childhood mortality and rare Mendelian disorders, the effects of inbreeding on late-onset traits are largely unknown (17–19).

In order to investigate the hypothesis that the heritable component of late-onset diseases includes a major class of deleterious recessive alleles, we have recently studied the effects of inbreeding on BP among 2760 adult individuals in a Dalmatian island isolate. The study demonstrated a large effect equivalent to a rise in SBP of ~20 mmHg and DBP of ~12 mmHg in offspring of first cousin marriages. The effect appeared to be mediated by several hundred recessive alleles as a result of increased homozygosity (11). We extended this observation by investigating the relationship between inbreeding and the prevalence of 10 late-onset complex diseases of public health importance. Our preliminary conclusions were that these data are consistent with our hypothesis that inbreeding causes an increase in homozygosity at many genetic loci with small and recessive deleterious effects on homeostatic pathways, resulting in non-specific breakdown of compensatory mechanisms, which, in turn, result in increased disease risk. This is consistent with the finding of greater influence of inbreeding on late-onset traits possibly due to greater sensitivity of homeostatic mechanisms in later life (6) and with reports of inbred animals having poorer health outcomes when in a natural habitat (20,21). In this study, we have generalized this finding to investigate the effects of  $h$  on a range of traits which underlie >20% (hypertension 11% and hypercholesterolaemia 8%) of the total burden of disease in industrialized countries in 2000 (10).

### Evidence for dominance variance in traits under study

Quantitative genetic theory predicts that a change in  $h_R$  should have most influence on late-onset traits that show significant dominance variance (7,10). The broad and narrow sense heritabilities of the traits, which we studied, have been reported in

other populations (9,22,23). We expect that the reduced variance in environmental exposures that is characteristic of all genetic isolate populations will act to increase heritability in our study population. In addition, owing to inbreeding and relationships through multiple lines of descent, we expect that our power to detect dominance variance will be increased in this population. Our findings are consistent with the published reports of dominance variance in BP, total cholesterol and LDL cholesterol (8,9). In contrast, height, body mass index, HDL cholesterol, creatinine, urate and forced expiratory volume (FEV<sub>1</sub>), which have not been shown to exhibit dominance variance in previous studies (8,9,24), showed no significant association with  $h_R$  in this study.

### Study results

Our experience in utilizing genetic short tandem repeat (STR) markers to estimate heterozygosity led to a finding of a positive correlation between the number of markers successfully typed and heterozygosity (negative correlation with  $h$ ). We excluded marker data from individuals for whom there were data on less than 750 markers and corrected for any remaining effect statistically. This finding is consistent with the recent report that 'missing data can also result in false positives (homozygotes) if samples with a particular genotype are more likely not to be classified during genotyping; many genotyping methods have lower success rates for heterozygotes, so this scenario is not unrealistic' (25). We believe that this is due to the identification of alleles that show large differences in size being more error prone than where there are only small differences in size. This can result from allelic outliers not being recognized and differences in peak height resulting from different efficiencies of amplification of alleles that differ in size. The correlation between number of successfully typed markers and relative heterozygosity is most likely due to differences in DNA quality. Poor DNA quality is more prone to inefficient amplification of larger alleles, which will tend to reduce  $h_r$  values. In this study, including marker data from all individuals would have introduced bias because of a relative loss of heterozygotes (increase of homozygotes) in individuals with the highest proportion of missing marker data. It is important that future studies are aware of and correct for this potential bias.

We selected individuals from a meta-population of genetic isolates to maximize the range of observed  $h_R$  and were able to confirm the large variation in the value of  $h_R$ . This variation in  $h_R$  is unlikely to be random or stochastic as it correlates closely with individual genetic histories based on three generation pedigrees and the knowledge of the birthplaces of the parents and grandparents of participants. Our measure of  $h_R$  is able to give a valid comparison of heterozygosity across groups, as it controls for differing allele frequencies across different communities. We suggest, therefore, that  $h_R$  influenced by population substructure, recent inbreeding and population admixture are the underlying mechanisms of the effects that we have observed.

