Vol 460 | 16 July 2009 | doi:10.1038/nature08149

LETTERS

Evolution of a malaria resistance gene in wild primates

Jenny Tung[1,2], Alexander Primus[1,2], Andrew J. Bouley[1], Tonya F. Severson[2], Susan C. Alberts[1,2,3,4]* & Gregory A. Wray[1,2,3]*

The ecology, behaviour and genetics of our closest living relatives, the nonhuman primates, should help us to understand the evolution of our own lineage. Although a large amount of data has been amassed on primate ecology and behaviour, much less is known about the functional and evolutionary genetic aspects of primate biology, especially in wild primates. As a result, even in well-studied populations in which nongenetic factors that influence adaptively important characteristics have been identified, we have almost no understanding of the underlying genetic basis for such traits. Here, we report on the functional consequences of genetic variation at the malaria-related *FY* (*DARC*) gene in a well-studied population of yellow baboons (*Papio cynocephalus*) living in Amboseli National Park in Kenya. *FY* codes for a chemokine receptor normally expressed on the erythrocyte surface that is the known entry point for the malarial parasite *Plasmodium vivax*[1–3]. We identified variation in the *cis*-regulatory region of the baboon *FY* gene that was associated with phenotypic variation in susceptibility to *Hepatocystis*, a malaria-like pathogen that is common in baboons[4,5]. Genetic variation in this region also influenced gene expression *in vivo* in wild individuals, a result we confirmed using *in vitro* reporter gene assays. The patterns of genetic variation in and around this locus were also suggestive of non-neutral evolution, raising the possibility that the evolution of the *FY cis*-regulatory region in baboons has exhibited both mechanistic and selective parallels with the homologous region in humans[6–8]. Together, our results represent the first reported association and functional characterization linking genetic variation and a complex trait in a natural population of nonhuman primates.

In humans, a transition from the wild-type T variant to a C variant at a single polymorphic site in the *FY cis*-regulatory region causally abolishes all expression of this gene in erythrocytic precursors. As a result, C homozygotes at this site are strongly protected from infection by *P. vivax*[9], and a lower level of protection is also conferred on C/T heterozygotes[10,11]. The C variant has apparently arisen independently at least twice in geographically distinct human populations (in Africa and in Papua New Guinea[2,11]), and has been driven to high frequencies on at least two haplotypic backgrounds within Africa[6]. Additionally, the pattern of variation in the *cis*-regulatory region as a whole strongly indicates a historical pattern of natural selection in different populations around the world, probably as the product of directional selection in some populations (for example, local positive selection), and a complex mix of selection and demographic history in others[6–8]. The unusual evolutionary history of this locus led us to investigate the pattern of genetic variation in its baboon homologue, and to explore the possibility that it might also explain phenotypic variation in parasite infection in a wild primate population, the well-studied baboon population of the Amboseli basin in East Africa[12–14].

Baboons are not generally infected by *Plasmodium* in the wild, but are vulnerable to infection by several closely related haematoprotozoans[4,5] including *Hepatocystis kochi*, a blood parasite nested within the paraphyletic *Plasmodium* genus[15]. *Hepatocystis* parasites do not produce the cyclical fever spikes typical of malaria in humans, but do produce anaemia and visible merocyst formation, followed by scarring on the liver[4].

We tested for the presence of *Hepatocystis* parasites by screening DNA samples extracted from baboon blood for 190 individuals in the Amboseli baboon population. We found a high incidence of *Hepatocystis* infection in the Amboseli population (61.9%), although rates of infection varied substantially between different social groups and over time, possibly because of differences in home range and hence exposure to the vector, a biting midge[16].

In 174 of 190 baboons that we screened for *Hepatocystis*, we also sequenced the region of baboon DNA homologous to the annotated human *FY cis*-regulatory region. We identified six single nucleotide polymorphisms (SNPs) in the baboon *FY cis*-regulatory region (Fig. 1 and Supplementary Fig. 1; the malaria-associated SNP documented in humans was invariant in the baboons). *Hepatocystis* infection was significantly associated with an A/G variable site in the *FY cis*-regulatory region, in a model that took social group (a significant source of variance in infection) and genetic background into account. The risk of infection decreased as the number of G alleles an individual carried increased ($P < 0.012$, $n = 174$; Fig. 2).

