BMC Medical Genetics

BioMed Central

Research article

Homozygosity and risk of childhood death due to invasive bacterial disease

Emily J Lyons[†1,4], William Amos*[†2], James A Berkley[3], Isaiah Mwangi[2], Mohammed Shafi[3], Thomas N Williams[3], Charles R Newton[3], Norbert Peshu[3], Kevin Marsh[3], J Anthony G Scott[3] and Adrian VS Hill[1]

Address: [1]The Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK, [2]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK, [3]Centre for Geographic Medicine Research (Coast), Kilifi, Kenya and [4]Current address: MRC Centre for Outbreak Analysis and Modeling, Department of Infectious Disease Epidemiology, Imperial College, St. Mary's Campus, Norfolk Place, London W2 1PG, UK

Email: Emily J Lyons - e.lyons@imperial.ac.uk; William Amos* - w.amos@zoo.cam.ac.uk; James A Berkley - jberkley@kilifi.kemri-wellcome.org; Isaiah Mwangi - imwangi@kemri-nuitm.or.ke; Mohammed Shafi - jshafi@africaonline.co.ke; Thomas N Williams - tom.n.williams@gmail.com; Charles R Newton - CNewton@kilifi.kemri-wellcome.org; Norbert Peshu - NPeshu@kilifi.kemri-wellcome.org; Kevin Marsh - KMarsh@kilifi.kemri-wellcome.org; J Anthony G Scott - ascott@ikilifi.net; Adrian VS Hill - avsh@gwmail.jr2.ox.ac.uk

* Corresponding author   †Equal contributors

Abstract

Background: Genetic heterozygosity is increasingly being shown to be a key predictor of fitness in natural populations, both through inbreeding depression, inbred individuals having low heterozygosity, and also through chance linkage between a marker and a gene under balancing selection. One important component of fitness that is often highlighted is resistance to parasites and other pathogens. However, the significance of equivalent loci in human populations remains unclear. Consequently, we performed a case-control study of fatal invasive bacterial disease in Kenyan children using a genome-wide screen with microsatellite markers.

Methods: 148 cases, comprising children aged <13 years who died of invasive bacterial disease, (variously, bacteraemia, bacterial meningitis or neonatal sepsis) and 137 age-matched, healthy children were sampled in a prospective study conducted at Kilifi District Hospital, Kenya. Samples were genotyped for 134 microsatellite markers using the ABI LD20 marker set and analysed for an association between homozygosity and mortality.

Results: At five markers homozygosity was strongly associated with mortality (odds ratio range 4.7 – 12.2) with evidence of interactions between some markers. Mortality was associated with different non-overlapping marker groups in Gram positive and Gram negative bacterial disease. Homozygosity at susceptibility markers was common (prevalence 19–49%) and, with the large effect sizes, this suggests that bacterial disease mortality may be strongly genetically determined.

Conclusion: Balanced polymorphisms appear to be more widespread in humans than previously appreciated and play a critical role in modulating susceptibility to infectious disease. The effect sizes we report, coupled with the stochasticity of exposure to pathogens suggests that infection and mortality are far from random due to a strong genetic basis.

Background

Many recent studies of natural populations report a correlation between genetic heterozygosity (heterozygosity-fitness correlation, HFC), measured at a small number (of the order of 10) of presumed neutral markers, and fitness [1,2]. Fitness measures range widely from survival [3] and reproductive success [4-6] to indirect traits such as song complexity [7] and territory size [8] in birds. Some of the most frequently reported traits relate to immune function [9] and susceptibility to micro- [10] and macroparasites [11-13]. Such studies raise obvious questions, both about the mechanism responsible, and whether similar patterns may affect humans.

Two primary mechanisms have been suggested to explain HFCs [14,15]. First, relatively homozygous individuals may be more susceptible to infection because they are inbred. Here, average heterozygosity at the panel of markers being genotyped estimates genome-wide heterozygosity, which in turn estimates the inbreeding coefficient, *F*. However, several theoretical treatments have come to the conclusion that such a mechanism is unlikely to operate in most real populations [16-18]. The problem is that random mating generates extremely few individuals with sufficiently high *F* for their heterozygosity to stand out when measured at tens or even hundreds of markers, unless the population is very small or highly polygynous. Humans may offer a further exception in cultures where cousin marriages are actively encouraged [19], potentially increasing the rate of heritable diseases [20,21].

The second mechanism that may generate HFCs involves chance linkage between one or more of the markers and a gene(s) experiencing balancing selection. Balancing selection has often been thought to be rather rare, particularly in humans [22] where the classical example is sickle cell anemia [23] remains one of very few examples. Moreover, while some argue that polymorphism at immune function genes is maintained by overdominant balancing selection [24], there is evidence that this is unlikely to be effective at maintaining more than two alleles [25-27]. Regardless of theory, a number of recent HFC studies report convincing associations between heterozygosity at one particular locus and the measured trait [13,28-31].