The inverse relationship between  $h_R$  and BP extends the findings of previous reports of inbreeding effects on this trait (11,26,27), whereas those with total and LDL cholesterol and FEV<sub>25</sub> are novel findings. The strong interaction in

effect on SBP and DBP between  $h_R$  and sex is also a novel finding. Sex-specific dominance effects have been reported for a variety of QTs in humans (8,9,28), pointing to the possible importance of variants affecting genetic regulation in determining trait values (29). Strong empirical support for frequent and substantial sex effects on heritability (including dominance components) in complex traits has been described in other organisms, including *Drosophila* (30). There are a number of possible explanations for the effect of gender on BP, although there is little direct evidence bearing on this issue. These include hormonal effects on endothelial function and oxidative stress. In males, the lack of oestrogenic protection against oxidative damage to endothelial cells when compared with pre-menopausal females may mask genetic differences in oestrogen-response genes.

The effects of  $h_R$  on BP are in the same direction as, but somewhat lower than those that we reported previously in the sample of 2760 subjects from Brac, Hvar and Korcula. These discrepancies could be explained, at least, in part by: (i) effects of medication, ignoring of which leads to underestimation of effect size (medication was not available at the time of the previous study); (ii) random variation, reflected in the large standard errors of the estimated effect sizes and (iii) regression dilution resulting from imprecision in the  $F$ -estimates [predicted underestimation of the regression coefficients of ~47% (see Materials and Methods)].

The reason why there is an inverse relationship between  $h_R$  and BP is unclear, but may be a consequence of directional dominance, which has been widely observed for complex traits in response to inbreeding. It may relate to the breakdown of homeostatic mechanisms with age. Virtually, all specific genetic effects identified to date in hypertension serve to increase rather than decrease BP with increasing age, providing further support for the presence of alleles showing directional dominance (31).

### Implications of study results

We propose that relative (genome-wide) heterozygosity ( $h_R$ ) could represent a new genetic risk factor in observational analytic studies in epidemiological research on common complex disease traits. The findings of this study are consistent with reported inbreeding effects (associated with lower  $h_R$ ) on disease traits in animals and plants (32,33) and in humans (12). However, this study extends previous reports of 'inbreeding depression' on fitness and health-related traits by demonstrating a positive effect of increased relative heterozygosity, which is consistent with the positive effects of heterosis and hybrid vigour, which have been demonstrated in animals and plants, especially in the F1 generation (34). Thus, the breakdown of past population substructures and greater admixture resulting from greatly increased human population mixing and urbanization may be leading to increasing  $h_R$  and beneficial effects on a range of traits associated with human health and disease.

The observed adverse effect of  $h_R$  on the prevalence of several different late-onset traits that represent major risk factors for common complex diseases of late onset is consistent with the presence of many deleterious recessive alleles, located throughout the genome (15). This implies that the

study of inbred populations would be advantageous as the increased gene dosage of such variants in inbred individuals will tend to amplify their phenotypic effects when compared with outbred populations, where most alleles are present in heterozygotes (13). It is also consistent with a more general effect of increased homozygosity at these loci, leading to an accumulation of small deleterious effects on homeostatic pathways, which cumulatively increase disease risk. This suggests a greater sensitivity of homeostatic mechanisms to inbreeding in later life, as predicted by findings in animals. Decay of homeostatic capacity would also be expected to lead to reduced capacity to respond appropriately to diverse stimuli. This is supported by the recent observations that the reduced survival found in inbred animals is greater in the natural habitat than in a controlled laboratory environment (35).

The adverse effect of inbreeding (with associated lower  $h_R$ ) on BP has been reported now in several large studies (11,26,27,36). The size of this effect is substantial and could be of public health significance. The sex-specific nature of this effect has not been reported previously, but is consistent with marked sex-specific differences in dominance variance of BP traits which have been reported in the Hutterite population (28). Further studies are warranted to investigate this effect in more detail.

Raised BP and total/LDL cholesterol are estimated to be responsible for ~20% of the total burden of disease in industrialized countries (10) so that our finding of beneficial effects of outbreeding on these and other health-related traits may be of public health significance. The association of higher  $h$  with lower SBP, DBP, total/LDL cholesterol is consistent with the well-known fitness effects of outcrossing in animals and plants. The breakdown of population structure resulting from urbanization may, therefore, have beneficial effects on a range of traits associated with human health and disease. Other traits, such as age at menarche, IQ (37) and lifespan, which have been changing in parallel with urbanization, could also be influenced by this factor.