We also investigated whether *FY cis*-regulatory variation in baboons causally influences gene expression, as the C/T variant does in humans. We collected 101 samples of RNA-preserved blood from adults in six baboon social groups between 2004 and 2008, and used

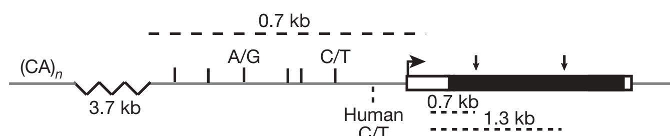

Figure 1 | Schematic of the baboon *FY* gene (not to scale). Boxed regions are regions of the gene present in mature messenger RNA (open boxes are untranslated regions; black boxes are protein coding sequence). Grey lines indicate untranscribed regions. The bent arrow indicates the start of transcription. The downward arrows show two baboon SNPs used as markers for the pyrosequencing assays. The vertical black bars mark *cis*-regulatory SNPs in baboons, with the *Hepatocystis*-associated SNP labelled A/G and the allelic-imbalance-associated SNP labelled C/T. The dashed vertical bar marks the location of the functional SNP known in humans. The dashed horizontal lines provide relative distances and/or sizes of features. The location of the $(CA)_n$ microsatellite used in the F_{st} analyses is also shown upstream.

[1]Department of Biology, Duke University, [2]Institute for Genome Sciences and Policy, Duke University, [3]Department of Evolutionary Anthropology, Duke University, Durham, North Carolina 27708, USA. [4]Institute of Primate Research, National Museums of Kenya, Nairobi, Kenya.
*These authors contributed equally to this work.

©2009 Macmillan Publishers Limited. All rights reserved

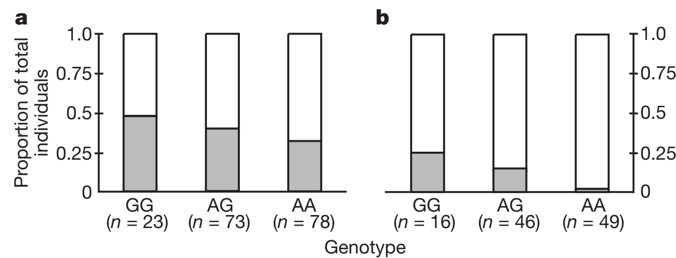

Figure 2 | Genotype at the *FY cis*-regulatory A/G SNP is associated with *Hepatocystis* infection. The proportion of uninfected individuals is shown in grey, and the proportion of infected individuals is shown in white. **a**, Results for the entire sample set ($n = 174$; $P < 0.012$); **b**, results only for members of the six groups with high prevalence (>75%) of *Hepatocystis* infection ($n = 111$; $P < 0.004$). Numbers below each genotype show the number of individuals for the given genotype.

these samples to measure allele-specific expression at the *FY* locus using pyrosequencing. Specifically, we investigated whether the level of *FY* expression driven by one *cis*-regulatory *FY* allele differed from the level of *FY* expression driven by the other *cis*-regulatory *FY* allele, within the same individual. Because allele-specific expression compares the relative amounts of gene expression within individuals, it controls for effects on gene expression operating in *trans*, such as those produced by genetic background or by environmental main effects[17,18]. If the two alleles within an individual drive expression differently (allelic imbalance), that individual is likely to harbour a functional *cis*-regulatory variant that influences gene expression.

We measured allele-specific expression in 38 individuals (all the individuals among the 101 RNA-sampled baboons that were heterozygous at a transcribed pyrosequencing assay SNP: see Fig. 1). Average $\log_2$ fold-change differences in expression between alleles within heterozygous individuals ranged from -0.002 (no difference between alleles) to 2.13 (substantial difference between alleles). This suggested that one or more common functional *cis*-regulatory variants influenced expression of the baboon *FY* gene in the Amboseli population. We predicted that if a *cis*-regulatory variant contributes to variation in gene expression, then individuals heterozygous at the *cis*-regulatory site would show significantly higher levels of allelic imbalance than individuals homozygous for the same variant.