Over the last five to ten years, association studies examining the genetic basis of human disease have switched overwhelmingly from microsatellite markers to single nucleotide polymorphisms (SNPs) [32]. SNPs are much less polymorphic than microsatellites, a deficiency that is usually compensated for by the vastly greater number of markers being genotyped. However, while there are many advantages to using SNPs for the assessment of local heterozygosity, microsatellites offer an arguably more direct approach that circumvents the need to reconstruct complex haplotypes. To assess the possible importance of HFCs in humans, we therefore conducted a case-control study in a population of Kenyan children, using a panel of microsatellite markers to quantify both local and genome-wide heterozygosity.

Results

All our samples were drawn from a prospective study in Kilifi District Hospital and were genotyped for 134 microsatellite markers using the ABI LD20 marker set (Applied Biosystems, USA) [see Additional file 1]. Cases (n = 148) comprised a consecutive series of children aged <13 years who died of invasive bacterial disease, (variously, bacteraemia, bacterial meningitis or neonatal sepsis, for details see methods), a major contributor to childhood mortality in the developing world [33]. Controls comprised 137 randomly selected healthy children matched on age to the cases. Microsatellite traces were scrutinised carefully to ensure homozygotes were identified with high accuracy.

For the study of HFCs a number of measures of heterozygosity have been proposed that offer potential benefits over straight heterozygosity, weighting scores variously by allele size (mean d^2)[3], allele frequencies (internal relatedness, IR) [6] and the variability of loci scored (HL) [34]. However, in automated high throughput studies, heterozygosity assessment can sometimes be problematic, particularly where time for scrutiny of every trace is limited. Thus, null (non-amplifying) alleles, allele drop-out and, at some loci, high levels of stutter-bands can all contribute to a tendency for a minority of loci to carry misleading genotypes where heterozygotes are called as homozygotes or vice versa. Issues have also been identified with allele binning, in some cases causing single alleles to be split between two length classes [35]. In an attempt to circumvent these problems we spent most empirical effort ensuring that heterozygotes and homozygotes were accurately scored and introduce a variant of the measure Standardised Heterozygosity (SH) [2], designed to be highly conservative. SH controls for missing data by expressing heterozygosity as the ratio of the observed heterozygosity in an individual relative to the expected value at the markers genotyped, assuming Hardy-Weinberg equilibrium. Our measure, Standardised Observed Homozygosity (SOH), follows the same principle but instead of calculating the expected homozygosity from the allele frequencies, we used the observed homozygosity at each locus. In this way, SOH measures the extent to which any given individual is more or less homozygous relative to the level expected if all genotypes were randomized among individuals, negating the requirement for accurate allele frequency estimates and reducing the impact of allele drop out, null alleles and other possible artefacts.

We first asked whether SOH varied significantly among disease categories by conducting a one-way ANOVA. Raw SOH values exhibit a slightly skewed distribution, but this is removed by a simple log transformation (Shapiro-Wilk normality test, W = 0.9941, p = 0.322). Following transformation, SOH revealed highly significant variation among disease classes ($F_{[5,281]}$ = 6.75, P = 5.89 × 10^{-6}) (Figure 1). However, when the control class was excluded, the ANOVA was no longer significant ($F_{[4,144]}$ = 0.785, P = 0.54), indicating that the main effect is driven by a difference in heterozygosity between cases and controls rather than between disease classes. The direction of the deviation is toward greater homozygosity in cases compared with the controls.