## MATERIALS AND METHODS

### Croatian island study population

The village populations of neighbouring islands in the eastern Adriatic, Middle Dalmatia, Croatia (Fig. 1) represent a well-characterized meta-population of genetic isolates (2,38–43). Although sharing similar environments, divergence in behaviour patterns for socio-cultural and economic reasons through history led them to adopt consanguinity practices ranging from extreme inbreeding to complete avoidance of inbreeding, providing an excellent setting for this study because we required to represent a broad range of heterozygosities,  $h$ , in order to maximize the power of our regression-based analysis.

### Selection of 1001 individuals in the overall study population

Nine island settlements were carefully chosen to represent a wide range of distinct and well-documented demographic histories. We have previously reported very high level of

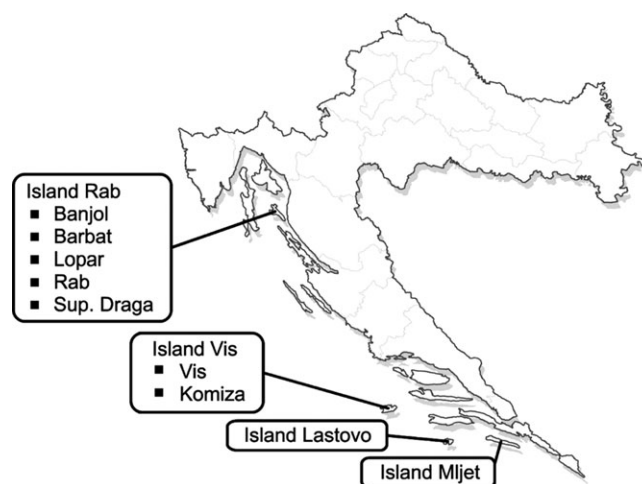


Figure 1. Map of Dalmatian genetic isolate.

differentiation between most of these island communities based on Wright's fixation indexes (43). We know from previous studies that the most common form of inbreeding in these villages is at the level of second cousins marriages (expected inbreeding coefficient in their progeny of 1.56%). The observed inbreeding coefficients ranged from rare progeny of first cousins ( $F$  of 6.25%) to more common progeny of third cousins ( $F$  of 0.4%). The mean inbreeding coefficients based on genealogy and marker data are presented in Supplementary Material, Table S1 and Figure S2, respectively, and the relationship between these two measures in Supplementary Material, Table S2 and Figure S2. The fieldwork was performed during 2002 and 2003 by a team that included employees of the School of Public Health of the University of Zagreb Medical School and the Institute for Anthropological Research in Zagreb, Croatia. In each of the nine villages, a random sample of 100 adult inhabitants was recruited. Sampling was based on computerized randomization of the most complete and accessible population registries in each village, which included medical records (Mljet and Lastovo islands), voting lists (Vis island) and household numbers (Rab island). An additional 101 examinees were recruited from second-generation immigrants into all nine villages to form a genetically diverse control population sharing the same environment. Ethical approval for this research was obtained from appropriate research Ethics Committees in Croatia and Scotland. Informed written consent was obtained from all participants in the study.

### Selection of a subsample of individuals for current analysis

Participants provided information on the names and place(s) of birth of their parents and grandparents. From these data, the percentage of grandparents born in the same village as the participant was estimated. In order to maximize the power of the regression analysis for a given genotyping cost, we sought to study individuals likely to represent a wide range of  $h$ . Thus, we identified 76 individuals for whom there was evidence of recent inbreeding (category a). These

comprised six from Banjol, 13 from Barbat, six from Lopar, three from Rab, four from Supetarska Draga, 13 from Vis, 18 from Komiza, two from Lastovo and 11 from Mljet villages. There were 36 men and 40 women. The mean (median) age of this sample was 58.4 (61) years with an age range from 21 to 83 years.

