Evolutionary History of the *ADRB2* Gene in Humans

To the Editor: Recently, Cagliani et al.[1] reported on the evolutionary history of the human β2-adrenoreceptor gene (*ADRB2* [MIM 109690]). The authors presented their analysis of the genealogy of inferred *ADRB2* haplotypes reconstructed through the use of a median-joining network approach. *ADRB2* shows a high level of polymorphism, including three coding variants with altered functional properties (at codons 16, 27, and 164, hereafter referred to as C16, C27, and C164), and several suspected regulatory variants. The network presented by Cagliani et al.[1] was prepared with software implementing a model that assumes no recombination but allows for recurrent mutation. We notice that most SNPs in their network are represented by recurrent mutations in nonreticulated parts of the network, which we find biologically implausible and difficult to interpret in an evolutionary context. We have prepared a network diagram for the evolution of common *ADRB2* haplotypes that can be accounted for by divergence via the accumulation of mutations and rare recombination events between diverged haplotypes (Figure 1). This revised network requires no recurrent mutations (homoplasy) and accounts for the major haplotype groups of Cagliani et al.,[1] Hawkins et al.,[2] and Drysdale et al.,[3] as well as those revealed by the HapMap data, with only one recombination event.

To construct this revised network, we generated an alignment of genomic sequences for human, chimpanzee, and macaque as an outgroup, using GenBank accession nos. DQ094845, NC_006472, and NC_007863, respectively. The alignment, generated with the use of ClustalX, spanned *ADRB2* from base position −4219 to +5479 relative to the first base of the initiation codon of the main polypeptide product (Figure S1, available online). It was found that the chimpanzee genome sequence used gave poor alignment in the region from −600 to −900. We used our own chimpanzee sequence generated by PCR amplification from genomic DNA to replace poorly aligned regions of NC_006472. In Figure 1 we have shown only the network connecting the major haplogroups. Our network accounts for the generation of the HC1 haplogroup in Cagliani et al.[1] (haplogroup 1.1 in Hawkins et al.[2]) by recombination between the 5′ end of a haplotype containing the ancestral G alleles at each of the common functional coding SNPs, C16 and C27 (i.e., a "GG" haplotype), and the 3′ end of a doubly-derived C16/C27 AC haplotype. This network, produced by eye, satisfies conditions of parsimony, requiring only one recombination event and no homoplasic mutations. Further mutation and recombination events, especially around the 3′ end of the coding region, would account for the full spectrum of rare and region-specific haplotypes observed in Cagliani et al.[1] and Hawkins et al.[2] (unpublished data).

The major difference in network topology that we find in relation to the common functional coding SNPs is that the first divergence from the most recent common ancestor (MRCA) of extant haplotypes separates the C16/C27 GG haplogroup from the remaining (GC and AC) haplogroups, with mutation at the C27 site creating a GC haplotype, followed by divergence of the GC and AC haplogroups, indicating that the C16 mutation that created the AC haplotype occurred on a GC haplotype background (Figure 1). The order of mutation events and consequent high degree of linkage disequilibrium between the neighboring C16 and C27 sites probably explains why AG haplotypes are never found in well-sequenced data sets worldwide.

All of the sites represented in the network by Cagliani et al.[1] are currently polymorphic, so their node labeled "chimp" may thus better be represented as the most recent common ancestral human node for this gene. In our revised network, we have used comparison with the macaque outgroup to assign all fixed mutations on the branches leading to the chimpanzee and to the ancestral human. Using our sequence for the chimpanzee, we find fewer fixed differences (50 sites) between the chimpanzee and the human common ancestor haplotype than the number reported in Cagliani et al.[1] (77 sites). Of these 50 differences, 32 occur on the chimpanzee branch and 18 occur on the human branch. The ratio of polymorphic to fixed sites within the human lineage will relate to the relative time to the most recent common ancestor (TMRCA) of all extant *ADRB2* haplotypes compared with time since the divergence from chimpanzee, although the true value would be affected by differences in generation times, population parameters, and per-generation substitution rates within each species and haplotype lineage. Cagliani et al.[1] estimate this TMRCA to be in the range of 1.05–1.65 million years. The revised network that we have generated points to a significantly more ancient MRCA for this gene, given that we find that 11, 8, and 13 currently polymorphic sites are found in the GG, AC, and GC haplogroups, respectively, compared with the 18 fixed human mutations that have occurred since the human/chimpanzee common ancestor.