We next asked whether there was evidence of local effects due to chance linkage between one or more markers and a gene(s) experiencing balancing selection. To test this proposition we calculated age-adjusted odds ratios of mortality at each locus in turn (Figure 2). Most markers show either a non-significant or borderline (at alpha = 0.05) association between homozygosity and risk of mortality. However, nine markers reveal a strong associations with experiment wide significance using full Bonferroni correction (p < 0.00037, see Table 1). This is a highly conservative threshold since where multiple markers are expected, the less stringent false discovery rate approach can be justified [36]. Although the spacing between markers is sufficient to ensure they behave as if unlinked, it is possible that multiple markers contribute to the same risk through linkage to related genes. Consequently, we then constructed a multivariable logistic regression model with mortality as the response and age, sex, locality, SOH and homozygosity at each of the nine largest-effect markers as explanatory variables. Sequential removal of terms that did not contribute significantly (likelihood ratio test, p <
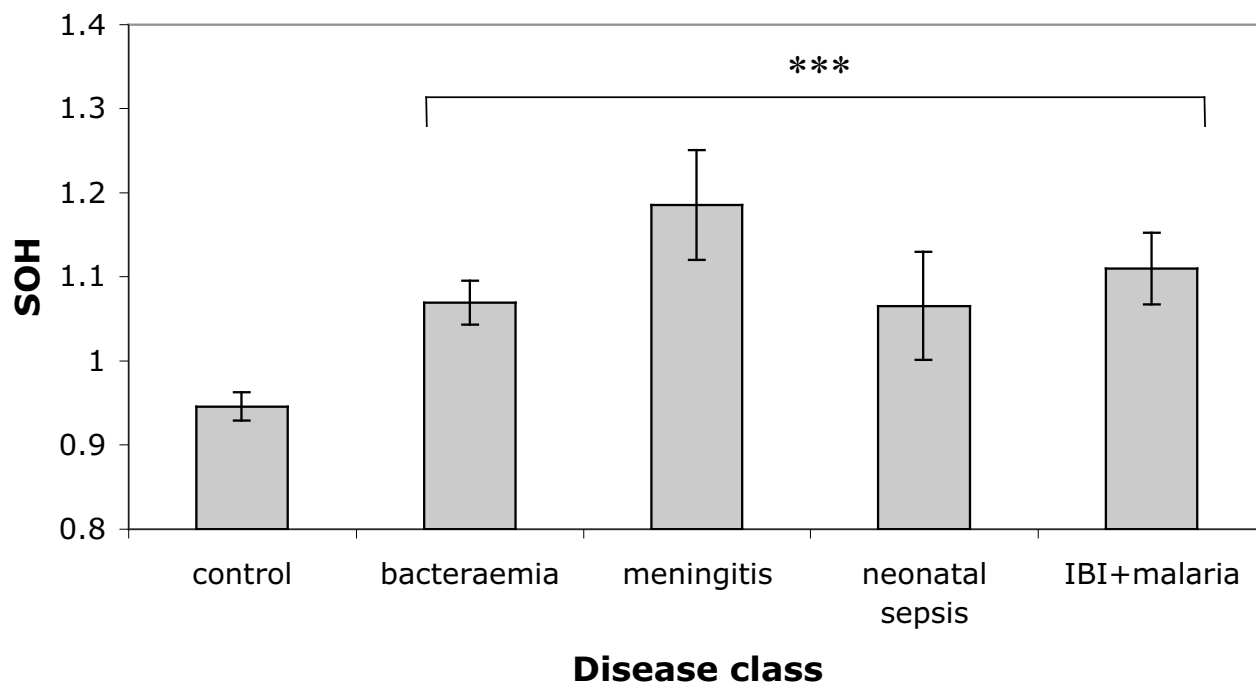

Figure 1

Analysis of variance of standardized observed homozygosity values for cases and controls. S_{OH} is the Standardised Observed Homozygosity for an individual genotyped for *i* loci, calculated as: $S_{OH} = \frac{N_{\text{hom}}}{\sum_i H_{oi}}$ where N_{hom} is the number of homozygote genotypes in the individual concerned and H_{oi} is the observed frequency of homozygotes at one of the *i* loci scored in this individual. ***indicates a highly significant test where P < 1 × 10^{-5}. The IBI + malaria group includes individuals who had invasive bacterial disease but also malaria parasitaemia so that the contribution of the latter to mortality could not be determined with certainty. Sample sizes for the disease classes are as follows: control = 183, bacteraemia = 71, meningitis = 18, neonatal sepsis = 26 and IBI + malaria parasitaemia = 34. IBI: invasive bacterial infection. Error bars are ± 1 standard error.
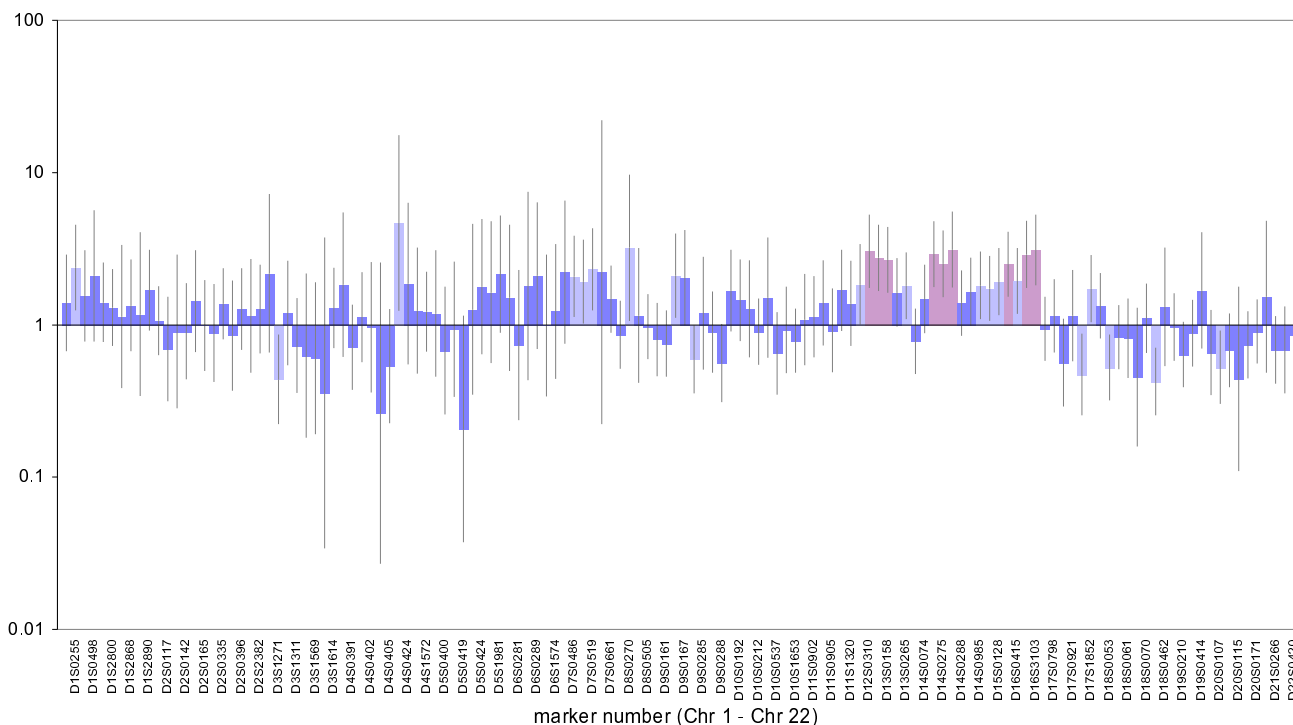
Figure 2
Odds ratios and 95% confidence intervals for mortality and homozygosity by marker, adjusting for age. Age-adjusted odds ratios (with 95% confidence intervals) of mortality at each locus. All markers were tested for significance using a chi-squared test based on a simple 2 × 2 contingency table (case/control vs homozygotes/heterozygotes). ORs shown in dark blue are non-significant. ORs shown in pale blue are significant at P < 0.05 and ORs shown in pink (n = 9) are significant at P < 0.00037 (i.e. significant experiment-wide at P < 0.05).