For each of these, we also identified 228 ( $76 \times 3$ ) adults who were individually matched [by village of residence, age ( $\pm 5$  years) and gender, level of education and basic socio-economic indicators], but with widely differing genetic backgrounds according to personal genetic history (categories b–d). We further identified another 28 individuals whose grandparents either came from diverse backgrounds ( $n = 25$ ) and who had no history of recent inbreeding or who were offspring of inbred individuals ( $n = 3$ ). All these subjects were selected independent of and blind to their phenotype measurements. This yielded adults in the following four groups for study:

- (a) inbred: selected when there was evidence of recent inbreeding from genealogical data;
- (b) 'endogamous': when all four of the subject's grandparents were born in the subject's village of residence, but there was no evidence of recent inbreeding from genealogical data;
- (c) 'exogamous': when two to three of the subject's grandparents were born in the subject's village of residence, but there was no evidence of recent inbreeding from genealogical data;
- (d) outbred: zero to one of the subject's grandparents born in subject's village of residence, and no two were born in the same small settlement.

Three individuals had a genotypic sex that was apparently incompatible with their reported sex. These may have been due to clerical errors and were excluded from analyses of all outcome variables, although they were retained for calculations of allele frequencies and heterozygosities.

#### **Phenotypic and other measurements performed in the sample**

All of the examinees were first interviewed by one of the trained surveyors, based on a questionnaire developed for this research programme. The questionnaire collected data on personal characteristics (name, date and place of birth, gender, marital status, education level and occupation), selected health-related lifestyle variables (such as diet and smoking status), health complaints, drug intake and hospitalization records. A socio-economic status score was defined by the number of positive answers to a set of four questions on ownership of household goods chosen to discriminate between groups within the study population. SBP and DBP in mmHg, lung function parameters, height (mm) and weight (kg) were each measured in local health centres and dispensaries between 8 and 11 a.m. All physiological measurements were made by a survey team with many years of experience in similar surveys and using standard methods. BP measurements were taken after a 10 min period of rest on two occasions in subjects lying down with the cuff at the

level of the heart. Korotkoff sound V was taken as the diastolic pressure. Biochemical analyses of creatinine, uric acid, HDL, LDL, total cholesterol, triglycerides and blood glucose were performed on fasting blood samples in fasting individuals (collected between 8:30 and 9:30 a.m.). Samples were aliquoted, stored in a  $-20^{\circ}\text{C}$  freezer without delay and then transported frozen to a single biochemical laboratory based in Zagreb. The laboratory employed stringent internal quality control procedures and participated in and was accredited for performing the analyses under study by an international external quality assurance (RIQAS) group. Indexes of smoking and alcohol consumption were obtained from standard questionnaires. All data were (double) entered into a database for analysis. The field work and data collection were carried out according to the procedures laid down by UK and Croatian Ethics Committees. These included informed consent from all participants, covering the genetic analyses in the study.

#### **DNA extraction and analysis**

DNA was extracted from EDTA whole blood specimens by the salting out procedure. DNA was stored at the MRC Human Genetics Unit in Edinburgh. Thirty micrograms of DNA in the required format was sent for genotyping to the Center for Medical Genetics, Marshfield Medical Research Foundation, USA.