Genotypes at four of the six SNPs in the baboon *FY cis*-regulatory region were sufficiently variable to test for an association with allelic imbalance. Genotype at the SNP closest to the start of transcription, a C/T transition, was significantly associated with allelic imbalance in the predicted direction: heterozygotes exhibited higher levels of allelic imbalance than homozygotes ($P < 0.002$, $n = 38$; Fig. 3). However, this site explained only 22.0% of the overall variance in the allelic imbalance samples, after taking into account year of sampling. These results suggested that this C/T SNP functionally influences gene expression of the *FY* gene within the baboon population, but that additional *cis*-regulatory variants probably also play a part.

We next investigated the A/G SNP in the *FY cis*-regulatory region that we had associated with *Hepatocystis* infection risk. Of the 38 individuals for whom we measured allele-specific expression, 37 were heterozygous for this A/G SNP. Hence, we could not compare allelic imbalance levels between heterozygotes and homozygotes at this site. We therefore tested this variant for possible functional effects using an *in vitro* approach in cell culture. We also used this framework to test further whether the C/T variant that was associated with allelic imbalance causally influenced *FY* expression. For each of these two regulatory SNPs (A/G and C/T), we built two plasmid constructs consisting of the *FY cis*-regulatory region linked to the firefly *luciferase* gene, such that the two constructs differed only at the variable site. We then tested the ability of these constructs to drive gene expression in a human erythroleukaemic (HEL) cell line. For the C/T SNP, the T allele construct drove significantly higher levels of expression than

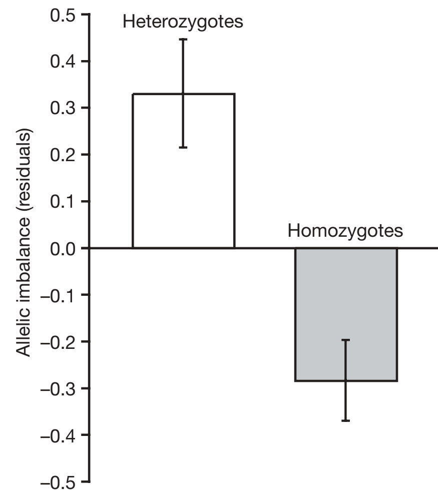

Figure 3 | Allelic imbalance is associated with *FY cis*-regulatory genotype. Height of the bars shows the mean for each genotypic class (heterozygotes: 0.329 ± 0.116 s.e.m.; homozygotes: -0.285 ± 0.086 s.e.m.). The *y* axis gives the residuals of $\log_2$-transformed allelic imbalance on year of sampling ($n = 38$). Heterozygotes at the C/T SNP exhibit high values of allelic imbalance relative to homozygotes at the C/T SNP after controlling for the effect of year of sampling. High values indicate that the two alleles of the *FY* gene are transcribed at different levels within individuals, and suggest a functional *cis*-regulatory role for this SNP.

the alternative C allele construct ($P < 0.0001$; Fig. 4a). Similarly, for the A/G SNP, the G allele construct drove significantly higher levels of expression than the alternative A allele construct ($P < 0.0001$; Fig. 4b). These results suggest that both SNPs have the capacity to drive differential expression of the *FY* gene. In the case of the C/T SNP, for which both *in vivo* and *in vitro* analyses were possible, the T allele was associated with higher levels of expression in both experiments. Unlike the human case, in which one regulatory SNP results in null expression, all the baboon haplotypes we tested drove robust expression of the gene.

Together, these data indicate that the *FY cis*-regulatory region is associated with parasite infection in a wild population of baboons, and that functional sequence variants within this region causally influence the level of expression of the *FY* gene. As in humans, variation in gene expression at the *FY* locus may therefore be important in parasite susceptibility, either through altering the direct access of *Hepatocystis* to baboon erythrocytes or, as has recently been demonstrated in humans, by altering a more general property of the immune system, such as relative white blood cell counts[19].

