## ARTICLE

# Distribution of ancestral chromosomal segments in admixed genomes and its implications for inferring population history and admixture mapping

Wenfei Jin[1,2], Ran Li[1,2], Ying Zhou[1] and Shuhua Xu[*,1]

The ancestral chromosomal segments in admixed genomes are of significant importance for both population history inference and admixture mapping, because they essentially provide the basic information for tracking genetic events. However, the distributions of the lengths of ancestral chromosomal segments (LACS) under some admixture models remain poorly understood. Here we introduced a theoretical framework on the distribution of LACS in two representative admixture models, that is, hybrid isolation (HI) model and gradual admixture (GA) model. Although the distribution of LACS in the GA model differs from that in the HI model, we demonstrated that the mean LACS in the HI model is approximately half of that in the GA model if both admixture proportion and admixture time in the two models are identical. We showed that the theoretical framework greatly facilitated the inference and understanding of population admixture history by analyzing African-American and Mexican empirical data. In addition, we found the peak of association signatures in the HI model was much narrower and sharper than that in the GA model, indicating that the identification of putative causal allele in the HI model is more efficient than that in the GA model. Thus admixture mapping with case-only data would be a reasonable and economical choice in the HI model due to the weak background noise. However, according to our previous studies, many populations are likely to be gradually admixed and have pretty high background linkage disequilibrium. Therefore, we suggest using a case-control approach rather than a case-only approach to conduct admixture mapping to retain the statistics power in recently admixed populations.
*European Journal of Human Genetics* (2014) **22**, 930–937; doi:10.1038/ejhg.2013.265; published online 20 November 2013

## INTRODUCTION

Admixed populations arise when two or more previously mutually isolated populations start interbreeding, which is a very common phenomenon in the human evolutionary history.[1–3] As recombination breaks and rejoins DNA molecules to form new ones, a chromosome from an admixed population resembles a mosaic of ancestral chromosomal segments from different parental populations. Furthermore, the distribution of ancestral chromosomal segments can be reshaped and rearranged by recombination in each generation, which essentially provides valuable information about population history.[1,4–6] In general, the ancestral chromosomal segments from different parental populations are spliced into shorter pieces as the number of generation increases, while the ancestral chromosomal segments in a recent admixed population are usually much long due to a limited number of recombination events occurred. With the availability of high-density single-nucleotide polymorphisms data in recent years, it is now feasible to infer population history based on the length of ancestral chromosomal segments (LACS).[1] In addition, a number of methods and software, such as PCAdmix[7] and ChromoPainter,[8] have been developed to identify ancestral chromosomal segments based on the high-density genomic data.[2,9–11] Application of these methods to the empirical data has significantly increased our knowledge of population history and admixture processes.[1,5,10,12,13] However, such studies are limited to comparisons between empirical and simulated data in particular; the features of LACS distributions in some classic admixture models are still not clear.

Indeed, theoretical distribution of LACS not only benefits the inference of population history but also has many implications in admixture mapping, which is a type of mapping strategy to identify disease-associated genetic variants in recent admixed populations.[14–16] Theoretical and experimental analyses have shown that admixture mapping is a potentially promising tool for the identification of genetic variants compared with other gene mapping methods.[2,15–18] The statistical power of admixture mapping relies on the elevated linkage disequilibrium (LD) and the ancestral chromosomal segments created by population admixture.[15,19,20] Although the features of LD pattern under different admixture models and their implications in admixture mapping have been investigated in various studies,[14,20,21] the influences of LACS distribution on the signatures of association have never been explored.

In this study, we inferred the theoretical LACS distribution in two simple representative admixture models: hybrid isolation (HI)

[1]Max Planck Independent Research Group on Population Genomics, Chinese Academy of Sciences and Max Planck Society (CAS-MPG) Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China
[2]These authors contributed equally to this work.
*Correspondence: Professor S Xu, Max-Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yueyang Road, Shanghai 200031, China. Tel: +86 021 54920479; Fax: +86 21 54920451; E-mail: xushua@picb.ac.cn

model[22] in which admixture occurs only in the first generation and gradual admixture (GA) model[23] in which admixture occurs at a fixed rate in each generation. We also inferred the relationship of mean LACS in the two classic admixture models and showed that the theoretical framework in this study facilitated the inference of population history. Finally, we compared the signatures of association in admixture mapping in the two admixture models and proposed strategies to retain the high statistical power in each model. In summary, we believe the theoretical framework and data analysis in this study will facilitate admixture mapping, detection of signatures of natural selection and inference of population history in the future.