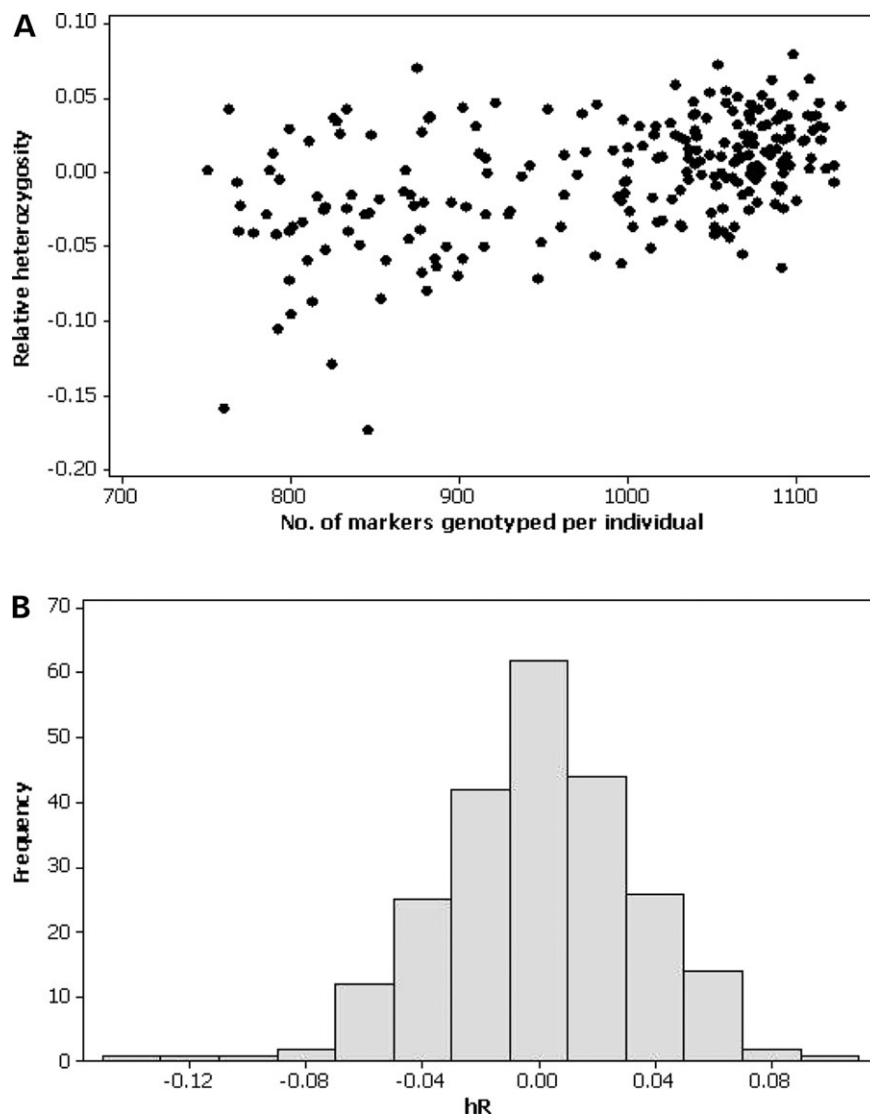
#### **Genotyping**

Samples were analysed with 1184 autosomal markers comprising microsatellite markers taken from the Marshfield Screening Set and indel markers. Details of the genotyping methods can be found at <http://research.marshfieldclinic.org/genetics>.

#### **Calculation of genomic $h_R$ values**

Values of  $h_R$  were defined as relative heterozygosity [relative heterozygosity ( $h_R = \text{excess heterozygosity/expected heterozygosity}$ ) calculated using the weighted version of a method we have recently described (14). Two methods of calculating the expected heterozygosities were used (14). The first was simply the proportion of heterozygotes observed at each locus among those individuals for which a genotype was reported. The second was the HW value, based on the observed allele frequencies from these individuals. The correlation between the values of  $h$  calculated by the two methods was 0.98. Their orderings and relative magnitudes were very similar, but the mean value using the second method considerably exceeded that using the first method. This presumably reflects the fact that the level of heterozygosity in this mixed population sample is expected to be lower than that predicted under HWE because of the genetic subdivisions within this meta-population. Because we are concerned only with relative heterozygosity values in this study, however, further results were restricted to those based on the first method.

We found a positive correlation between the number of STR markers successfully typed and heterozygosity. This finding is consistent with a recent report by Hirschhorn and Daly (25). In this review, including marker data from all individuals would have introduced bias because of a relative loss of



**Figure 2.** (A) Relation between relative heterozygosity and numbers of markers genotyped per individual ( $N = 235$ ). (B) Histogram of relative heterozygosity values ( $h_R$ ), after removing the linear trend shown in Figure 2(A).

heterozygotes (increase in homozygotes) in individuals with the highest proportion of missing marker data. We therefore excluded marker data from individuals for whom there were data on fewer than 750 markers. There remained a dependency, albeit much reduced, between  $h$  and the numbers of reported genotypes, even within this restricted sample of individuals (Fig. 2A and B). This dependency was eliminated by regressing the  $h$  values on the number of genotypes, weighted by the inverse of their variance, and using the residuals ( $h_R$ ) as a predictor variable in further analysis.

### Statistical analysis and modelling

**Variation in  $h$  across sample groups.** Data from 1184 autosomal short tandem repeat and indel markers were used to estimate genome-wide heterozygosity ( $h$ ) in each individual. We explored the range of  $h$  across groups by general linear

modelling by estimating relative heterozygosity ( $h_R = \text{excess heterozygosity/expected heterozygosity}$ ), where expected heterozygosity is the average heterozygosity across all samples. The differences in relative heterozygosity between the four study groups were explored by general linear modelling with  $h_R$  as the dependent variable, and  $\log(\text{height})$ ,  $\log(\text{weight})$ , years of schooling, socio-economic status score, age, sex, smoking and alcohol indexes and sample group as predictors. Starting with the full model, variables not significant at the  $P = 0.05$  level were dropped one at a time until no further simplification was possible. The final model contained sample group as the only significant predictor ( $P < 0.0005$ ).