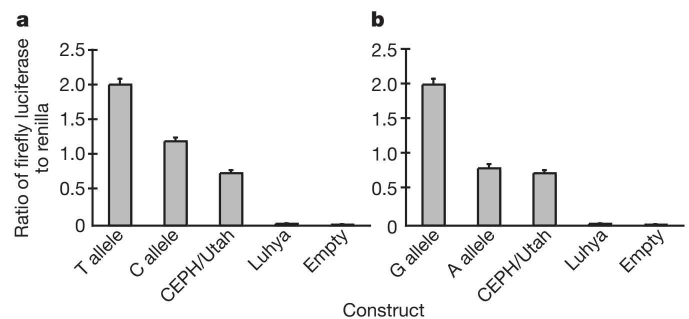

Figure 4 | *FY cis*-regulatory variation drives differential expression *in vitro*. **a**, The C/T SNP identified through the allelic imbalance measurements. **b**, The A/G SNP that associates with *Hepatocystis* infection drives differential gene expression compared to the alternative allele of the same SNP in cell culture. Values on the *y* axis give the relative ratio of firefly luciferase luminescence to a control renilla luciferase reporter. Human constructs from a normal-expressing individual (haplotype from the CEPH/Utah HapMap panel) and a null-expressing individual (Luhya) and an empty vector are shown for comparison. Error bars show s.e.m. for each construct.

©2009 Macmillan Publishers Limited. All rights reserved

These results suggest that the genetic basis of phenotypic variation in different primate species can exhibit a remarkable degree of parallelism. In this case, not only are these similarities present on the molecular level or on the level of trait association, as shown by previous work[20,21], but they also extend to the mechanism that links molecular and phenotypic variation (which is probably gene expression). Given these parallelisms, baboons, like humans, may also exhibit evidence of non-neutral evolution at the *FY cis*-regulatory region. We detected an increased level of population differentiation among East African baboon populations around *FY*, by comparing a *FY*-linked microsatellite with 35 neutral microsatellites ($F_{st} = 0.31$, $P < 0.029$; range of F_{st} for the neutral markers was 0.008–0.346; F_{st} is a metric describing genetic divergence between populations based on allele frequency differences at variable sites; Supplementary Fig. 3). We also detected a higher value for the Tajima's D statistic ($D = 1.26$) in this region relative to nine of nine other resequenced putative *cis*-regulatory regions in the Amboseli population and 11 of 12 resequenced transcribed regions (range of D for all other loci was -1.60 to 2.12). The only locus with a higher value of D, a transcribed portion of the gene *MSR1*, exhibited an even more extreme value than that identified for the *MHC DQA1* promoter in baboons[22], which is known to evolve under strong *trans*-specific balancing selection[20].

Interestingly, a sliding window analysis showed that the peak values of D corresponded well with the *Hepatocystis* and allelic-imbalance-associated SNPs (Supplementary Fig. 3). Given that rates of *Hepatocystis* infection appear to vary across different populations (30% in the Masai Mara Reserve, Kenya, $n = 10$; 90% in Mikumi National Park, Tanzania, $n = 20$; see Supplementary Tables 1 and 2), these results suggest that the baboon *FY* cis-regulatory region may be subject to a complex selective history similar to the case described in humans[7,23] in which differing levels of pathogen pressure across populations are associated with high levels of population differentiation around the *FY* gene, and varying signatures of selection within populations.

In spite of the parallels that we have documented for baboon and human *FY*, the functional variants we have identified in baboons are not homologous to the known functional variant in humans, which reveals that phenotypic variation in different primate species may show similar, but not precisely convergent, patterns of evolution. Indeed, while in humans the *FY*–malaria relationship is Mendelian, both *FY* expression and infection by *Hepatocystis* in baboons are clearly complex traits: even individuals homozygous for the *Hepatocystis* 'resistance' variant (the G allele at the A/G SNP) suffer from parasitism, albeit at a lower rate (52.2% of GG homozygotes were infected, versus 67.9% of AA homozygotes, across all study groups: see Fig. 2 and Supplementary Fig. 2). Additionally, the *in vitro* cell culture experiments suggest that the G allele of this variant actually drives higher expression of *FY* than the alternative A allele, even though the G allele is associated with a lower risk of *Hepatocystis* infection. The relationship between *FY* gene expression and *Hepatocystis* in baboons is therefore clearly different from that in humans, perhaps owing to balancing the cost of infection by other blood parasites, some of which are not known to co-occur with, and might be excluded by, *Hepatocystis*[24]. Alternatively, while the *in vitro* data on the A/G variant strongly suggest that this site has the capacity to influence *FY* gene expression, the direction and magnitude of its effects may differ in its natural cellular context.