The mutation rate analysis in Cagliani et al.[1] is constrained by the use of polymorphism data only. Further evidence that *ADRB2* is unusually polymorphic comes from a comparison between the nucleotide-substitution frequency in the lineage to the macaque outgroup and those to the fixed and polymorphic sites along the other branches. To illustrate the changes in substitution frequency across *ADRB2* and the differences between lineages, we counted the number of substitutions (of all

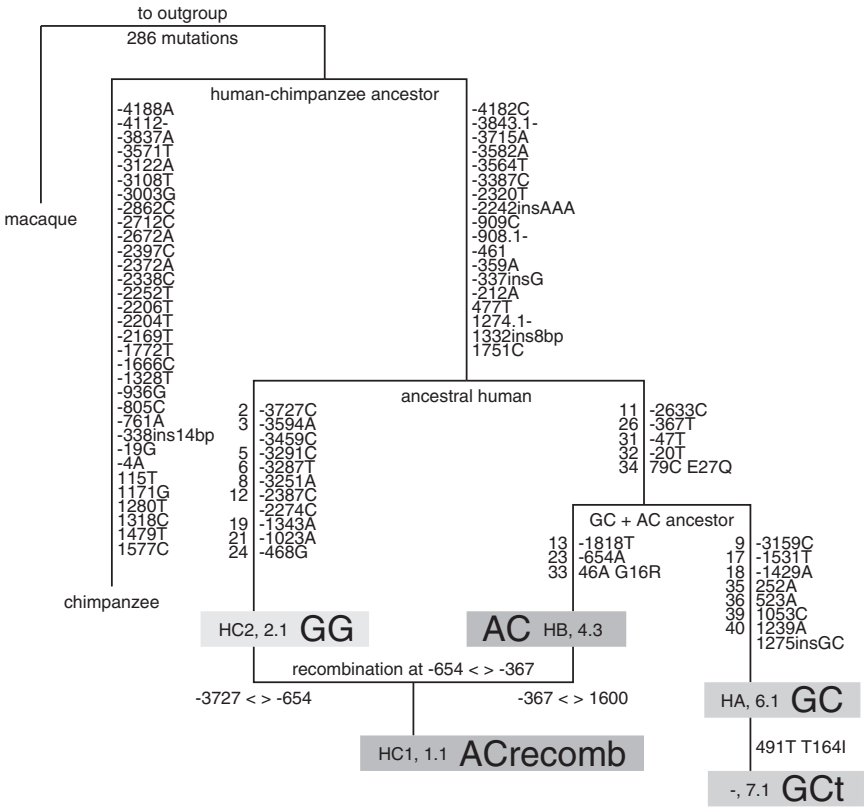

Figure 1. Evolutionary Network for the Common Haplotypes of Human *ADRB2*

Substituted and polymorphic sites in *ADRB2* are listed along the branches the mutation occurred in. Sites different in macaque but identical in human and chimpanzee are counted but not listed. Nodes connecting lines represent common ancestral haplotypes and are labeled. Sites are coded by position in the aligned human sequence relative to the first base of the initiation codon, derived allele, and whether the change involved insertion or deletion. Insertions and deletions were treated as single mutational events for calculation of the network regardless of the number of bases altered. Within the C tract at position 1275, the coappearance of adjacent G and C inserted bases were treated as a single mutation (see Table 1). Major haplogroups are indicated in shaded boxes. Within each box, these are coded as Cagliani et al. haplogroup (e.g., "HC2"), Hawkins et al. haplogroup (e.g., "2.1"), and C16/C27 functional SNP haplogroup (e.g., "GG"). GCt refers to the rare C164 derived haplotype not considered by Cagliani et al.[1] Branches rejoining below a node represent formation of a new haplotype via recombination. Regions contributed by each parental haplotype and regions of recombination are indicated with "<" and ">." Rare, population-specific polymorphic sites have not been shown on this network.

