

The adaptive variant *EDARV370A* is associated with straight hair in East Asians

Jingze Tan · Yajun Yang · Kun Tang ·
Pardis C. Sabeti · Li Jin · Sijia Wang

Received: 12 February 2013 / Accepted: 9 June 2013 / Published online: 23 June 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract Hair straightness/curliness is a highly heritable trait amongst human populations. Previous studies have reported European specific genetic variants influencing hair straightness, but those in East Asians remain unknown. One promising candidate is a derived coding variant of the ectodysplasin A receptor (*EDAR*), *EDARV370A* (*370A*), associated with several phenotypic changes of epidermal appendages. One of the strongest signals of natural selection in human genomes, *370A*, has risen to high prevalence in East Asian and Native American populations, whilst being almost absent in Europeans and Africans. This striking frequency distribution and the pleiotropic nature of *370A* led us to pursue if hair straightness, another epidermal appendage-related phenotype, is affected by this variant. By studying 1,718 individuals from four distinctive East Asian populations (Han, Tibetan, Mongolian, and Li), we found a significant association between *370A* and the straight hair type in the Han ($p = 2.90 \times 10^{-6}$), Tibetan

($p = 3.07 \times 10^{-2}$), and Mongolian ($p = 1.03 \times 10^{-5}$) populations. Combining all the samples, the association is even stronger ($p = 5.18 \times 10^{-10}$). The effect of *370A* on hair straightness is additive, with an odds ratio of 2.05. The results indicate very different biological mechanisms of straight hair in Europe and Asia, and also present a more comprehensive picture of the phenotypic consequences of *370A*, providing important clues into the potential adaptive forces shaping the evolution of this extraordinary genetic variant.

Introduction

Recent genome-wide scans of natural selection in human populations have identified hundreds of candidate adaptive alleles for further study. One of the most compelling candidates is a derived coding variant of the Ectodysplasin A receptor (*EDAR*), *EDARV370A* (*370A*) (Sabeti et al. 2007; Grossman et al. 2010; Kamberov et al. 2013). Analysis of the geographical distribution of *370A* revealed that the

J. Tan and Y. Yang contributed equally to this work.

P. C. Sabeti, L. Jin, and S. Wang are cosenior authors of this work.

J. Tan · Y. Yang · L. Jin (✉)
State Key Laboratory of Genetic Engineering and MOE Key
Laboratory of Contemporary Anthropology, School of Life
Sciences, Fudan University, Shanghai 200433, China
e-mail: lijn.fudan@gmail.com

J. Tan · Y. Yang · L. Jin
CMC Institute of Health Sciences, 1 Yaocheng Road,
Taizhou 225300, Jiangsu, China

K. Tang · L. Jin · S. Wang (✉)
CAS-MPG Partner Institute for Computational Biology,
Shanghai Institutes for Biological Sciences, Chinese Academy of
Science, Shanghai 200031, China
e-mail: wangsijia@picb.ac.cn

P. C. Sabeti · S. Wang
The Broad Institute of MIT and Harvard, Cambridge,
MA 02142, USA

P. C. Sabeti · S. Wang
Department of Organismic and Evolutionary Biology, Center for
Systems Biology, Harvard University, Cambridge, MA 02138,
USA

P. C. Sabeti
Department of Immunology and Infectious Diseases, Harvard
School of Public Health, Boston, MA 02115, USA

allele is at a high frequency in East Asian and Native American populations, but is nearly absent in Europeans and Africans (Sabeti et al. 2007). Previous studies using various computational methods demonstrated that the corresponding non-synonymous SNP (rs3827760) shows one of the strongest signals of selection in the human genome (Sabeti et al. 2007; Bryk et al. 2008; Xue et al. 2009; Grossman et al. 2010). In support of this hypothesis, *370A* was also shown to associate with several phenotypic changes of epidermal appendages, including hair thickness (Fujimoto et al. 2008a, b), dental morphology (Kimura et al. 2009; Park et al. 2012; Kamberov et al. 2013), and sweat gland density (Kamberov et al. 2013). Equivalent phenotypic changes have been reported in a *370A* knock-in mouse model (Kamberov et al. 2013).