## MATERIALS AND METHODS

### Admixture models

In this study, we first attempted to explore the distribution of ancestral chromosomal segments in two typical admixture models representing two general cases of population admixture (Figure 1): HI model[22] and GA model.[23] A number of previous studies have investigated the genetic structure and LD pattern of the admixed population based on the two typical models.[20,22–24] In the admixture models, $m$ is the proportion of genetic contribution from one parental population and bounds between 0 and 1 inclusive. Here, $m$ and $1-m$ represent the genetic contribution of parental populations, pop1 and pop2, to the admixed population, respectively. $T$ represents the total number of generations since the first admixture event happened. $t$, ranging from 1 to $T$, represents the current number of generation the admixed population has experienced. In the HI model, admixture occurs only in the first generation without further genetic contribution from parental populations (Figure 1a). In the GA model, admixture occurs gradually and the parental populations contribute gene flow to the admixed population in each generation. The rates of gene flow from pop1 and pop2 to the admixed population in each generation are $m/T$ and $(1-m)/T$, respectively (Figure 1b). This ensures that the genetic contribution of a given parental population to the admixed population is constant in each generation.

### Population samples and data

In this study, we investigated the population history of African-Americans and Mexicans based on a three-way admixture model, that is, African, European and Amerindian/East Asian. Detailed information about the African-American and Mexican data sets has been described in previous studies.[1,25] Briefly, the African-American data set contained 2214 African-Americans and 268 samples from their three putative parental populations. The 268 samples representing the three putative parental populations consisted of 112 European samples (Utah residents with northern and western European ancestry from the CEPH collection; CEU), 112 African samples (Yoruba in Ibadan, Nigeria; YRI) and 44 Amerindian samples.

## RESULTS

### Theoretical distribution of LACS in the HI model

In the HI model (Figure 1a), the two parental populations, pop1 and pop2, intermarried $T$ generations ago, and their genetic contributions to the admixed population were $m$ and $1-m$, respectively. We used genetic distance (Morgan) to measure LACS. Recombination events occurring along the chromosome would follow a Poisson process at a rate of 1 in each generation if the ends of the chromosomes were ignored. Recombination occurring between two ancestral chromosomal segments, both from pop1, could be ignored if the genetic contribution of pop1 ($m$) was very small. For a particular chromosome from pop1, each recombination event would cut the ancestral chromosomal segments into smaller pieces. As a result, the ancestral chromosomal segments entering the admixed population $T$ generations ago followed an exponential distribution with mean $1/T$: $f(x; T) = Te^{-Tx}$.[4]

However, $m$ is usually not a small value in typical admixture models, and influences of recombination on ancestral chromosomal segments from the same parental population should not be ignored. We treated pop1 as the given parental population and analyzed the ancestral chromosomal segments from it to illustrate the distribution of LACS in this study. The probability that a given ancestral segment from pop1 could recombine with those from the same parental population was $m$. However, recombination among ancestral chromosomal segments from the same parental population could not change the LACS distribution, while recombination among chromosomal segments from distinct parental populations tends to cut the segments into smaller pieces. The probability that a particular ancestral segment from pop1 recombined with ancestral chromosomal segments from pop2 was $1-m$. Thus, chromosome segments from pop1 recombining with those from pop2 followed a Poisson process at a rate of $1-m$. Then after $T$ generations, the distribution of LACS from pop1 in the admixed population would follow an exponential distribution with mean $\frac{1}{(1-m)T}$, as shown below.

$$f(x; T) = (1-m)Te^{-(1-m)Tx} \tag{1}$$

Based on this formula, we found that LACS distribution in the HI model was influenced by the genetic contribution of parental population ($m$) and the number of generations ($T$). Briefly, the less the genetic contribution from parental population ($m$) the shorter the average LACS and the larger the number of generations the shorter the average LACS.