mutation categories) in each lineage in 500 bp windows across the entire region, moving along the gene in 200 bp steps, and plotted these values as substitution frequencies per 1000 bp on a line graph (Figure 2). To look for statistical differences in substitution frequency between lineages, we used Fisher's exact test to compare the frequency of substitutions in each 500 bp window for chimpanzee fixed, human fixed, and human polymorphic sites, respectively, with that in the macaque/outgroup. Individual 500 bp windows giving $p < 0.05$ in these tests were investigated further by extending the window in each direction and retesting with Fisher's exact test until no further decrease in p value was observed (i.e., to give the minimum p value). The substitution frequency varied significantly (more than 5-fold) across *ADRB2* in the macaque/outgroup (500 bp windows, $p < 10^{-6}$ for heterogeneity, 18 degrees of freedom, χ-squared test). For the other branches, ten out of 144 individual 500 bp windows showed differences in

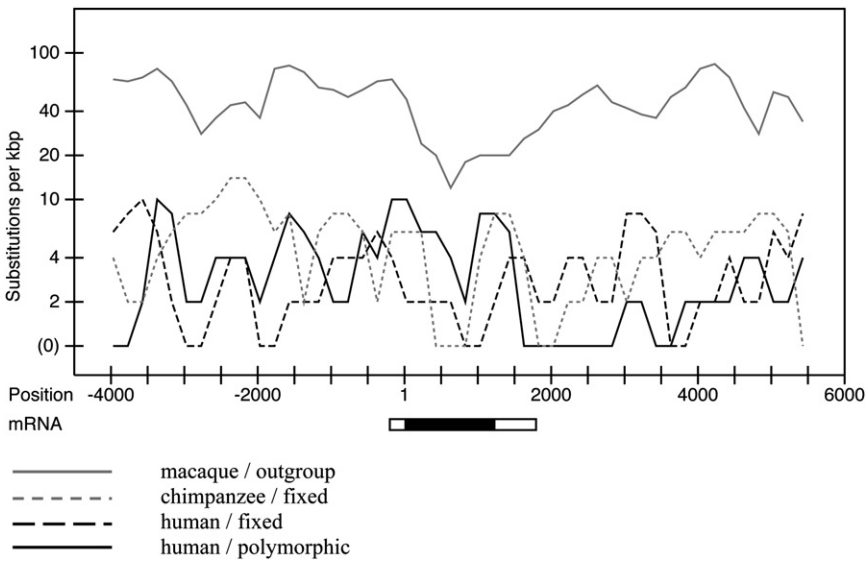

Figure 2. Substitution and Polymorphism Rates across the *ADRB2* Region

Number of substitutions per kbp are plotted, on a logarithmic scale, versus position along the genomic region analyzed for central points in a sliding window of 500 bp, with step size of 200 bp. Base position 1 is the first base of the initiation codon. The black and white bar represents the *ADRB2* coding region and UTRs, respectively.

Table 1. Allelic Composition of 3′ Polymorphic Sites in *ADRB2*

Position										
C16/C27 Haplogroup	46	79	491	1053	1239	1263–1282	C Tract[a]	R?	N	Percentage of Haplogroup[b]
GG	G	G	C	G	G	AGACCCCCCC–– – –CCCAAC	10		2	4
GG	G	G	C	G	G	AGACCCCCCCC–– – –CCCAAC	11		48	90
GG	G	G	C	G	**A**	AGACCCCCCCC–– – –CCCAAC	11	R	2	4
GG	G	G	C	**C**	**A**	AGACCCCCCCC–**CG**CCCAAC	13 +G	R	1	2
GC	G	**C**	C	**C**	**A**	AGACCCCCCCC–**CG**CCCAAC	13 +G		37	97
GC	G	**C**	C	**C**	G	AGACCCCCCCC–**CG**CCCAAC	13 +G	R	1	3
GCt	G	**C**	**T**	**C**	**A**	AGACCCCCCCC–**CG**CCCAAC	13 +G		3	N/A[c]
AC	**A**	**C**	C	G	G	AGACCCCCCCC–– – –CCCAAC	11		35	56
AC	**A**	**C**	C	G	G	AGACCCCCCCC**C**– –CCCAAC	12		5	8
AC	**A**	**C**	C	**C**	**A**	AGACCCCCCCC–**CG**CCCAAC	13 +G	R	19	31
AC	**A**	**C**	C	G	**A**	AGACCCCCCCC–**CG**CCCAAC	13 +G	R	1	2
AC	**A**	**C**	C	**C**	**A**	AGACCCCCCCC**CCG**CCCAAC	14 +G	R	1	2

Derived alleles are shown in bold, including additional C residues appearing within the C tract. Abbreviations are as follows: N, no. of chromosomes observed; R, recombinant after position 79. In some rare cases, recombination cannot easily be distinguished from back mutation or gene-conversion events.