Although *370A* is a genetic variant of wide interest, one phenotype that has not been adequately explored amongst potential phenotypic changes of epidermal appendages is hair straightness. While an overexpression model of the *Edar* gene in mice showed that it led to straight hair (Mou et al. 2008), the *370A* knock-in mouse model (Kamberov et al. 2013) demonstrated that this gross overexpression does not adequately model the phenotypic changes driven by the *370A* single amino-acid change. Thus, there has yet been any direct evidence on whether *370A* allele is associated with straight hair, in either humans or mice. Indeed, Fujimoto and colleagues (2008a, b) found no significant association when they examined hair index, a phenotype often used as a proxy to hair straightness in human populations. Those previous human studies, however, were limited by a relatively small sample size. By increasing the sample size and studying different East Asian populations, we aim to adequately test the hypothesis whether *370A* is associated with straight hair in human populations.

Materials and methods

Populations

We collected 1,718 samples from four East Asian populations in different geographical regions of China, including 973 Han from Taizhou, Eastern China; 226 Tibetan from Shigatse, Western China; 323 Mongolian from Hulunbuir, Northern China; and 196 Li (Hlai) from Hainan, Southern China (Fig. 1; Table 1).

Hair straightness rating

We rated hair straightness by eye on a three-point scale (straight, wavy, and curly) (Medland et al. 2009a), after confirmation that the participant had never had their hair treated. Since curly hair was at a very low frequency in our

populations (10 cases in total from all samples of four populations), we grouped wavy and curly hair into one category “non-straight” in the analyses presented in the rest of the paper.

Genotyping

We collected blood samples from the participants, and extracted DNA using standard phenol/chloroform method. For Han samples, we performed genotyping of the SNP of interest, rs3827760, using a SNaPshot assay and called genotypes with GeneMapper v2.0. For Tibetan, Mongolian and Li samples, we performed genotyping of rs3827760 with Sequenom MassARRAY and called genotypes with MassARRAY Typer Version 4.0. We completed this work under approved protocols from the Institutional Research Board at Fudan University.

Statistical analysis

We applied logistic regression to test the association between the *370A* genotypes and hair straightness, implemented using PLINK (Purcell et al. 2007). We adopted the additive model, as supported by previous *370A* studies in humans (Fujimoto et al. 2008a, b; Kimura et al. 2009; Park et al. 2012; Kamberov et al. 2013) and transgenic mice (Mou et al. 2008; Chang et al. 2009; Kamberov et al. 2013). We controlled for the confounding effect of age and sex by applying both age and sex as covariates in the logistic regression. In our analysis of combined samples from the four different populations, we introduced another covariate, the collection location, to account for potential substructure in the combined samples.

Results

In the Han, Mongolian and Li populations, the frequency of straight hair is higher in the derived homozygote *370A* than in the heterozygote *370V/370A* (Fig. 2c–f). This trend is statistically significant in Han (two-tailed Fisher’s exact test, $p < 0.0001$) and Mongolian (two-tailed Fisher’s exact test, $p = 0.0002$), but not in Li population (two-tailed Fisher’s exact test, $p = 0.3425$). In the Tibetans, where we have collected 25 ancestral homozygote *370V* samples, the frequency of straight hair is the highest in *370A*, and the lowest in *370V*, with the heterozygote *370V/370A* at an intermediate frequency (Fig. 2d). This supports an additive model of *370A*, consistent with previous reports in other associated phenotypes in humans and mice. This pattern is further supported by the combined analysis of all four populations (Fig. 2b).