This possibility is supported by the differences in magnitude of the effect of the C/T *cis*-regulatory variant in the *in vitro* transfection assays and the *in vivo* allelic imbalance measurements. *In vivo* gene expression measurements are complicated by variation in genetic background and in the environment, both of which can modify functional *cis*-regulatory effects[25,26]. Indeed, our results show that even baboons that are homozygotes at the C/T site sometimes exhibit allelic imbalance in *FY* expression, suggesting that other, unidentified functional *cis*-regulatory variants are also segregating in the population. In contrast, in the *in vitro* comparisons, only a single *cis*-regulatory site differed between the experimental constructs, thus controlling for both environment and genetic backgrounds. Using both approaches in tandem can be synergistic: while *in vitro* experiments can help pin down specific functional sites, *in vivo* results demonstrate that these effects are relevant to the biology of individuals in the wild.

Thus, although identifying the genetic basis for phenotypic variation in wild primates poses substantial challenges, we present this study as a model to motivate additional evolutionary genetic research on natural primate populations. This work is essential if we hope to integrate an evolutionary and functional genetic perspective into the rich tradition of organismal research on these species. Our results demonstrate that patterns of variation in nonhuman primates can provide unique insights into the influence of ecological and environmental factors on genetic and trait variation in humans. Integrative research on nonhuman primates should also help us develop a better understanding of the evolution of our own species.

METHODS SUMMARY

We screened for *Hepatocystis* in 190 baboons, using *Hepatocystis* mitochondrial DNA specific primers. We sequenced the region homologous to the human *FY cis*-regulatory region in 174 of these baboons. We identified predictors of *Hepatocystis* infection in these 174 individuals using a generalized linear mixed model with a binomial error structure, fitting genotype and an estimate of genetic background as fixed effects and study group as a random effect.

Allelic imbalance was assessed via pyrosequencing using RNA collected from the study population, and was measured as the $\log_2$ transformed ratio of the expression of one allele versus the other allele, within individuals. We modelled allelic imbalance using a general linear mixed model and evaluated the significance of the genotype effect on allelic imbalance using a permutation test.

We co-transfected human erythroleukaemic cells with experimental constructs or a control construct and a normalization construct. A total of 24 measurements were made per construct in three independent experiments. We compared the measurements for each pair of constructs separately using general linear mixed models.

For the F_{st} comparison, we genotyped up to 36 polymorphic microsatellite loci (35 neutral markers plus one located near the *FY cis*-regulatory region[23]) in ten baboons from the Masai Mara Reserve, Kenya, 20 baboons from Mikumi National Park, Tanzania, and 12 baboons from the Amboseli basin, Kenya (Supplementary Information). We compared the F_{st} value for the *FY*-linked microsatellite locus with those for the 35 neutral loci. The probability of observing a more extreme level of genetic differentiation at the *FY*-linked locus was assessed by fitting a gamma distribution to the F_{st} values for the neutral loci, and calculating the average value of P based on different parameters of the gamma (weighted by the probability of each parameter set, given the data). Tajima's D values were calculated for samples from Amboseli using DnaSP version 4.9 (ref. 27).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 March; accepted 15 May 2009.
Published online 24 June 2009.

1. Barnwell, J. W., Nichols, M. E. & Rubinstein, P. *In vitro* evaluation of the role of the Duffy blood group in erythrocyte invasion by *Plasmodium vivax*. *J. Exp. Med.* **169**, 1795–1802 (1989).
2. Miller, L. H., Mason, S. J., Clyde, D. F. & Mcginniss, M. H. Resistance factor to *Plasmodium vivax* in blacks—Duffy blood group genotype *Fyfy*. *New Engl. J. Med.* **295**, 302–304 (1976).
3. Miller, L. H., Mason, S. J., Dvorak, J. A., Mcginniss, M. H. & Rothman, I. K. Erythrocyte receptors for *Plasmodium knowlesi* malaria—Duffy blood group determinants. *Science* **189**, 561–563 (1975).
4. Garnham, P. C. C. *Malaria Parasites and Other Haemosporidia* (Blackwell, 1966).
5. Myers, B. J. & Kuntz, R. E. A checklist of parasites reported for the baboon. *Primates* **6**, 137–194 (1965).
6. Hamblin, M. T. & Di Rienzo, A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am. J. Hum. Genet.* **66**, 1669–1679 (2000).
7. Hamblin, M. T., Thompson, E. E. & Di Rienzo, A. Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet.* **70**, 369–383 (2002).
8. Sabeti, P. C. *et al.* Positive natural selection in the human lineage. *Science* **312**, 1614–1620 (2006).
9. Tournamille, C., Colin, Y., Cartron, J. P. & Levankim, C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy negative individuals. *Nature Genet.* **10**, 224–228 (1995).