**Effect of  $h_R$  on QT values.** A total of 15 QT variables were analysed as outcome variables. For the residuals to conform approximately to normality, as required by the standard least-squares regression, five variables were log-transformed

and one was log(log)-transformed (Table 3). Two extreme outliers were also removed. To explore the possible effects of  $h_R$  on each of these, a multiple regression model was constructed. In this model, age, sex, ln height, ln weight, years of schooling, socio-economic status score, smoking and alcohol indexes and  $h_R$  were forced as predictors. As mentioned earlier, the full model was fitted first and then variables not significant at the  $P = 0.05$  level were dropped one at a time until no further simplification was possible (apart from age and sex which were forced into the model, regardless of significance). SBP and DBP were taken as the mean of two SBP readings or of two DBP readings, where two readings were available (208 cases), or as a single reading where only one was available (25 cases).

We estimated the extent of underestimation of the regression coefficients for BP because of regression dilution resulting from imprecision in  $h_R$  estimates ( $\sim 47\%$ ) as follows:  $b_{\text{true}} = b_{\text{est}} \times (1 + v_{\text{est}}/v_F)$ , where  $b_{\text{true}}$  is the true regression coefficient,  $b_{\text{est}}$  the estimated regression coefficient,  $v_{\text{est}}$  the estimation variance of  $h_R$  values and  $v_F$  the real variation of  $h_R$  values in the sample (44); in this case, the mean  $v_{\text{est}}$  was 0.0005585 and the overall variance of  $F$  values was 0.0011935, giving  $v_F = 0.0011935 - 0.0005585 = 0.000635$ .

The analysis was repeated with interaction terms for sex  $\times$   $h_R$  and age  $\times$   $h_R$  included as predictors. To explore the possible effects of hypertension medication on the relationship between  $h_R$  and SBP and DBP, subjects were recorded as taking/not taking antihypertensive medication. Duration and dose of treatment were not recorded. We postulated that the effect of medication would be to decrease mean SBP by 15 mm, then explored this effect by adding 15 mm to the mean SBP for those recorded as taking treatment as recommended by Tobin *et al.* (45).

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

## ACKNOWLEDGEMENTS

We would like to thank staff of the Institute for Anthropological Research in Zagreb, Croatia for their help during the field work and sample preparation. We thank Dr P. Visscher and Dr J. Wilson for their helpful comments and discussion. This work was supported in part by the National Heart Lung and Blood Institute Mammalian Genotyping Service (contract no. HV48141), European Commission FP6 STREP grant number 018947 (LSHG-CT-2006-01947), the Wellcome Trust, the UK Medical Research Council, the Royal Society (UK), the British Council and the Croatian Ministry of Science, Education and Sport (grants I.R. 108-1080315-0302; P.R. 196-1962766-2751; B.J. 196-1962766-2763; N.SN 196-1962766-2747).

*Conflict of Interest statement.* The authors declare that they have no competing financial interests.