Fig. 1 Population locations and allele frequencies. *Pie charts* present the allele frequencies of *EDARV370A*, with 370V in white and 370A in black

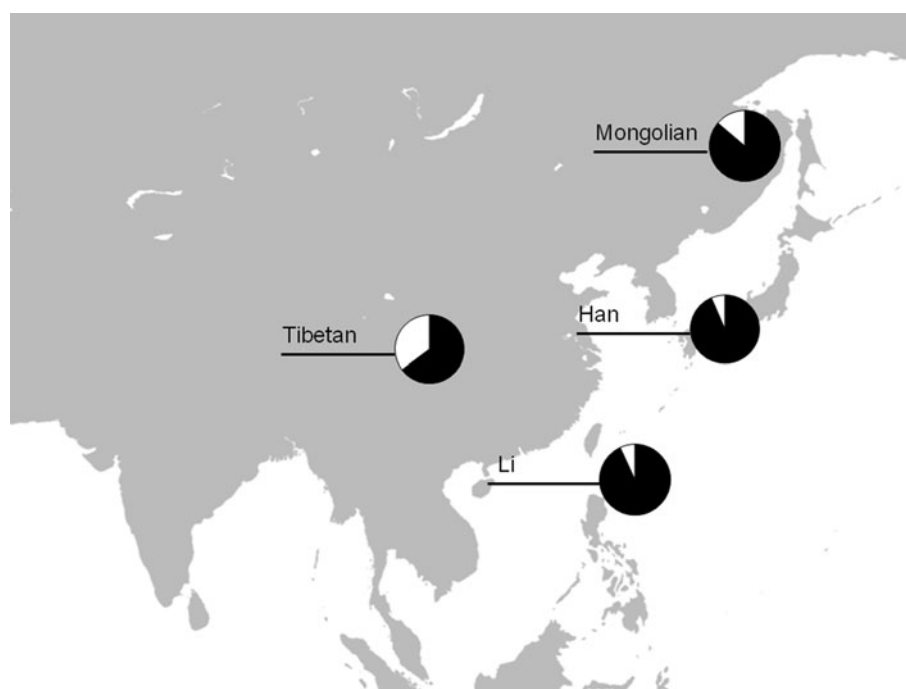


Table 1 Characteristics of the four populations

	Han	Tibetan	Mongolian	Li
Average age (range)	19.8 (17–25)	17.7 (15–22)	40.4 (14–80)	24.1 (18–58)
No. of females	579	154	164	98
No. of males	394	72	159	98
Percentage of females with straight hair (%)	85.1	76.0	90.9	80.6
Percentage of males with straight hair (%)	74.4	68.1	89.3	64.3
Percentage of females with non-straight hair (%)	14.9	24.0	9.1	19.4
Percentage of males with non-straight hair (%)	25.6	31.9	10.7	35.7

In a logistic regression model assuming an additive model, the association between straight hair and 370A is significant in three of the four populations ($p = 2.90 \times 10^{-6}$ in Han, $p = 3.07 \times 10^{-2}$ in Tibetan, $p = 1.03 \times 10^{-5}$ in Mongolian, and $p = 0.489$ in Li; Table 2). In our combined analysis, the association is even stronger ($p = 5.18 \times 10^{-10}$), with an odds ratio of 2.05. All the associations remain significant after controlling for the confounding effects of age, sex, and collection location whenever applicable (Table 2).

Discussion

This study, assisted by a large sample size, found an association between 370A and straight hair in East Asians. Unpublished data from a separate study of ~7000 Mestizos in five Latin American countries has independently verified the association between 370A and straight hair reported in this paper (Ruiz-Linares, personally communication). Previously, Fujimoto and colleagues examined the hair index [the ratio of small diameter to large diameter of the cross-sections of hair, thought to be an indicator of hair straightness (Bean 1911)] in Indonesian, Thai (Fujimoto et al. 2008a), and Japanese (Fujimoto et al. 2008b) populations, but did not find a significant association between hair index and 370A. This is likely due to limitations in power with small sample sizes (121 Indonesian, 65 Thai, and 189 Japanese). In the present study, with a much larger sample size, we observed a significant association between hair straightness and 370A. Indeed in the case of Li, the smallest of four population samples in our study (with 196 individuals) is the only one in which we did not observe a significant association between straight hair and 370A (Table 2). This is to be expected. For a population like the Li with 73.5 % straight hair cases in samples with genotype 370A, and 64 % straight hair cases in samples with the heterozygote genotype 370V/370A, the power to detect an association given a sample size of 196 (evenly distributed within two genotypes), is only 0.3. Further studies with