### Theoretical distribution of LACS in the GA model

In the GA model (Figure 1b), parental populations gradually contributed their genetic material to the admixed population over $T$ generations. Relative gene flow from parental populations, pop1 and pop2, to the admixed population at each generation were $m/T$ and $(1-m)/T$, respectively, with rest of the genetic materials from previous generations of the admixed population. Therefore, the final genetic contributions of pop1 and pop2 to the admixed population
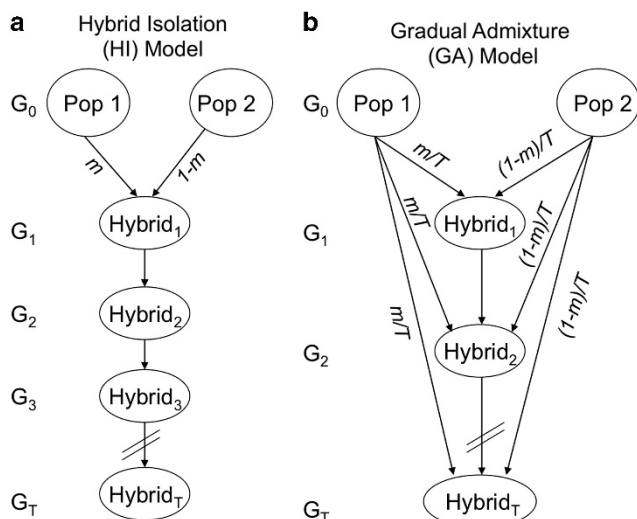


Figure 1 Schematic diagram of HI model (a) and GA model (b) for inferring the distribution of LACS. HI and GA models were adopted from previous studies.[1] In each model, the genetic contributions of pop1 and pop2 were $m$ and $1-m$, respectively. The admixed population experienced $t$ generations ranging from 1 to $T$ generation.

were $m$ and $1-m$, respectively. Similar to the HI model, we treated pop1 as the given parental population and analyzed the ancestral chromosomal segments from pop1 to illustrate the distribution of LACS. If a chromosome from pop1 entered the admixed population $t$ $(1 \le t \le T)$ generations ago, the distribution of LACS derived from this chromosome would be the same as that in the HI model: $f(x;t) = (1-m)te^{-(1-m)tx}$. If chromosome ends were ignored, a chromosome from pop1 would be expected to split into $E(k(t)) = (1-m)t$ pieces per unit length. Then the contribution of ancestral segments from pop1 to the admixed population is proportional to $(1-m)t$ after $t$ generations. Therefore, ancestral segments from pop1 included segments from different times over $T$ generations,

$$f(x;T) = \frac{\int_0^T (1-m)te^{-(1-m)tx}(1-m)tdt}{\int_0^T (1-m)tdt}$$
$$= \frac{4-2e^{-(1-m)Tx}\left((1-m)^2T^2x^2 + 2T(1-m)x + 2\right)}{(1-m)^2T^2x^3} \quad (2)$$

Based on this formula, the distribution of LACS in the GA model was observed to be influenced by the genetic contribution of parental population ($m$) and the number of generations ($T$).

**Consistency between theoretical distributions and simulated data**
We simulated data under several scenarios with forward-time simulation program (Supplementary Data). The genetic contribution of a given parental population to the admixed population was set at

50%. We assumed that the effective population size ($Ne$) of parental and admixed populations was 5000 each, although $Ne$ would not influence the distribution.[1] The number of generations since admixture was set at 10, 20, 50 and 100.