0.05) yielded a final model containing age, location and five markers (D12S310, D13S158, D14S275, D16S3103, D16S423). SOH is dropped as a marginally significant term (LHR, p~0.07) whether fitted as a continuous variable or as a factor with five levels. By implication, it seems that genome-wide effects (inbreeding depression) are minimal or absent.

Beginning with the final model derived above, we explored further possible interactions between markers, and also between each marker and age. Among all possible pair-wise combinations of markers, two revealed significant interactions, both of which were retained in the model regardless of the order in which they were added (Table 2, last two columns). No significant interactions with age were detected. Our data contain approximately equal numbers of individuals who died from Gram positive (n = 63), gram negative (n = 79) or both (n = 6) infections and this allowed us to ask whether our markers identify genes that impact differently on diseases caused by bacteria of different classes. We therefore repeated the logistic regression approach above on each bacterial class separately, including dual infections in both analyses.

Given the smaller datasets, criteria for initial inclusion in the model were relaxed to an initial OR significant at p < 0.005. The final models are summarized in Table 2 and reveal surprising complexity, with susceptibility to gram positive and gram negative infections associated mostly with non-overlapping genomic locations. Only marker D12S310 is significant in all three models. Marker D9S164 reveals an interaction with age, infants being more likely to die if homozygous (OR = 1.65) and older children less (OR = 0.18). Interactions between marker pairs in the whole dataset suggest that homozygosity at both markers together confers no greater risk than homozygosity at either one alone. However, in the gram negative model the interaction of homozygosity at two markers, D7S486 and D16S423, indicates a significantly synergistic risk of mortality (odds ratio 40.7) where homozygosity at either of the markers alone confers no risk.

Finally, to assess the magnitude of homozygosity effects with respect to the population, we calculated the population attributable risk fraction (PARF) [37] for each marker (Table 3). PARFs indicate the proportional reduction in

Table 1: Nine microsatellites showing the strongest association between heterozygosity and mortality due to invasive bacterial disease.

Marker	OR	lower 95% CI	upper 95% CI	chi sq	p value
D12S310	2.86	1.68	4.86	15.420	0.000086
D12S352	2.66	1.64	4.31	16.194	0.000057
D13158	2.67	1.66	4.3	16.496	0.000049
D14S261	3.14	1.94	5.09	21.998	0.000003
D14S275	2.59	1.59	4.22	15.034	0.000106
D14S280	3.57	2.05	6.21	21.421	0.000004
D15S1007	2.64	1.63	4.28	16.093	0.000060
D16S423	2.60	1.6	4.21	15.368	0.000088
D16S3103	2.89	1.75	4.77	17.611	0.000027

The table shows odds ratios (OR), adjusted for age in 6 strata and geographical location in 8 strata, of death due to bacterial invasive disease, along with lower and upper 95% confidence intervals (CI). Significance was assessed using 2 × 2 contingency table of heterozygote/homozygote vs case/control, yielding a chi-squared value (chi sq) and its associated P value. All loci comfortably exceed the experiment-wide significance threshold of 0.00037.

mortality due to bacterial infection that would result if homozygosity at the locus could be eliminated. However, such direct interpretation in our case is problematic for many reasons, including the fact that the risk is probably driven by specific alleles whose prevalence differs greatly from that of homozygotes in general. None the less, the ORs of the full model indicate that the risks we describe are sizeable, particularly since the markers provide only indirect measures of homozygosity at the genes themselves. Furthermore, given that the population prevalence of homozygosity at the relevant markers is high, the population-wide effects of homozygosity are likely to impact

Table 2: Age- and geographic location-adjusted odds ratios for invasive bacterial death with homozygosity at specific microsatellite markers in multivariable models restricted to cases of Gram positive sepsis, gram negative sepsis or including all invasive bacterial deaths combined.