## REFERENCES

1. Clark, D. (2004) *Urban World, Global City*. Routledge, UK.

2. Vitart, V., Carothers, A.D., Suffolk, R., Hayward, C., Teague, P., Hastie, N.D., Campbell, H. and Wright, A.F. (2005) Increased level of linkage disequilibrium in rural compared to urban communities: a factor to consider in association study design. *Am. J. Hum. Genet.*, **76**, 763–772.
3. Darvasi, A. and Shifman, S. (2005) The beauty of admixture. *Nat. Genet.*, **37**, 118–119.
4. Helgason, A., Ingvadottir, B., Hrafnkelsson, B., Gulcher, J. and Stefansson, K. (2005) An Icelandic example of the impact of population structure on association studies. *Nat. Genet.*, **37**, 90–95.
5. Kobylanski, E. and Livshits, G. (1989) Age-dependent changes in morphometric and biochemical traits. *Ann. Hum. Biol.*, **16**, 237–247.
6. Charlesworth, B. and Hughes, K.A. (1996) Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence. *Proc. Natl Acad. Sci. USA*, **93**, 6140–6145.
7. Falconer, D.S. and Mackay, T.F.C. (1996) *Introduction to Quantitative Genetics*, 4th edn. Prentice Hall, Harlow, UK.
8. Ober, C., Abney, M. and McPeck, M.S. (2001) Familial studies of medical and anthropometric variables in a human isolate. *Am. J. Hum. Genet.*, **69**, 1068–1079.
9. Abney, M., McPeck, M.S. and Ober, C. (2001) Broad and narrow heritabilities of quantitative traits in a founder population. *Am. J. Hum. Genet.*, **68**, 1302–1307.
10. World Health Organisation (2002) *World Health Report 2002: reducing risks, promoting healthy life*. WHO, Geneva.
11. Rudan, I., Campbell, H., Carothers, A., Wright, A., Smolej-Narancic, N., Skaric-Juric, T. and Rudan, P. (2003) Inbreeding and the genetic complexity of human hypertension. *Genetics*, **163**, 1011–1021.
12. Rudan, I., Rudan, D., Campbell, H., Carothers, A., Wright, A., Deka, R., Smolej-Narancic, N., Janicijevic, B. and Rudan, P. (2003) Inbreeding and risk of complex chronic diseases. *J. Med. Genet.*, **40**, 925–932.
13. Wright, A., Charlesworth, B., Rudan, I., Carothers, A. and Campbell, H. (2003) A polygenic basis for late-onset disease. *Trends. Genet.*, **19**, 97–106.
14. Carothers, A.D., Rudan, I., Hayward, C., Teague, P., Vitart, V., Rudan, P., Polasek, O., Kolcic, O., Campbell, H., Weber, J. and Wright, A.F. (2006) Estimating human individual inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. *Ann. Hum. Genet.*, **70**, 666–676.
15. Jimenez-Sanchez, G., Childs, B. and Valle, D. (2000) Human disease genes. *Nature*, **400**, 853–855.
16. Brown, W.M., Beck, S.R., Lange, E.M., Dvid, C.C., Kay, C.M., Langefeld, C.D. and Rich, S.S. (2003) Framingham Heart Study. Age-stratified heritability estimation in the Framingham Heart Study families. *BMC Genet.*, **4**(Suppl. 1), S32.
17. Bittles, A.H., Mason, W.M., Greene, J. and Rao, N.A. (1991) Reproductive behaviour and health in consanguineous marriages. *Science*, **252**, 789–794.
18. Bittles, A.H. and Neel, J.V. (1994) The costs of human inbreeding and their implications for variations at the DNA level. *Nat. Genet.*, **8**, 117–121.
19. Rudan, I. and Campbell, H. (2004) Five reasons why inbreeding may have considerable effect on post-reproductive human health. *Coll. Antropol.*, **28**, 943–950.
20. Jimenez, J.A., Hughes, K.A., Alaks, G., Graham, L. and Lacy, R.C. (1994) An experimental study of inbreeding depression in a natural habitat. *Science*, **266**, 271–273.
21. Acevedo-Whitehouse, K., Gulland, F., Greig, D. and Amos, W. (2003) Inbreeding: disease susceptibility in Californian sea lions. *Nature*, **422**, 35.
22. Brenn, T. (1994) Genetic and environmental effects on coronary heart disease risk factors in northern Norway. The cardiovascular disease study in Finnmark. *Ann. Hum. Genet.*, **58**, 369–379.
23. Abney, M., McPeck, M.S. and Ober, C. (2000) Estimation of variance components of quantitative traits in inbred populations. *Am. J. Hum. Genet.*, **66**, 629–650.
24. Wilk, J.B., Djousse, L., Borecki, I., Atwood, L.D., Hunt, S.C., Rich, S.S., Eckel, J.H., Arrett, D.K., Rao, D.C. and Myers, R.H. (2000) Segregation analysis of serum uric acid in the NHLBI Family Heart Study. *Hum. Genet.*, **106**, 355–359.
25. Hirschhorn, J.N. and Daly, M.J. (2005) Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.*, **6**, 95–108.