For comparisons, we have displayed the theoretical and simulated LACS distributions in the same figure (Figure 2). As shown, the theoretical distribution was essentially consistent with the simulated distribution in both the HI and GA models in all scenarios (Figures 2a and c). Moreover, we observed that the theoretical LACS distribution with 100 generations almost perfectly fit the simulated data based on Quantile-Quantile (Q-Q) plot (Supplementary Figures S1 and S2). Further analysis showed no significant difference between theoretical and simulated distributions ($P > 0.05$, Kolmogorov–Smirnov tests[26]). Further analysis showed that the simulated data lacked long ancestral chromosomal segments compared with theoretical distribution when the number of generation was short ($t < 10$). These differences were essentially caused by assuming an infinite chromosome length in theoretical distribution, while the simulated data were based on real length of human chromosomes with fixed and finite lengths. With $t = 1$, LACS in theoretical distribution was ill-defined, and the distribution become more accurate when $t$ became larger. We also found that the distributions of LACS among different generations were significantly different from each other in each tested model. ($P < 10^{-16}$, Kolmogorov–Smirnov tests[26]). The difference of distributions observed among different generations was more pronounced when the 10-based logarithm of LACS was obtained (Figures 2c and d).
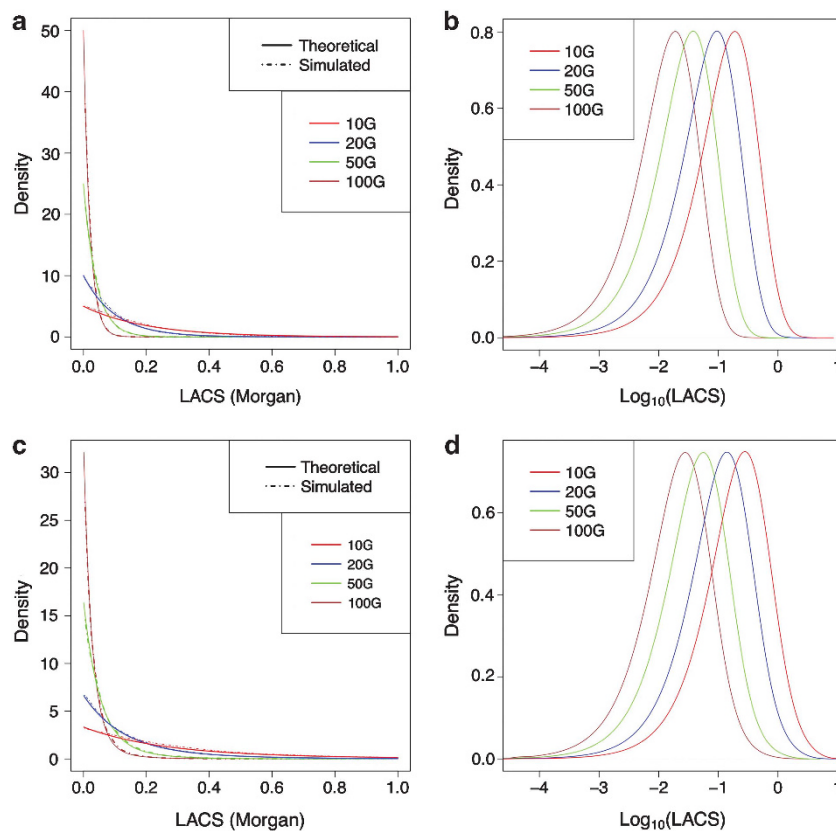


**Figure 2** Distribution of LACS in the HI and GA models. The genetic contribution from parental populations to the admixed population was assumed to be 50%. The number of generations since admixture was assumed to be 10, 20, 50 and 100, respectively. *G* denotes number of generations since admixture. Solid lines represent theoretical distribution and dashed lines represent simulated distribution. Distribution of LACS in the HI (**a**) and GA (**c**) models. Distribution of Log$_{10}$ (LACS) in the HI (**b**) and GA (**d**) models.

### Comparison of LACS distributions between HI and GA models

Although we deduced the theoretical distribution of ancestral chromosomal segments in both the HI and GA models separately, we were also interested in the relationship between these distributions. As the mean of LCAS in the HI and GA models were $E(x; T) = \frac{1}{(1-m)T}$ and $\hat{E}(x; T) = \frac{2}{(1-m)T}$ (see Supplementary Data), respectively, we could observe that the mean LACS of a given parental population in the GA model is twice as that in the HI model if both the number of generation ($T$) and genetic contribution ($m$) in the two models were identical, which also indicated that the mean LACS in the GA model was equal to that in the HI model with half the admixture time.

As the variance of LCAS in the HI and GA models are $D(x; T) = \frac{1}{(1-m)^2 T^2}$ and $\hat{D}(x; T) \approx \frac{4(\sum_{t=1}^{t=T} \frac{1}{t} - 1)}{(1-m)^2 T^2}$ (see Supplementary Data), respectively, we could observe that the variance of LACS in the GA model was higher than that in the HI model if $T > 1$ (GA model assumes at least two generations since admixture). This was reasonable considering that admixed population in the GA model contained both the long ancestral chromosomal segments that entered the admixed population recently and short ancestral chromosomal segments that entered much earlier. Besides, the mean and SD of LACS in the HI model were identical (Figure 3a) as they followed an exponential distribution, while the SD of LACS in the GA model was larger than the mean. As variance of LACS distribution in the GA model was larger than that in the HI model with the same generation (Figure 3a), we conjectured the LACS distribution in the GA model

could be flatter than that in the HI model, which was also supported by the observations (eg, Figure 3b). Although the mean LACS in the GA model was the same as that in the HI model with half the number of generation, there was a much higher proportion of long ancestral chromosomal segments in the GA model compared with that in the HI model (eg, Figure 3b).