Markers	Interaction	Gram Negative OR	Gram Negative 95% CI	Gram Positive OR	Gram Positive 95%CI	Combined OR	Combined 95% CI
D7S486		1.02	0.16 – 6.53				
D7S486	D16S423	40.7	4.28 – 387				
D12S310		14.0	2.70 – 72.7	4.73	1.59–14.1	4.94	2.27 – 10.8
D13S158		6.11	1.45 – 25.8			4.66	1.92 – 11.3
D15S1007		7.28	1.89 – 28.1				
D16S423		1.61	0.31 – 8.33			7.65	2.55 – 22.9
D9S164	infants			0.18	0.018 – 1.90		
D9S164	children			1.65	0.249 – 10.9		
D14S275				3.93	1.35 – 11.4	12.2	4.44 – 33.3
D14S275	D13S158					10.1	3.84 – 26.5
D16S3103				3.70	1.20 – 11.4	7.04	2.56 – 19.4
D16S3103	D16S423					10.2	4.10 – 25.2

The table shows odds ratios, adjusted for age in 6 strata and for geographical location in 8 strata, of death due to all invasive bacterial disease, of death due to Gram negative invasive bacterial disease and of death due to Gram positive invasive bacterial disease, for homozygosity at microsatellite markers either alone or paired (where interactions were noted with LRT p values < 0.05). Variables excluded, with LRT p values = 0.05, in addition to sex and SOH, were markers d12s352, d14s261, d14s275, d16s3103 and d18s452 in the gram negative model, markers d13s158, d14s261, d14s280 and d16s423 in the gram positive model and markers d12s352, d14s261, d14s280 and d15s1007 in the combined model. Variables and interaction terms in the final models all had LRT p values < 0.02.

Table 3: Population attributable risk fractions (PARF) for homozygosity at five microsatellite markers in a final multivariable model of bacterial diseases death.

Microsatellite marker	population prevalence of homozygosity	OR for bacterial disease death	PARF
D12S310	0.187	3.65	0.331
D13S158	0.387	2.04	0.287
D14S275	0.282	4.68	0.509
D16S3103	0.490	3.52	0.553
D16S423	0.418	2.75	0.423

PARFs were calculated using all markers retained in the final multivariable logistic regression model but, for simplicity, without interaction effects. For this reason the Odds Ratios quoted here are not identical to those in table 2.

very considerable on the total burden of invasive bacterial disease mortality, with the majority of deaths being genetically determined.

Discussion

Here we conduct what we believe is the first systematic analysis of the association between heterozygosity and infectious diseases in humans. Although cases exhibit generally increased homozygosity relative to controls, more detailed analysis indicates that this is largely due to a small subset of markers, each of which contributes a significant risk factor when homozygous. We conclude that heterozygosity at a minimum of five loci contributes ORs of up to 40, and that the most important loci vary depending on the type of pathogen.

There is currently a debate as to whether the benefits of heterozygosity accrue mainly through genome-wide effects (inbreeding) [38,39] or through individual balanced polymorphisms [13,14]. We found that inbreeding effects are either small or absent in this population. This is perhaps not surprising because, in contrast to some other populations such the Fulani [40] and some Arab communities [19], consanguineous marriages tend to be discouraged, with a preference for marriages between rather than within clans [41]. In contrast, five loci independently contribute significant risk factors, lending strong support to the local effects model. However, it should be remembered that human populations differ greatly in their structure and that, in contrast to most animals populations, some human populations actually favour consanguineous marriages [19,40,42]. In such populations a rather different pattern may well emerge.

To find several balanced polymorphisms in a relatively small study of just 134 markers is surprising, given how few have been identified previously in humans [22]. Two factors may contribute to this discrepancy. First, a large majority of genome scans focus on complex, non-infectious diseases, and these are likely to differ from infectious diseases mechanistically. Most heritable non-infectious diseases involve mutant alleles at one or more loci where function is removed or disrupted, and hence are mostly

recessive. In contrast, the efficacy of immune-function genes is widely though to benefit from high diversity, a larger palette of alleles increasing the range of pathogen types that can be recognised, and therefore these loci tend naturally towards heterosis. Second, classical association studies tend to be applied to diseases that are known to run in families [43,44], and hence susceptibility will tend to have an appreciable additive component. As such, patterns where heterozygosity is important will tend to be overlooked because heterozygosity *per se* tends not to be heritable. Instead there is a strong focus on searching for associations between particular alleles and disease [43,45,46]. It will be interesting to see the extent to which future studies reveal a much higher prevalence of balancing selection, thereby supporting results from many non-human systems.