©2009 Macmillan Publishers Limited. All rights reserved

10. Michon, P. *et al.* Duffy-null promoter heterozygosity reduces DARC expression and abrogates adhesion of the *P.vivax* ligand required for blood-stage infection. *FEBS Lett.* **495,** 111–114 (2001).
11. Zimmerman, P. A. *et al.* Emergence of FY*A(null) in a *Plasmodium vivax*-endemic region of Papua New Guinea. *Proc. Natl Acad. Sci. USA* **96,** 13973–13977 (1999).
12. Alberts, S. C., Buchan, J. C. & Altmann, J. Sexual selection in wild baboons: from mating opportunities to paternity success. *Anim. Behav.* **72,** 1177–1196 (2006).
13. Buchan, J. C., Alberts, S. C., Silk, J. B. & Altmann, J. True paternal care in a multi-male primate society. *Nature* **425,** 179–181 (2003).
14. Tung, J., Charpentier, M. J. E., Garfield, D. A., Altmann, J. & Alberts, S. C. Genetic evidence reveals temporal change in hybridization patterns in a wild baboon population. *Mol. Ecol.* **17,** 1998–2011 (2008).
15. Perkins, S. L. & Schall, J. J. A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. *J. Parasitol.* **88,** 972–978 (2002).
16. Garnham, P. C. C., Heisch, R. B., Minter, D. M., Phipps, J. D. & Ikata, M. *Culicoides adersi* Ingram and Macfie, 1923, a presumed vector of *Hepatocystis* (= *Plasmodium*) *kochi* (Laveran, 1899). *Nature* **190,** 739–741 (1961).
17. Wittkopp, P., Haerum, B. & Clark, A. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430,** 85–88 (2004).
18. Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B. & Kinzler, K. W. Allelic variation in human gene expression. *Science* **297,** 1143 (2002).
19. Reich, D. *et al.* Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genet.* **5,** e1000360 (2009).
20. Loisel, D. A., Rockman, M. V., Wray, G. A., Altmann, J. & Alberts, S. C. Ancient polymorphism and functional variation in the primate MHC-DQA1 5′ *cis*-regulatory region. *Proc. Natl Acad. Sci. USA* **103,** 16331–16336 (2006).
21. Wooding, S. *et al.* Independent evolution of bitter-taste sensitivity in humans and chimpanzees. *Nature* **440,** 930–934 (2006).
22. Loisel, D. *Evolutionary Genetics of Immune System Genes in a Wild Primate Population*. PhD thesis, Duke Univ. (2007).
23. Seixas, S., Ferrand, N. & Rocha, J. Microsatellite variation and evolution of the human Duffy blood group polymorphism. *Mol. Biol. Evol.* **19,** 1802–1806 (2002).
24. Moore, J. A. & Kuntz, R. E. *Entopolyploides macaci* Mayer, 1934 in the African baboon (*Papio cynocephalus* L. 1766). *J. Med. Primatol.* **4,** 1–7 (1975).
25. Brem, R. B., Storey, J. D., Whittle, J. & Kruglyak, L. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436,** 701–703 (2005).
26. Smith, E. N. & Kruglyak, L. Gene-environment interaction in yeast gene expression. *PLoS Biol.* **6,** e83 (2008).
27. Rozas, J., Sanchez-Del Barrio, J. C., Messeguer, X., Rozas, R. & Dna, S. P. DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19,** 2496–2497 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Office of the President of the Republic of Kenya and the Kenya Wildlife Service for permission to work in Amboseli National Park, and the Institute of Primate Research for local sponsorship. We thank the wardens and staff of Amboseli National Park, and the pastoral communities of Amboseli and Longido for cooperation. We thank J. Altmann for providing access to long-term data and contributing samples, and J. Altmann and Y. Gilad for providing comments on the manuscript. R. S. Mututua, S. Sayialel, and J. K. Warutere assisted with sample collection. The Integrated Primate Biomaterials and Information Resource, the Coriell Institute, J. Rogers and R. Sapolsky provided access to DNA samples from Mikumi and Masai Mara respectively. G. Gibson, T. F. C. Mackay, L. Goering and D. Tan provided access to a pyrosequencer at NC State University. M. Akinyi assisted with sample collection and analysis. A. D. Pfefferle assisted with sequencing. S. Mukherjee advised and assisted with statistical tests. Financial support came from the National Science Foundation (to S.C.A. and J.T.); the American Society of Primatologists (to J.T.); Duke University and the Duke chapter of Sigma Xi (to J.T.); and the Duke Institute for Genome Sciences and Policy (to G.A.W.).