[a] Length of C tract, including polymorphic G insertion if present; "+G" indicates that G insertion is present.

[b] Proportion of chromosomes from each of the three major haplogroups (GG, GC, and AC); GCt indicates the rare haplotype containing the derived T allele at the third functional nonsynonymous site, codon 164, nucleotide position 491.

[c] The GCt chromosomes are not counted as part of the GC haplogroup for this analysis because they were specifically selected.

substitution frequency (at $p < 0.05$) when compared to the macaque/outgroup branch (Table S1). After extension of these ten windows to find the minimum p values, we found that the two most significant regions were the human polymorphic sites from position −50 to +1300 ($p = 0.00001$) and from position +1300 to +4050 ($p = 0.0013$). The macaque/outgroup substitution frequencies demonstrate a long-term reduction in substitution rate over the coding region, signaling constraint presumably associated with stabilizing selection on the ADRB2 functional phenotype. In agreement with Cagliani et al.,[1] we note a relative increase in the density of surviving human polymorphic mutations in the coding region relative to the macaque/outgroup branch. This high diversity in our species may be associated with long-term balancing selection or with relaxed constraint. We also note the low diversity downstream of the transcript, which may be indicative of a selective sweep, although the possibility of incomplete SNP ascertainment in this region cannot be discounted.

We have also characterized the haplotype structure at the 3′ end of *ADRB2*, which was not addressed in detail by Cagliani et al.[1] or by Hawkins et al.,[2] nor addressed completely by another study by the latter group.[4] Table 1 shows the haplotypes obtained by sequencing 83 European individuals from our earlier study of Greek teenagers.[5] This study was approved by both Greek and UK local research ethics committees, and all subjects gave parentally-sanctioned informed consent. All sequenced individuals were homozygous for haplotypic combinations at codons 16 and 27 (i.e., GG homozygotes, GC homozygotes, or AC homozygotes), apart from one individual heterozygous for a GCt haplotype. Homozygosity facilitated the reading of the sequence electropherogram across the polyC tract beginning at position 1266. We found almost complete linkage disequilibrium between the C16/C27 coding polymorphisms and the 3′ polymorphisms within the GG and GC haplotypes, but considerable diversity within the AC haplotype. This 3′-end diversity is best explained by a combination of recombination events between haplotypes and variation in C-tract homopolymer length. In addition to the variability in the length of the C tract, we would emphasize the hitherto unappreciated insertion of a G at position 1275 within the C tract in certain haplotypes. The functional importance of length variation in the C tract in relation to mRNA stability has been highlighted by Panebra et al.,[4] but no alleles with a C tract interrupted by the 1275G insertion were tested in their study. The insertion of a G into the homopolymer C tract would be expected to influence C tract conformation and interactions involved in mRNA stability at least as much as changes in homopolymer length. Both the C tract length and the G insertion should be taken into account in future functional studies of variation at the 3′ end of the gene.

The degree of linkage disequilibrium between the 3′ variants and upstream variants may be a reflection of possible functional interactions between variants in different regions of *ADRB2* in relation to protein function, transcriptional regulation, and mRNA stability. There may be

additional effects mediated by the polymorphic site at position 1239, immediately 5′ to the stop codon, affecting translation termination. A complete and accurate picture of variation across the entire gene is required before such interactions can be studied effectively.

Richard H. Wilson,[1] Colin N. Moran,[1] John J. Cole,[1] Yannis P. Pitsiladis,[1] and Mark E.S. Bailey[1,*]
[1]Molecular Genetics and Institute of Diet, Exercise and Lifestyle (IDEAL), Faculty of Biomedical & Life Sciences, University of Glasgow, Glasgow, G12 8QQ, UK
*Correspondence: m.bailey@bio.gla.ac.uk

Supplemental Data

Supplemental Data include one table and one figure and can be found with this article online at http://www.ajhg.org.