Our current study is relatively small-scale, with several of the smaller chromosomes being scored for only three or four markers. Consequently, there are large tracts of the genome where further loci could be located with the potential to contribute even further to genetic susceptibility, and implying that the five regions we identify are not the complete set of the loci that could potentially be identified in a larger study. This is surprising because the loci we have uncovered exhibit large individual and combined effect sizes, to the extent that mortality appears highly non-random. Moreover, it should be remembered that the overall risk factor combines both genetic susceptibility and variation in exposure. Unless exposure to pathogens is highly uniform, the impact of genetic factors will be even higher than we report and could rise further if our study has missed further contributory loci.

The effect sizes we report appear much larger than expected. Across the five loci identified as having highest impact, population attributable risk fractions (PARFs) all lie in the range 25–55%. PARFs provide an indication of the proportion of total risk that can be attributed to each genetic factor, given the local prevalence of exposure. Since the calculations assume overlapping effects, these do not sum to one. None the less, our analysis suggests that half or more of the observed deaths would probably

not have occurred if the individuals concerned had been heterozygous for these loci, a figure that would surely be even higher if we had been able to genotype SNPs in the genes concerned rather than at linked microsatellites.

The idea that pathogens could play a major role in driving balancing selection at many different locations across the genome is reinforced by the difference we found between Gram negative and Gram positive bacteria. Immune defense mechanisms against Gram positive and Gram negative pathogens vary significantly [47,48], and while there may be some degree of overlap in genetic regulation of immunity to different classes of pathogens, the difference we find between Gram negative and Gram positive strains would help to explain why so many different regions appear to be involved.

Conclusion

We believe our study is the first to apply to humans the sorts of analysis that commonly reveal single locus heterosis maintained by pathogens in natural populations. We reveal several discrete genomic locations where heterozygosity confers some degree of protection from lethal bacterial infection. Together these loci contribute a substantial risk factor that makes mortality from infection highly non-random. Our study has obvious implications for epidemiology and could lead to the development of simple tests for individuals who are most at risk from infection. High density SNP mapping is under way in order to identify relevant genes.

Methods

Meningitis is defined by a positive cerebrospinal fluid culture. Neonatal sepsis is defined as bacteraemia or meningitis from day 0 to 59 of life. Malaria parasitemia was concurrently present in some cases and these are analysed as a separate class because malaria may have contributed to mortality.

Control selection

Controls were selected at random from among a set of healthy subjects who had originally been selected from the community living near a case using the "spinning pencil" technique and individually matched to cases on age, sex and date of presentation to hospital in a case-control study of both surviving and fatal cases of bacteraemia. For ethical reasons, no controls were recruited among young infants (age <60 days). Cases and controls were restricted to the Mijikenda ethnic group indigenous to Coastal Kenya. The subset of controls selected for the present study was frequency-matched on age to cases in the present study. In all multi-variable logistic regression models age and administrative location of residence were included. Age was specified in six strata (0–5 m, 6–11 m. 12–23 m, 24–35 m, 36–59 m, 60–151 m) each of which

contained between 13–19% of the observations. To control for ethnic diversity we stratified by administrative authority, the best form of 'address' we could obtain, yielding eight geographical locations each of which contained between 4–26% of the data. These partitions allow for some degree of geographic substructure and correspond loosely with seven long established sub-groups of the Mijikenda ethnic group, each of which has a different language, and who tend to live in geographically defined clusters.

Standardized Observed Homozygosity

$$S_{OH} = \frac{N_{\text{hom}}}{\sum\limits_{i} H_{oi}}$$

S_{OH} is the standardized observed homozygosity for an individual genotyped for i loci. N_{hom} is the number of homozygote genotypes in the individual concerned and H_{oi} is the observed frequency of homozygotes the ith locus scored in this individual, calculated across the full sample set.

Population Attributable Risk Fractions

The PARFs were estimated as prev(OR-1)/(1+prev(OR-1) for each marker in the final model of all invasive bacterial disease deaths combined but, for simplicity, excluding the interaction terms. The prevalence of homozygosity in the population was estimated in the control population after standardizing on age to the known age-distribution of the population around the hospital. This was provided by the Kilifi Demographic Surveillance Study, which has conducted 2–3 household visits each year to enumerate the population in an area accommodating 230,000 people living closest to the hospital since 2000.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

EL conducted the genotyping, participated in the analysis and helped write the paper. WA helped conceive the study, led the analysis and wrote the paper. JB, IM, KM, NP and AS conducted the case control study and defined the clinical syndromes, JB and AS participated in the study design and interpretation of clinical data and AS further conducted the statistical analysis and helped write the paper. MS TRW CRN participated in study coordination, sample acquisition and processing, and interpretation of the data. AH initiated the genetic programme and directed the Oxford research activities.