After investigation of the overall pattern of LACS distribution in both the HI and GA models, we further examined LACS distribution in a specific genomic region in the two typical models. A genetic locus was randomly selected from a gradually admixed population (admixed population under the GA model) and a hybrid-isolated population (admixed population under the HI model). We found that the genetic contribution of loci transmitted with the given locus along the chromosome decreased quickly as the distance to the given locus increased in the hybrid-isolated population, whereas it decreased much slower in the gradually admixed population compared with that in hybrid-isolated population (Figures 3c and d). These results indicated that different admixture dynamics could have a strong influence on the pattern of local ancestral chromosomal segments. In particular, the change of genetic contribution of loci transmitted with the given locus in the gradually admixed population became slower as the distance to the locus increased. The genetic contribution of a given parental population to the admixed population was hardly reduced to 0, because some recent ancestral chromosomal segments from the given parental population were very long and even spread through the whole chromosome in the admixed population (Figure 3d).
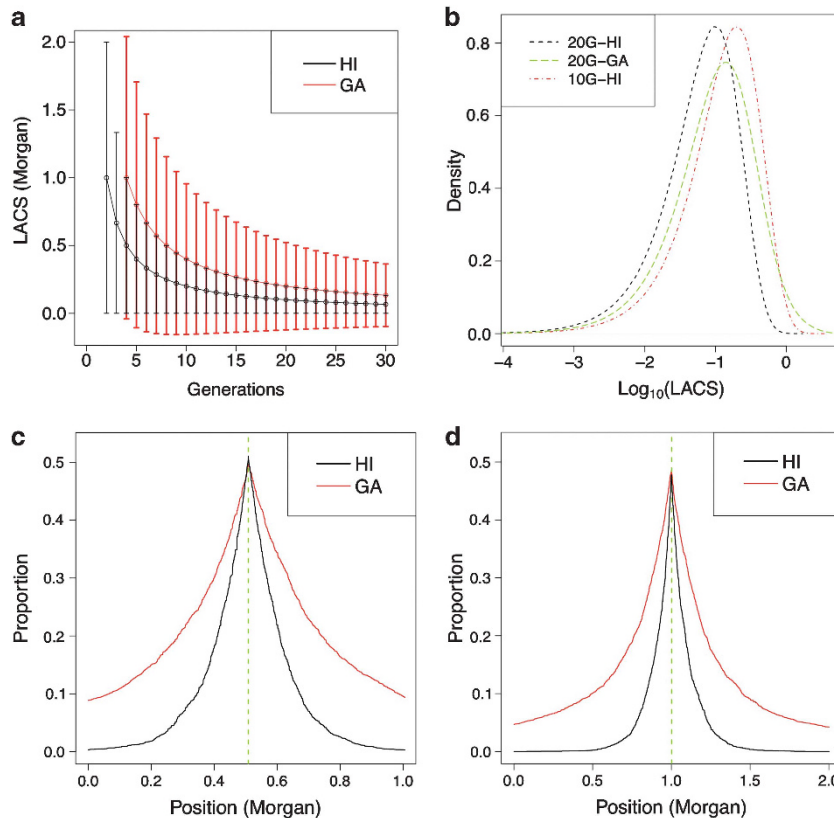


**Figure 3** Comparison of LACS distributions between HI and GA models. (**a**) Comparison of mean and SD of LACS between HI and GA models. Error bars and circles represent SD and mean, respectively. (**b**) Comparison of LACS distribution between HI and GA models when the number of generation is 10 or 20. (**c**) The change of genetic contribution that transmitted with the given locus along the chromosome in the HI and GA models. (**d**) The change of genetic contribution transmitted with the given locus along the chromosome in the HI and GA models in a longer chromosome. Green vertical dashed line represents the reference locus.