26. Martin, A.O., Kurczynskim, T.W. and Steinberg, A.G. (1973) Familial studies of medical and anthropometric variables in a human isolate. *Am. J. Hum. Genet.*, **25**, 581–593.
27. Krieger, H. (1986) Inbreeding effects on metrical traits in Northeastern Brasil. *Am. J. Hum. Genet.*, **21**, 537–546.
28. Weiss, L.A., Pan, L., Abney, M. and Ober, C. (2005) The sex-specific genetic architecture of quantitative traits in humans. *Nat. Genet.*, **38**, 218–222.
29. Ranz, J.M., Castillo-Davis, D.I., Meiklejohn, C.D. and Hartl, D.L. (2003) Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science*, **300**, 1742–1745.
30. Mackay, T.F.C. (2001) Quantitative trait loci in *Drosophila*. *Nat. Rev. Genet.*, **2**, 11–20.
31. Lifton, R.P., Gharavi, A.G. and Geller, D.S. (2001) Molecular mechanisms of human hypertension. *Cell*, **104**, 545–556.
32. Keller, L.F. and Waller, D.M. (2002) Inbreeding effects in wild populations. *Trends Ecol. Evol.*, **17**, 230–242.
33. Kristensen, T.N. and Sorensen, A.C. (2005) Inbreeding—lessons from animal breeding, evolutionary biology and conservation genetics. *Animal Sci.*, **80**, 121–133.
34. Charpentier, M., Setchell, J.M. and Prugnolle, F. (2005) Genetic diversity and reproductive success in mandrills (*Mandrillus sphinx*). *Proc. Natl Acad. Sci. USA*, **102**, 16723–16728.
35. Joron, M. and Brakefield, P.M. (2003) Captivity masks inbreeding effects on male mating success in butterflies. *Nature*, **424**, 191–194.
36. Badaruddoza (2004) Inbreeding effects on metrical phenotypes among North Indian Children. *Coll. Antropol.*, **28**(Suppl. 2), 311–319.
37. Mingroni, M.A. (2004) The secular rise in IQ: giving heterosis a closer look. *Intelligence*, **32**, 65–83.
38. Rudan, P., Simic, D., Smolej-Narancic, N., Bennett, L.A., Janicijevic, B., Jovanovic, V., Lethbridge, M.F., Milicic, J., Roberts, D.F. and Sujoldzic, A. (1987) Isolation by distance in Middle Dalmatia-Yugoslavia. *Am. J. Phys. Anthropol.*, **74**, 417–426.
39. Barac, L., Pericic, M., Klaric, I.M., Rootsi, S., Janicijevic, B., Kivisild, T., Parik, J., Rudan, I., Villems, R. and Rudan, P. (2003) Y chromosomal heritage of Croatian population and its island isolates. *Eur. J. Hum. Genet.*, **11**, 535–542.
40. Klaric, I.M., Pericic, M., Lauc, L.B., Janicijevic, B., Kubat, M., Pavicic, D., Rudan, I., Wang, N., Jin, L., Chakraborty, R., Deka, R. and Rudan, P. (2005) Genetic variation at nine short tandem repeat loci among islanders of the eastern Adriatic coast of Croatia. *Hum. Biol.*, **77**, 471–486.
41. Saftic, V., Rudan, D. and Zgaga, L. (2006) Mendelian diseases and conditions in Croatian island populations: historic records and new insights. *Croat. Med. J.*, **47**, 543–552.
42. Rudan, I., Biloglav, Z., Carothers, A.D., Wright, A.F. and Campbell, H. (2006) Strategy for mapping quantitative trait loci (QTL) by using human metapopulations. *Croat. Med. J.*, **47**, 532–542.
43. Vitart, V., Biloglav, Z., Hayward, C., Janicijevic, B., Smolej-Narancic, N., Barac, J., Pericic, M., Martinovic Klaric, I., Polasek, O., Kolcic, I. *et al.* (2006) 3,000 years of solitude: extreme level of differentiation in the island isolates of the Dalmatian coast. *Eur. J. Hum. Genet.*, **14**, 478–487.
44. Snedecor, G.W. and Cochran, W.G. (1967) *Statistical Methods*, 6th edn. Iowa State University Press, Ames, Iowa.
45. Tobin, M.D., Sheehan, N.A., Scurrah, K.J. and Burton, P.R. (2005) Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat. Med.*, **24**, 2911–2935.