Cell

No Positive Selection for G Allele in a p53 Response Element in Europeans

In a recent publication, a SNP, rs4590952, in a functional *KITLG* p53-response element (p53-RE) was claimed to have undergone adaptive evolution in Europeans (Zeron-Medina et al., 2013). This conclusion was based on the experimental finding that the G allele exhibits stronger p53-RE activity than the A allele and that the observation that the G allele is found to be in much higher frequency in Europeans than Africans. This SNP is in high linkage disequilibrium with variants that are strongly associated with cancer risk in genome-wide association studies. If confirmed, it would serve as one of a handful of examples where adaptive variation is strongly associated with disease, as opposed to the vast majority of disease variants that are currently explained by the action of purifying selection (Dudley et al., 2012). However, our molecular evolutionary and population genomic analyses do not support the conclusion that the G allele has risen to high frequencies in Europeans by positive selection.

Our multispecies and population genomic analyses clearly establish that the G allele is the ancestral allele in modern humans. This is evident from a survey of the genomes of all nonhuman primates, including great apes, monkeys, and ancient primates, all of which have the G allele (Figure 1A). Two Neanderthal genomes with sequences available for this segment also have a G allele. In fact, we find a complete lack of population polymorphism in data from diverse chimpanzees, gorillas, and orangutans (Prado-Martinez et al., 2013) at the position containing rs4590952. These evolutionary and population variation patterns indicate a strong negative selection against non-G alleles. This is particularly remarkable because the position of this particular SNP appears to be part of a hypermutable CpG dinucleotide in old world monkeys and apes, as the adjacent 5′ position has undergone a CpG to TpG transition in orangutan and the rs4590952 SNP in modern humans is a CpG to CpA transition (Figure 1A).

We confirmed the G allele to be the ancestral allele in modern humans in a haplotype network analysis. Using the genomic segment flanking SNP rs4590952, we inferred multiple distinct haplotypes from chimpanzees, gorillas, and 1,092 human individuals (1000 Genomes data). Two predominately African haplotypes containing the G allele (H_1 and H_2) flank the root haplotype present in great apes (H_0; Figure 1B). Therefore, evolutionary and population genomic analyses designate G as the ancestral allele in modern humans. This predicts the G allele to be predominant in human populations, which is indeed the case for not only Europeans, but also Asians and Americans (Figure 1C). This expectation obviates the need to require positive selection for the G allele to rise to high frequency in Europeans. Furthermore, an ancestral G allele invalidates the Zeron-Medina et al. (2013) interpretation of the high iHS test statistic in Europeans, because it assumes that G is the derived allele. In light of these results, the experimental finding of stronger p53-RE activity of the G allele is best interpreted as an ancestral feature, where strong purifying selection has operated against other alleles with weaker activity (e.g., A allele) for tens of millions of years in primate evolution leading to humans.

Now, the question is why do Africans have the derived A allele in high frequency? There are two possible explanations. First, there may be positive selection for the weaker functional activity of A alleles, resulting in its increase in frequency in Africans. The second possibility is that purifying selection to preserve the G allele has been relaxed in modern humans, and, thus, the fate of alleles at the rs4590952 position is now largely determined by genetic drift and its local genomic context. The F_{st} statistic cannot distinguish between these alternatives, as large differences in ancestral allele frequencies result in high F_{st} values, even for positions that are likely to be evolving neutrally (Figure 1D; see also Maruki et al., 2012). So, we used site-specific extended haplotype homozygosity (EHHS) profiles to detect signatures of positive selection around SNP rs4590952 in Africans. Because EHHS compares the EHH profile of a SNP position rather than that of individual alleles, its interpretation does not depend on the knowledge of the ancestral state. The EHHS profiles (Figure 1E) show a rapid decay in African populations, which is consistent with the patterns that would be generated by neutral evolutionary trends. However, homozygosity in Europeans decays at a slower rate. We noticed that this pattern is not unique to rs4590952, rather it is present in other similar SNPs in the same genomic segment that show high F_{st} and ancestral alleles in high frequency in Europeans and low frequency in Africans (Figure 1E). This genomic segment is already known to be under positive selection for pigmentation phenotypes (Lao et al., 2007). So it appears that the relaxation of purifying selection on the position harboring rs4590952 is allowing the frequency of alternative alleles to be dictated by random genetic drift and hitchhiking.