Fig. 2 Frequency of straight and non-straight hair types as a function of *EDARV370A* genotypes. **a** Representative images of straight and non-straight hair. **b** Frequency of straight and non-straight hair as a function of *EDARV370A* genotype in the combined samples ($n_{370A} = 1,351$; $n_{370V/370A} = 332$; $n_{370V} = 35$). **c–f** Frequency of straight and non-straight hair as a function of *EDAR370* genotype in the Han ($n_{370A} = 849$; $n_{370V/370A} = 121$), Tibetan ($n_{370A} = 95$; $n_{370V/370A} = 103$; $n_{370V} = 28$), Mongolian ($n_{370A} = 237$; $n_{370V/370A} = 83$), and Li ($n_{370A} = 170$; $n_{370V/370A} = 25$) population samples. The sample sizes of *370V* (n_{370V}) in Han, Mongolian and Li are too small, and therefore are not shown

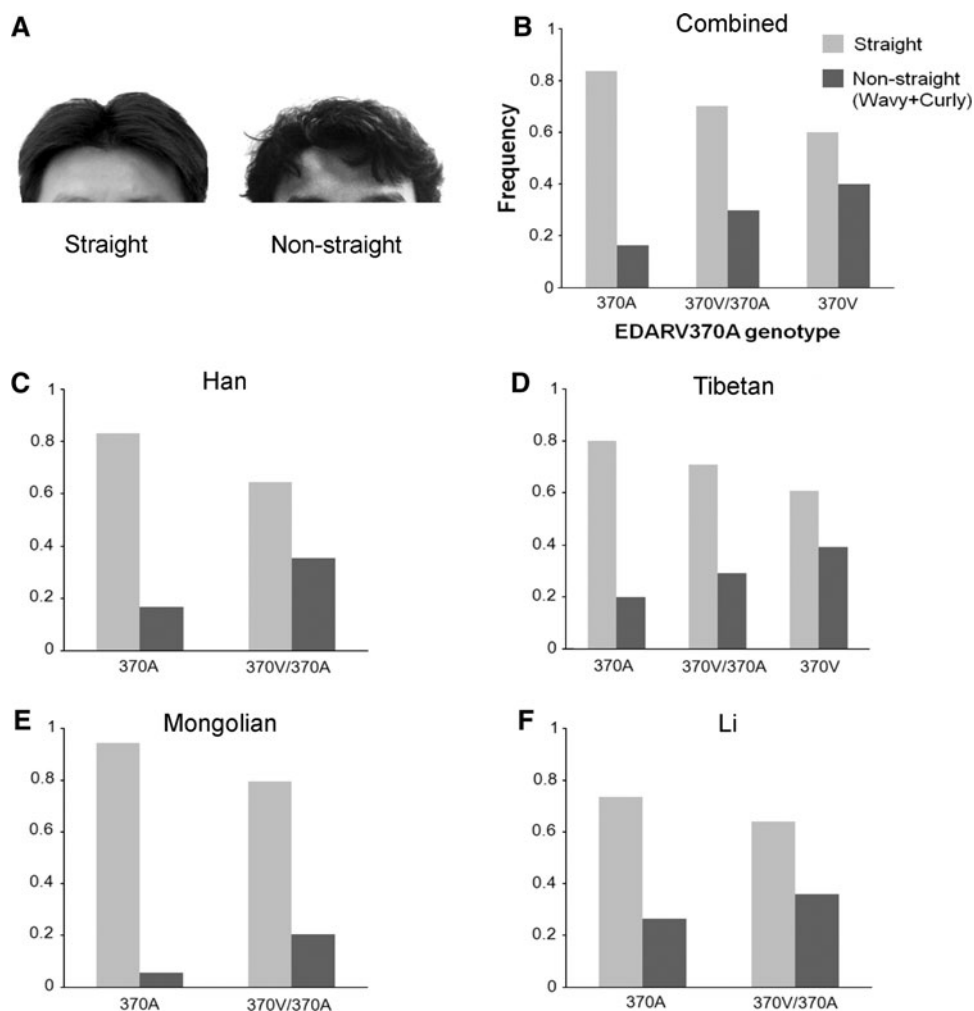


Table 2 Genotype-phenotype association of hair straightness and *EDARV370A*

	Han	Tibetan	Mongolian	Li	Combined
Sample no.	973	226	323	196	1718
Minor allele frequency (%)	6.53	35.18	13.78	6.89	11.70
Odds ratio	2.567	1.616	4.767	1.342	2.050
<i>p</i> Value	2.90×10^{-6}	3.07×10^{-2}	1.03×10^{-5}	0.489	5.18×10^{-10}
<i>p</i> Value after controlling covariates*	4.62×10^{-6}	3.53×10^{-2}	4.48×10^{-6}	0.219	1.33×10^{-8}

* We controlled for age and sex as covariates. In the combined test, we added an additional covariate, the collection location

increased sample sizes should be able to convincingly test whether *370A* is associated with straight hair in the Li population.

Previous studies indicate that hair straightness is a complex trait and that the underlying genes could be very different in distinct ethnic populations. Medland and colleagues (2009a) reported that European specific genetic variants in *Trichohyalin* gene (*TCHH*) are associated with straight hair in the Caucasians. The present study is the first unambiguous report on a specific common variant in *EDAR* associated with straight hair in Chinese populations. It has been shown that

the *EDAR370A* variant originated in Central China about 30,000 years ago (Kamberov et al. 2013), and to be absent in Europe and Africa. Thus, this *370A* variant cannot explain hair straightness in Europe and Africa. Straight hair in different populations may well involve distinctive genes, and hence distinctive biological mechanisms.

The *TCHH* variant explained ~6 % of the total variance of hair straightness in the Caucasians (Medland et al. 2009a). In our study, the *EDAR* variant explains an even smaller fraction of the total variance (in a quantitative model, the regression coefficient is approximately 0.14,