Author Contributions J.T., S.C.A. and G.A.W. designed the study, analysed the results, and wrote the paper. J.T. and S.C.A. collected blood samples; S.C.A. provided the long-term data on Amboseli National Park. J.T. gathered the allelic imbalance and sequence data; A.P., T.F.S. and J.T. collected the transfection assay data; T.F.S. and J.T. collected the genotyping data; A.J.B. and J.T. collected the *Hepatocystis* data. G.A.W. and S.C.A. provided funding support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Sequence data have been deposited in NCBI GenBank under the accession numbers FJ952954–FJ955880, FJ955882–FJ955885, FJ955887–FJ955896 and FJ955899–FJ956699. Correspondence and requests for materials should be addressed to J.T. (jt5@duke.edu).

©2009 **Macmillan Publishers Limited. All rights reserved**

doi:10.1038/nature08149

METHODS

DNA and RNA sampling. Blood samples for DNA extraction and *Hepatocystis* screening were collected from 190 Amboseli baboons between 1989 and 2008 (Supplementary Fig. 2)[28]. DNA was extracted using standard methods. DNA for some individuals was whole genome amplified (Qiagen Repli-G Kit). Blood samples for RNA extraction were collected in PaxGene RNA tubes from 101 adult Amboseli baboons darted between 2004 and 2008. RNA was extracted using the PaxGene RNA Blood kit (Qiagen), and reverse transcribed into complementary DNA (ABI High Capacity cDNA Archive Kit).

Sequencing. We amplified and sequenced the region homologous to the annotated *FY cis*-regulatory region in humans in 174 individuals and sequenced or genotyped the two pyrosequencing assay SNPs in 150 individuals. To assess congruence between the *in vitro* and *in vivo* gene expression results for the C/T SNP, we inferred haplotype phasing using PHASE 2.1.1 (ref. 29).

Hepatocystis screen and association with *FY*. We screened for *Hepatocystis* in all 190 baboons using *Hepatocystis* mtDNA specific primers, which produced a band of approximately 251 base pairs on an agarose gel in the presence of *Hepatocystis* (Supplementary Methods).

We then fitted the following generalized linear mixed model for 150 individuals, using a binomial error structure:

$$P(y_{ij} = 1 | G_{ij}) = \operatorname{logit}\left(\beta G_{ij} + \sum_{u=1}^{5} D_{iu} v_u + S_j + b + \varepsilon\right)$$

where the logit link function has been used to change the expression in parentheses into the probability of observing the parasite, given the explanatory variables. y is *Hepatocystis* infection status ($y = 1$ corresponds to infected; $y = 0$ corresponds to uninfected), individuals are indexed by i, and study group is indexed by j. β is a fixed effect of genotype, G_{ij}; v_u is the fixed effect of the projection D_{iu} on the uth principal component of population structure; S_j is a random effect of study group; b is the intercept; and ε represents model error (Supplementary Information). Principal-components analysis of population structure was conducted on data from 47 unlinked loci (33 SNPs and 14 microsatellites, Supplementary Table S4). We also included study group as a random effect because group is a common source of structure in baboons, and infection rates clearly differed between groups. We evaluated the significance of β, the SNP effect, as evidence for association between infection and *cis*-regulatory variation.