## LACS distributions facilitated inference of African-Americans population history

Thus far, previous studies on African-Americans focusing on either population history or admixture mapping have deliberately ignored the Amerindian ancestral component.[27–30] Recent studies, however, have shown considerable fractions of Amerindian ancestral component in many African-American individuals.[1,13] Using STRUCTURE,[31,32] we estimated the genetic contribution of African, European and Amerindian/East Asian populations to the 2114 African-Americans to be 0.725, 0.245 and 0.029, respectively ($K=3$). The genetic contribution of Amerindians to African-Americans observed in our study was similar to that in a previous study using African-American samples from Washington, DC, USA.[33]

We identified the ancestral chromosomal segments of each parental population in the 2114 African-Americans and considered only the ancestral chromosomal segments with length $>0.005$ Morgan. Among the three parental populations, the African ancestral component contained the highest proportion of long ancestral chromosomal segments, while Amerindian ancestral component contained the highest proportion of short ancestral chromosomal segments (Figure 4a), which suggested that the less the genetic contribution the shorter the ancestral chromosomal segments. For Amerindian ancestral component, the mean LACS in the 13-generation HI model was very close to the empirical value while the mean LACS in the GA model required $>20$ generations (400 years, assuming 20 years per generation) to reach the empirical value (Figure 4b), which contradicted with the recorded history that most African ancestors arrived in America after eighteenth century. Therefore, we concluded that the admixture of Amerindians with other parental populations was likely to be similar to the 13-generation HI model.

Regarding the European and African ancestral components, the means of LACS in both 7-generation HI and 13-generation GA models were very close to the empirical values (Figures 4c and d). Although it is difficult to distinguish a 13-generation GA model from a 7-geneation HI model based on the empirical distribution of African ancestral component (Supplementary Figure S3), the empirical distribution of European ancestral component showed high proportion of long LCAS and was concordant with the 13-generation GA model (Supplementary Figure S4). In addition, the recorded history of African-Americans is much longer than 7 generations,[1,25,34] which supported a 13-generation GA model to explain the primary admixture pattern of African and European ancestral components. However, the number of generations indicated by the mean and the SD of LACS were not perfectly consistent for each parental population (Figure 4), suggesting the complicated population admixture history of African-Americans.

Finally, we have proposed a model to explain the primary admixture pattern of the three parental populations of Urban African-Americans as explained below. The primary African ancestors of current African-Americans arrived in North America in the
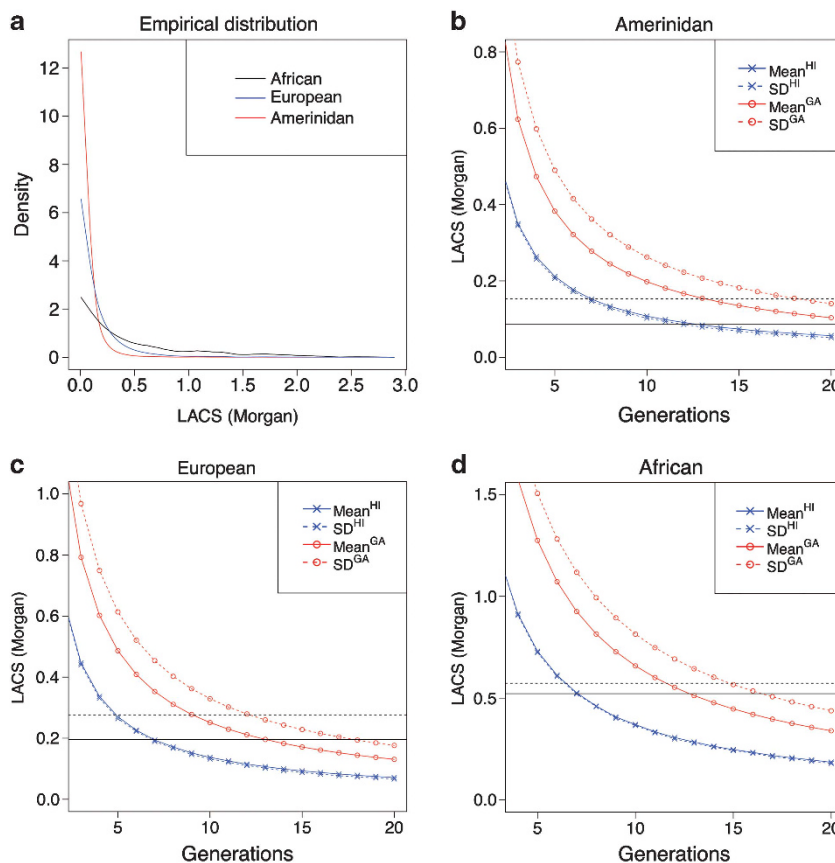