which means the association explains $\sim 2\%$ of total variance). This smaller amount of explained total variance could be partly explained by smaller sample sizes in this study than the *TCHH* study. It can also be partly contributed by the smaller variation of the hair straightness phenotype (a good majority of the East Asian samples have straight hair), and the lack of curly hair type in the studied populations. In addition, the East Asian populations included in the present study are genetically more differentiated than the Caucasians. Therefore, within East Asia, the amount of explained total variance by *370A* may vary between populations. In any case, the small amount of explained total variance in Caucasians and East Asians is a sharp contrast to the high estimated heritability of hair straightness (85–95 %) (Medland et al. 2009b). This suggests that hair straightness involves a large number of causal genes, and that the biology of hair straightness could be more complex than expected.

Our study offers new insights into the relevant selection force and biological mechanisms underlying the emergence of the *EDAR 370A* allele. Apart from being associated with hair straightness, *370A* is also associated with hair thickness (Fujimoto et al. 2008a, b), dental morphology (Kimura et al. 2009; Park et al. 2012; Kamberov et al. 2013), and sweat gland density (Kamberov et al. 2013). Since *370A* presents one of the strongest signatures of selection in humans, there is great interest in further revealing the actual selection force(s) driving the spread of this adaptive genetic variant. By reporting the association with hair straightness, we are providing a more complete understanding of the phenotypic consequences and the pleiotropic nature of *370A*. The new finding may lead to new adaptive hypothesis. For example, it is possible that the straight hair may exhibit an advantage in mate choice, and thus play a role in sexual selection (Hitsch et al. 2010). On the other hand, straight hair may facilitate the evaporation of sweat, leading to advantages in thermoregulation. Of course, another plausible scenario is that *370A* is under selection due to other phenotypic consequences (e.g. an increased density of sweat glands). In this case, hair straightness would be only a “side effect” that hitchhiked with *370A* to high frequency in East Asians, itself bearing no selective advantage or disadvantage. Whatever the adaptive trait, it is becoming clear that *EDAR 370A* allele is one of the most compelling and complex adaptive traits in humans.