Additional material

Additional file 1

Names and chromosomal assignments of all microsatellites genotyped in this study.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2350-10-55-S1.doc]

Acknowledgements

References

1. David P: **Heterozygosity-fitness correlations: new perspectives on old problems.** *Heredity* 1998, **80**:531-537.
2. Coltman DW, Slate J: **Microsatellite measures of inbreeding: a meta-analysis.** *Evolution* 2003, **57**:971-983.
3. Coulson TN, Pemberton JM, Albon SD, Beaumont M, Marshall TC, Slate J, Guiness FE, Clutton-Brock TH: **Microsatellites reveal heterosis in red deer.** *Proc R Soc Lond B* 1998, **265**:489-495.
4. Hoffman JI, Boyd IL, Amos W: **Exploring the relationship between parental relatedness and male reproductive success in the Antarctic fur seal *Arctocephalus gazella*.** *Evolution* 2004, **58**:2087-2099.
5. Slate J, Kruuk LEB, Marshall TC, Pemberton JM, Clutton-Brock TH: **Inbreeding depression influences lifetime breeding success in a wild population of red deer (*Cervus elaphus*).** *Proc Roy Soc Lond B* 2000, **267**:1657-1662.
6. Amos W, Worthington Wilmer J, Fullard K, Burg TM, Croxall JP, Bloch D, Coulson T: **The influence of paternal relatedness on reproductive success.** *Proc Roy Soc Lond B* 2001, **268**:2021-2027.
7. Garamszegi LZ, Møller AP, Erritzoe J: **The evolution of immune defense and song complexity in birds.** *Evolution* 2003, **57**:905-912.
8. Seddon N, Amos W, Tobias JA: **Heterozygosity predicts territory size and song structure in a co-operatively breeding bird.** *Proc Roy Soc Lond B* 2004, **271**:1823-1829.
9. Reid JM, Arcese P, Keller LF: **Inbreeding depresses immune response in song sparrows (*Melospiza melodia*): direct and inter-generational effects.** *Proc R Soc Lond B* 2003, **270**:2151-2157.
10. Rijks J, Hoffman JI, Kuiken T, Osterhaus ADME, Amos W: **Heterozygosity and lungworm burden in harbour seals (*Phoca vitulina*).** *Heredity* 2008, **100**:587-593.
11. Coltman DW, Pilkington JG, Smith JA, Pemberton JM: **Parasite-mediated selection against inbred Soay sheep in a free-living island population.** *Evolution* 1999, **53**:1259-1267.
12. Acevedo-Whitehouse K, Gulland F, Greig D, Amos W: **Disease susceptibility in California sea lions.** *Nature* 2003, **422**:35.
13. Acevedo-Whitehouse K, Spraker TR, Lyons E, Melin SR, Gulland F, DeLong RL, Amos W: **Contrasting effects of heterozygosity on survival and hookworm resistance in California sealion pups.** *Mol Ecol* 2006, **15**:1973-1982.
14. Hansson B, Westerberg L: **On the correlation between heterozygosity and fitness in natural populations.** *Mol Ecol* 2002, **11**:2467-2474.
15. Pemberton JM: **Measuring inbreeding depression in the wild: the old ways are the best.** *Trends Ecol Evol* 2004, **19**:613-615.
16. Balloux F, Amos W, Coulson TN: **Does heterozygosity estimate inbreeding in real populations?** *Mol Ecol* 2004, **13**:3021-3031.
17. Slate J, David P, Dodds KG, Veenvliet BA, Glass BC, Broad TE, McEwan JC: **Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data.** *Heredity* 2004, **93**:255-265.
18. DeWoody YD, DeWoody JA: **On the estimation of genome-wide heterozygosity using molecular markers.** *J Hered* 2005, **96**:85-88.
19. Jaber L, Shohat T, Rotter JI, Shohat M: **Consanguinity and common adult diseases in Israeli Arab communities.** *Am J Medic Genet* 1997, **70**:346-348.
20. Becker S, Al Halees Z, Molina C, Paterson RM: **Consanguinity and congenital heart disease in Saudi Arabia.** *Am J Medic Genet* 2001, **99**:8-13.
21. Roberts DF: **Consanguinity and multiple sclerosis in Orkney.** *Genet Epidem* 1991, **8**:147-151.
22. Bubb KL, Bovee D, Buckley D, Haugen E, Kibukawa M, Paddock M, Palmieri A, Subramanian S, Zhou Y, Kaul R, *et al.*: **Scan of human genome reveals no new loci under ancient balancing selection.** *Genetics* 2006, **173**:2165-2177.
23. Pasvol G, Weatherall DJ, Wilson RJ: **Cellular mechanism for the protective effect of haemoglobin S against P. falciparum malaria.** *Nature* 1978, **274**:701-703.
24. Doherty PC, Zinkernagel RM: **Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex.** *Nature* 1975, **256**:50-52.