Acknowledgments

We are grateful to M. Sironi for unpublished data that enabled us to interpret Cagliani et al.'s Figure 1. This research was supported in part by a BBSRC research grant, no. 17/D17566.

Web Resources

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/SNP/index.html
HapMap, http://www.hapmap.org
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/omim/

Accession Numbers

The chimpanzee *ADRB2* gene sequence generated by us has been given GenBank accession number GU373723.

References

1. Cagliani, R., Fumagalli, M., Pozzoli, U., Riva, S., Comi, G.P., Torri, F., Macciardi, F., Bresolin, N., and Sironi, M. (2009). Diverse evolutionary histories for β-adrenoreceptor genes in humans. Am. J. Hum. Genet. *85*, 64–75.
2. Hawkins, G.A., Tantisira, K., Meyers, D.A., Ampleford, E.J., Moore, W.C., Klanderman, B., Liggett, S.B., Peters, S.P., Weiss, S.T., and Bleecker, E.R. (2006). Sequence, haplotype, and association analysis of ADRbeta2 in a multiethnic asthma case-control study. Am. J. Respir. Crit. Care Med. *174*, 1101–1109.
3. Drysdale, C.M., McGraw, D.W., Stack, C.B., Stephens, J.C., Judson, R.S., Nandabalan, K., Arnold, K., Ruano, G., and Liggett, S.B. (2000). Complex promoter and coding region β2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. Proc. Natl. Acad. Sci. USA *97*, 10483–10488.
4. Panebra, A., Schwarb, M.R., Swift, S.M., Weiss, S.T., Bleecker, E.R., Hawkins, G.A., and Liggett, S.B. (2008). Variable-length poly-C tract polymorphisms of the β2-adrenergic receptor 3′-UTR alter expression and agonist regulation. Am. J. Physiol. Lung Cell. Mol. Physiol. *294*, L190–L195.
5. Moran, C.N., Yang, N., Bailey, M.E.S., Tsiokanos, A., Jamurtas, A., MacArthur, D.G., North, K., Pitsiladis, Y.P., and Wilson, R.H. (2007). Association analysis of the *ACTN3* R577X polymorphism and complex quantitative body composition and performance phenotypes in adolescent Greeks. Eur. J. Hum. Genet. *15*, 88–93.

DOI 10.1016/j.ajhg.2010.01.031.

Response to Wilson et al.

To the Editor: Wilson et al. noticed that the *ADRB2* (MIM 109690) haplotype network we have previously presented[1] displays some recurrent mutations; therefore, they propose an alternative genealogy for *ADRB2* haplotypes. Apparent homoplasies may result from different processes, namely recurrent mutation (true homolplasy) and recombination or gene conversion. It is often difficult to discriminate among these possibilities. As an example, three recurrent mutations in the *ADRB2* gene region occurred at CpG sites, suggesting that they may represent true homoplasies. Nonetheless, we agree with the authors that the network that they present might better satisfy parsimony conditions as compared to the one that we provided. We would like to mention that network analysis may not be robust in regions where recombination events are evident, especially when times to the most recent common ancestor (TMRCAs) need to be estimated. For this reason, we had performed an additional analysis in our previous publication[1] by the use of GENETREE.[2,3] This program is based on a maximum-likelihood coalescent analysis and assumes an infinite-site model so that SNPs and haplotypes that do not follow this rule have to be removed (see below).

We are grateful to Wilson et al. for noticing that the chimpanzee GenBank accession NC_006472 contained a few sequencing errors. Specifically, a 300 bp region upstream the transcription start site of *ADRB2* displayed a relatively poor alignment to the human sequence due to the presence of several single base pair mismatches. Following Wilson et al.'s indication, we resequenced the same region in three unrelated chimpanzees, and we confirmed that the sequence in this region is identical to the one obtained by Wilson et al. and nonpolymorphic (at least in this small number of individuals). With the use of this sequence, the GENETREE estimation of the TMRCA resulted equal to 1.90 million years (standard deviation 0.53 million years) (Figure 1). As expected, this TMRCA is deeper that the one that we had previously calculated,[1] therefore providing further support to our previous conclusion; i.e., that *ADRB2* has been evolving under a balancing-selection regime.

With respect to the sliding window analysis presented by Wilson et al., it should be noted that the evaluation of statistical significance in multiple, nonindependent