Acknowledgments This project received funding support from the National Natural Science Foundation of China (31271338), Ministry of Science and Technology (2011BAI09B00), Ministry of Health (201002007), Ministry of Education (311016) to L. J.; the Packard Foundation Fellowship in Science and Engineering and an National Institutes of Health Innovator Award (1DP2OD006514-01) to P. C. S.; National Natural Science Foundation of China (31071102), Philosophy and Social Science Foundation of Shanghai

(2010BZH005) to J. T.; and National Natural Science Foundation of China (31071096), National High-Tech Research and Development Program (2012AA021802) to Y. Y. We are deeply grateful to the subjects participated in this study. S.W. acknowledges support from the Max Planck Foundation and the Chinese Academy of Sciences.

References

- Bean RB (1911) Heredity of hair form among the Filipinos. *Am Nat* 45:524
- Bryk J, Hardouin E, Pugach I, Hughes D, Strotmann R, Stoneking M et al (2008) Positive selection in East Asians for an *EDAR* allele that enhances NF- κ B activation. *PLoS One* 3:e2209
- Chang SH, Jobling S, Brennan K, Headon DJ (2009) Enhanced *Edar* signalling has pleiotropic effects on craniofacial and cutaneous glands. *PLoS One* 4:e7591
- Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, Batubara L et al (2008a) A scan for genetic determinants of human hair morphology: *EDAR* is associated with Asian hair thickness. *Hum Mol Genet* 17:835–843
- Fujimoto A, Ohashi J, Nishida N, Miyagawa T, Morishita Y, Tsunoda T et al (2008b) A replication study confirmed the *EDAR* gene to be a major contributor to population differentiation regarding head hair thickness in Asia. *Hum Genet* 124:179–185
- Grossman SR, Shlyakhter I, Karlsson EK, Byrne EH, Morales S, Frieden IJ et al (2010) A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* 327:883–886. doi:10.1126/science.1183863
- Hitsch GJ, Hortacsu A, Arieli D (2010) What makes you click? Mate preferences and matching outcomes in online dating. *Quant Mark Econ* 8:393–427
- Kamberov YG, Wang S, Tan J, Gerbault P, Wark A, Tan L et al (2013) Modeling recent human evolution in mice by expression of a selected *EDAR* variant. *Cell* 152:691–702
- Kimura R, Yamaguchi T, Takeda M, Kondo O, Toma T, Haneji K et al (2009) A common variation in *EDAR* is a genetic determinant of shovel-shaped incisors. *Am J Hum Genet* 85:528–535
- Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G et al (2009a) Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am J Hum Genet* 85:750–755
- Medland SE, Zhu G, Martin NG (2009b) Estimating the heritability of hair curliness in twins of European ancestry. *Twin Res Hum Genet* 12:514–518. doi:10.1375/twin.12.5.514
- Mou C, Thomason HA, Willan PM, Clowes C, Harris WE, Drew CF et al (2008) Enhanced ectodysplasin-A receptor (*EDAR*) signaling alters multiple fiber characteristics to produce the East Asian hair form. *Hum Mutat* 29:1405–1411
- Park JH, Yamaguchi T, Watanabe C, Kawaguchi A, Haneji K, Takeda M et al (2012) Effects of an Asian-specific nonsynonymous *EDAR* variant on multiple dental traits. *J Hum Genet* 57:508–514. doi:10.1038/jhg.2012.60
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
- Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C et al (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918. doi:10.1038/nature06250
- Xue Y, Zhang X, Huang N, Daly A, Gillson CJ, Macarthur DG et al (2009) Population differentiation as an indicator of recent positive selection in humans: an empirical evaluation. *Genetics* 183:1065–1077. doi:10.1534/genetics.109.107722