Pyrosequencing. We designed pyrosequencing assays based on two variable SNPs in the transcribed region of the *FY* gene. 38 individuals for which cDNA samples were available were also heterozygous at one or both of these sites. For each of these individuals, we performed six to eight pyrosequencing reactions across two plates (mean number of measurements per individual = 7.05, range = 3–8, excluding failed reactions). The resulting values were expressed as the $\log_2$ transformed ratio of the expression of transcripts carrying one versus the other allele at one of the assay SNPs (values based on the alternative assay SNP were converted on the basis of linkage between the two sites).

We identified an effect of one of the upstream *cis*-regulatory sites by modelling variation in allelic imbalance using the following general linear mixed model:

$$y_{ij} = \beta G_{ij} + Y_j + b + \varepsilon$$

where y is allelic imbalance, indexed by individual i and year of sampling j; β is a fixed effect of homozygous or heterozygous genotype, G_{ij}; Y_j is a random effect of year of sampling (2006, 2007 or 2008; one individual was sampled in 2005 and grouped with the 2006 samples); b is the intercept; and ε is the model error. We assessed the significance of β using a permutation test: all measurements for an individual were grouped as a block and permuted 1,000 times over individual identity. We assigned a P value to the original SNP parameter estimate by ranking it among the corresponding estimates for the permuted data sets. The estimate of variance explained by the C/T site is based on modelling the residuals of allelic imbalance on year of sampling, using the C/T site alone.

Three pairs of individuals in the allelic imbalance analysis were related at $r = 0.5$; removing any set of three individuals so that no individuals were closely related did not qualitatively change our results (see also Supplementary Methods).

Transfection assays. Human erythroleukaemic cells (HEL 92.1.7) were maintained using the ATCC protocol. The wells of 24-well cell culture plates were seeded with 2×10^5 cells in 500 µl media, transfected, incubated for 48 h, and lysed. Cells were co-transfected with experimental constructs or empty firefly luciferase vector as control (pGL4.10; 180 ng per well) and the CMV Renilla normalization construct (pGL4.75, 20 ng per well; Promega) using Fugene 6 (Roche). Expression levels were measured with a dual-luciferase reporter assay (DLR1000 assay kit, Promega) and reported as relative ratios of luminescence (firefly:Renilla). Eight replicate wells were transfected for each experimental and control vector within an assay, with the assay repeated three times ($n = 24$ total measurements per construct).

We compared the measurements for each pair of constructs (A versus G for the *Hepatocystis*-associated SNP and C versus T for the SNP associated with allelic imbalance) separately using the following model:

$$y_{ijk} = \beta C_j + E_k + b + \varepsilon$$

where y is the relative ratio of luminescence for replicate i of construct j in experiment k; β is a fixed effect of construct, C_j; E_k is a random effect of experiment; b is the intercept; and ε is model error.

Signature of selection. For comparison of the F_{st} values, we genotyped up to 36 polymorphic microsatellite loci in ten baboons from the Masai Mara Reserve, Kenya; 20 baboons from Mikumi National Park, Tanzania; and 12 baboons from Amboseli National Park (Supplementary Information). One locus is a polymorphic microsatellite located ~3.7 kb upstream of the sequenced *FY cis*-regulatory region[23]; the other 35 loci reside in putatively neutral sites dispersed around the baboon genome[12,13]. We calculated F_{st} values for each of these loci independently using Arlequin 3.1 (ref. 30), and compared the F_{st} value for the *FY*-linked microsatellite locus with the F_{st} values for the 35 neutral loci (Supplementary Figs 3, 4 and Supplementary Methods). For comparison of Tajima's D, we resequenced 21 other regions in and around genes in the same or a subset of the individuals resequenced at the *FY cis*-regulatory region (Supplementary Fig. 3 and Supplementary Table 5). We calculated Tajima's D for all loci using the program DnaSP version 4.9 (ref. 27), assuming no recombination.

28. Altmann, J. *et al.* Behavior predicts genetic structure in a wild primate group. *Proc. Natl Acad. Sci. USA* **93**, 5797–5801 (1996).
29. Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
30. Excoffier, L., Laval, G. & Schneider, S. Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50 (2005).

©2009 Macmillan Publishers Limited. All rights reserved