**Figure 4** Admixture history of African-Americans based on three-way admixture. (**a**) Distributions of LACS for African, European and Amerindian ancestral components in African-American. (**b**) Mean and SD of LACS for Amerindian ancestral component and theoretical distribution in the HI and GA models. (**c**) Mean and SD of LACS for European ancestral component and theoretical distribution in the HI and GA models. (**d**) Mean and SD of LACS for African ancestral component and theoretical distributions in the HI and GA models. Black solid and dashed lines represent the empirical mean and SD of LACS, respectively.

eighteenth century and interbred with the Amerindians who were laborers in the southeast European colonies. However, there was no further Amerindian gene flow after the end of the Native American slave trade around 1730.[35,36] Therefore, we proposed that the admixture of Amerindian with other parental populations was more likely to follow the HI model. Then the Europeans gradually interbred with African-Americans generation after generation.

### LACS distributions facilitated the inference of Mexican population history

Although we systematically analyzed the population admixture history of Mexicans using European and African ancestral components,[1] we expected to infer a detailed population history when all three continental ancestral components (Africans, Europeans and Amerindians) were considered simultaneously. Using STRUCTURE,[31,32] we inferred that the genetic contributions of African, European and Amerindian populations to the 458 Mexicans to be 0.033, 0.451 and 0.517, respectively ($K=3$). We identified ancestral chromosomal segments of each parental population in the 458 Mexicans and considered only the ancestral chromosomal segments with length $>0.01$ Morgan. Further analysis showed that the distributions of Amerindian and European ancestral components were almost identical (Figure 5a). Both Amerindian and European ancestral components contained a higher proportion of long ancestral chromosomal segments compared with African ancestral component (Figure 5a).

In the HI model, the estimated number of generations of Amerindian, European and African ancestral components based on the mean LACS was 11, 11 and 10, respectively (Figure 5). In contrast, in the GA model, the estimated number of generations of Amerindian, European and African ancestral components based on the mean LACS was 22, 21 and 20, respectively (Figure 5). Further analysis showed the empirical LACS distributions of European and Amerindian ancestral components were concordant with the GA model (Supplementary Figures S5 and S6). Therefore, formation of Mexican could be explained by a 22-generation GA model between European and Amerindian at the beginning, and subsequently, African further admixed with the admixed population. The model was essentially consistent with the history of Mexican as reported.

### Influence of LACS distribution on admixture mapping

As different admixture dynamics can result in different LACS distribution, it was necessary to elucidate the influence of LACS distribution on admixture mapping. We believe this effort could help improve the statistical power in identifying disease-associated genetic variants using admixture mapping. Therefore, we conducted a series of simulations to examine the signatures of association in admixture mapping (see Supplementary Data). We compared the pros and cons of case-only and case-control designs in gradually admixed populations and hybrid-isolated populations.

To compare the admixture mapping in the HI and GA models, we used identical parameters for simulations by varying only the
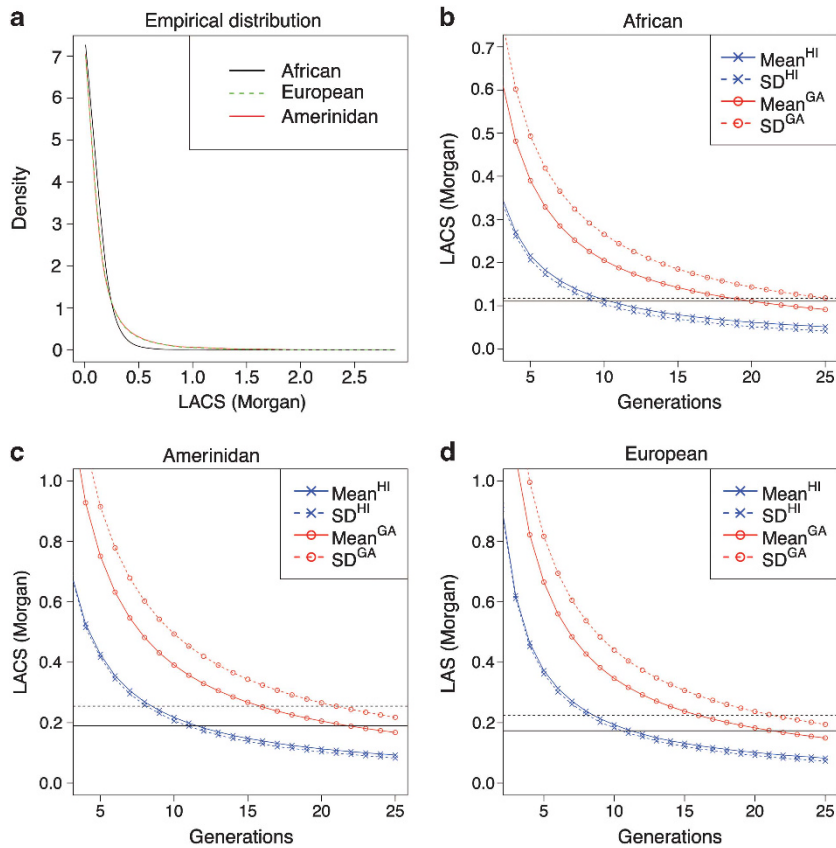