In summary, analysis of population variation and experimental functional genomics data in light of the multispecies evolutionary patterns establish that the high frequency of the G allele and its stronger functional activity are evolutionary features that do not require invocation of positive selection in Europeans.

AUTHOR CONTRIBUTIONS

S.K. designed the study. L.L. performed the analysis. S.K. and L.L. wrote the manuscript.

Sudhir Kumar[1,2,*] **and Li Liu**[1]
[1]Center for Evolutionary Medicine and Informatics, Biodesign Institute, Arizona State University, Tempe, AZ 85287-5301, USA
[2]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA 19122, USA
*Correspondence: s.kumar@temple.edu (S.K.)
http://dx.doi.org/10.1016/j.cell.2014.05.033

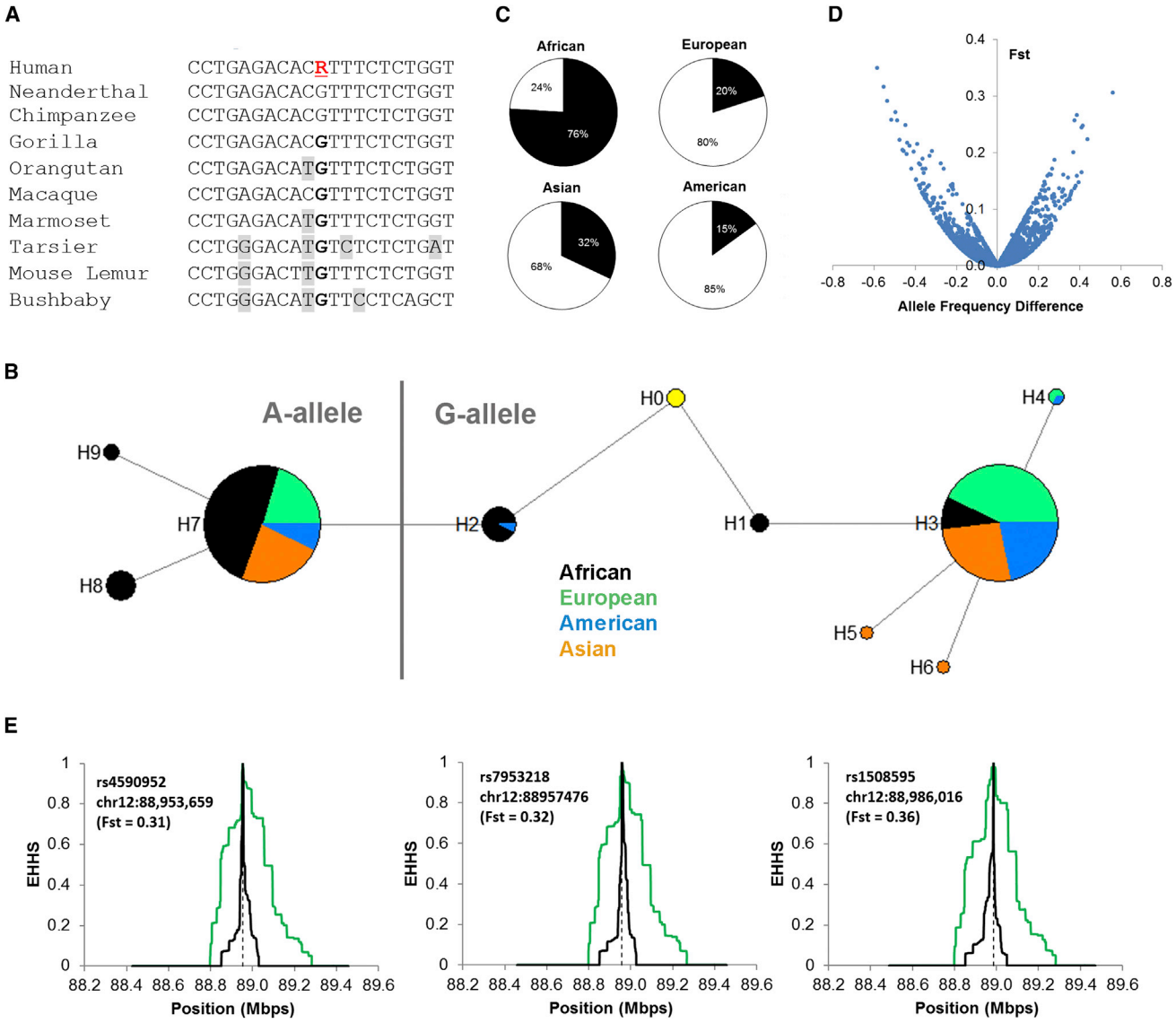

Figure 1. Multispecies and Population Genomic Analyses

(A) A multispecies alignment of flanking sequences around SNP rs4590952 (highlighted in red) retrieved from the Ensembl Variation Database. Human population shows [G/A] polymorphism (represented by R). Bases that differ from the human reference sequence are highlighted in gray. Based on the adjacent neighboring 5′ nucleotide context, the SNP of interest is likely to be part of a hypermutable CpG dinucleotide that has undergone deamination-caused transitions at the first base (CpG→TpG) in orangutan, and at the second base in humans (CpG→CpA).

(B) Haplotypes relationships constructed using Network 4.6 based on 1 kb flanking sequences around rs4590952. Haplotypes were inferred using the PHASE program based on the sequences of chimpanzee, gorilla, and 1,092 human individuals surveyed in the 1000 Genomes Project. Haplotypes containing the G allele and those containing the A allele separate into two distinct groups. The network was rooted on chimpanzee and gorilla haplotype (H_0). Each circle represents a haplotype and its size is proportional to the number of chromosomes with that haplotype. Colors within each circle represent a corresponding population.