25. Penn DJaP, WK : **The evolution of mating preferences and major histocompatibility complex genes.** *Am Nat* 1999, **153**:145-164.
26. Penn DJ, Damjanovich K, Potts WK: **MHC heterozygosity confers a selective advantage against multiple-strain infections.** *Proc Natl Acad Sci USA* 2002, **99**:11260-11264.
27. van Oosterhout C: **A new theory of MHC evolution: beyond selection on the immune genes.** *Proc Roy Soc Lond B* 2009, **276**:657-665.
28. Hoffman JI, Amos W, Trathan PN, Forcada JP: **Female fur seals show active choice for males who are heterozygous and unrelated.** *Nature* 2007, **445**:912-914.
29. Bierne N, Launey S, Naciri-Graven Y, Bonhomme F: **Early effect of inbreeding as revealed by microsatellite analyses on *Ostrea edulis* larvae.** *Genetics* 1998, **148**:1893-1906.
30. Hansson B, Bensch S, Hasselquist D, Åkesson M: **Microsatellite diversity predicts recruitment of sibling great reed warblers.** *Proc R Soc Lond B* 2001, **268**:1287-1291.
31. Hollox EJ, Armour JAL: **Directional and balancing selection in human beta-defensins.** *BMC Evol Biol* 2008, **8**:113.
32. Akey JM, Zhang G, Zhang K, Jin L, Shriver M: **Interrogating a high density SNP map for signatures of natural selection.** *Genomes Res* 2002, **12**:1805-1814.
33. Berkley JA, Lowe BS, Mwangi I, Williams T, Bauni E, Mwarumba S, Ngetsa C, Slack MP, Njenga S, Hart CA, *et al.*: **Bacteremia among children admitted to a rural hospital in Kenya.** *N Engl J Med* 2005, **352**:39-47.
34. Aparicio JM, Ortego J, Cordero PJ: **What should we weigh to estimate heterozygosity, alleles or loci?** *Mol Ecol* 2006, **15**:4659-4665.
35. Amos W, Hoffman JI, Frodsham AJ, Zhang L, Best S, Hill AVS: **Automated binning of microsatellite alleles: problems and solutions.** *Mol Ecol Notes* 2007, **7**:10-14.
36. Benjamini Y, Hochberg Y: **Controlling the false discovery rate – a practical and powerful approach to multiple testing.** *J Roy Stat Soc B* 1995, **57**:289-300.
37. Takei N, Mortensen PB, Klaening U, Murray RM, Sham PC, O'Callaghan E, Munk-Jørgensen P: **Relationship between in utero exposure to influenza epidemics and risk of schizophrenia in Denmark.** *Biol Psychiatry* 1996, **40**:817-824.
38. Ross-Gillespie A, O'Riain MJ, Keller LF: **Viral epizootic reveals inbreeding depression in a habitually inbreeding mammal.** *Evolution* 2007, **61**:2268-2273.
39. Spielman D, Brook BW, Briscoe DA, Frankham R: **Does inbreeding and loss of genetic diversity decrease disease resistance?** *Cons genet* 2004, **5**:439-448.
40. Hampshire KR, Smith MT: **Consanguineous marriages among the Fulani.** *Hum Biol* 2001, **73**:597-603.
41. Parkin DJ: *The sacred void: spatial images of work and ritual among the Giriama of Kenya* Cambridge: Cambridge University Press; 1991.
42. Abdulrazzaq YM, Bener A, Al-Gazali LI, Al-Khayat AI, Micallef R, Gaber T: **A study of possible deleterious effects of consanguinity.** *Clin Genet* 1997, **51**:167-173.
43. Schulze TG, McMahon FJ: **Genetic association mapping at the crossroads: which test and why? Overview and practical guidelines.** *Am J Medic Genet (Neuropsych Genet)* 2002, **114**:1-11.

44. Kimmel G, Shamir R: **A fast method for computing high-significance disease association in large population-based studies.** *Am J Hum Genet* 2006, **79**:481-492.
45. Ohashi J, Yamamoto S, Tsuchiya N, Hatta Y, Komata T, Matsushita M, Tokunaga K: **Comparison of statistical power between 2 × 2 allele frequency and allele positivity tables in case-control studies of complex disease genes.** *Ann Hum Genet* 2001, **65**:197-206.
46. Collins A, Morton NE: **Mapping a disease locus by allelic association.** *Proc Natl Acad Sci USA* 1998, **95**:1741-1745.
47. Wurfel MM, Gordon AC, Holden TD, Radella F, Strout J, Kajikawa O, Ruzinski JT, Rona G, Black RA, Stratton S, *et al.*: **Toll-like receptor i polymorphisms affect innate immune responses and outcomes of sepsis.** *Am J Respir Crit Care Med* 2008, **178**:710-720.
48. Puliti M, Uematsu S, Akira S, Bistoni F, Tissi L: **Toll-like receptor 2 deficiency is associated with enhanced severity of group B streptococcal disease.** *Infection Immun* 2009, **77**:1524-1531.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2350/10/55/prepub