(C) Pie charts illustrating the frequencies of the G allele (black slices) and A allele (white slices) of rs4590952 in various human populations surveyed in the 1000 Genomes Project.

(D) The relationship between F_{st} and the difference in the frequency of the ancestral allele in Europeans and Africans. 10,000 SNPs found at noncoding CpG dinucleotides were chosen randomly from chromosome 1 in humans. Genotypes of 246 African individuals and 379 European individuals were retrieved from the 1000 Genomes data. Ancestral alleles were designated based on their presence in the great apes genomes.

(E) EHHS plots of the SNP of interest (rs4590952), a CpG SNP (rs7953218) nearby within the KITLG gene and a CpG SNP (rs1508595) outside the KITLG gene for the European population (green lines) and the African population (black lines). These two SNPs were selected based on their high F_{st} values (>0.3) and the presence of the ancestral allele in Europeans and the derived allele in Africans, as such a comparison allows one to control the differences in the ages among populations and alleles. EHHS values were calculated using the R-rehh package that was also employed by Zeron-Medina et al. (2013).

REFERENCES

Dudley, J.T., Kim, Y., Liu, L., Markov, G.J., Gerold, K., Chen, R., Butte, A.J., and Kumar, S. (2012). Genome Res. *22*, 1383–1394.

Lao, O., de Gruijter, J.M., van Duijn, K., Navarro, A., and Kayser, M. (2007). Ann. Hum. Genet. *71*, 354–369.

Maruki, T., Kumar, S., and Kim, Y. (2012). Mol. Biol. Evol. *29*, 3617–3623.

Prado-Martinez, J., Sudmant, P.H., Kidd, J.M., Li, H., Kelley, J.L., Lorente-Galdos, B., Veeramah, K.R., Woerner, A.E., O'Connor, T.D., Santpere, G., et al. (2013). Nature *499*, 471–475.

Zeron-Medina, J., Wang, X., Repapi, E., Campbell, M.R., Su, D., Castro-Giner, F., Davies, B., Peterse, E.F., Sacilotto, N., Walker, G.J., et al. (2013). Cell *155*, 410–422.