**Figure 5** Admixture history of Mexicans based on three-way admixture. (**a**) Distribution of LACS for African, European and Amerindian ancestral components in Mexican. (**b**) Mean and SD of LACS for African ancestral component and theoretical distributions in the HI and GA models. (**c**) Mean and SD of LACS for Amerindian ancestral component and theoretical distributions in the HI and GA models. (**d**) Mean and SD of LACS for European ancestral component and theoretical distributions in the HI and GA models. Black solid and dashed lines represent the empirical mean and SD of LACS, respectively.

admixture model. Therefore, the main difference in the signature of association between hybrid-isolated population and gradually admixed population should result from different admixture dynamics. In each study, we simulated 2000 cases and 2000 controls for admixture mapping, with genetic contribution of the given parental population to the admixed population $\theta = 20\%$, number of generations since the admixture $\lambda = 20$ and the increased risk of 2 for containing alleles from the given parental population at the susceptibility locus. Although the highest ancestral deviations at the susceptibility locus in both hybrid-isolated population and gradually admixed population were identical (40%), we found that the peak of association in the hybrid-isolated population was narrower and sharper than that in gradually admixed population (Figure 6), indicating that the identification of putative causal allele in hybrid-isolated population could be more efficient than that in gradually admixed population. In contrast, the peak of signatures in gradually admixed population was wider than that in hybrid-isolated population (Figure 6), indicating that admixture mapping on a genome-wide scale in gradually admixed population required fewer markers than that in hybrid-isolated population.

In the case-control designed admixture mapping, P-values were calculated by comparing the deviation of genetic contribution in cases with that in controls through phenotype-association analysis.[17,18] P-values of admixture mapping were determined from the ancestral deviation between cases and controls. Therefore, the P-value of susceptibility locus in the GA model could be the same as that in HI, because the highest ancestral deviations in both the HI and GA models were identical. In contrast, the P-values in case-only designed admixture mapping were calculated based on the empirical distribution of LACS in cases.[17,18] As the distribution of signatures in the GA model was wider than that in the HI model, the P-value in the GA model could be larger than that in the HI model, which indicated that the signatures in HI were more likely to be significant compared with those in the GA model. Therefore, we suggest case-control-designed study rather than case-only-designed study to improve the statistical power of admixture mapping in gradually
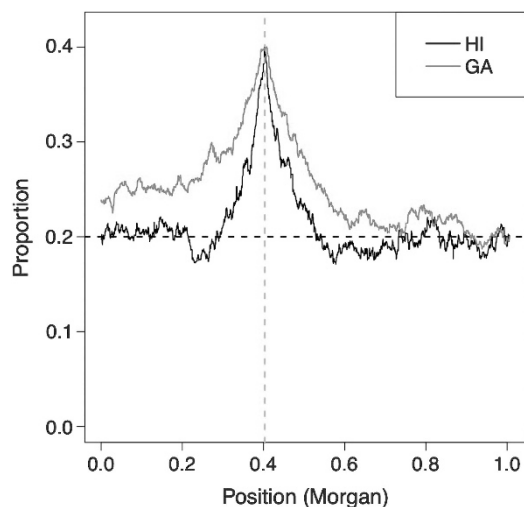
admixed populations. In contrast, we suggest the case-only-designed admixture mapping in hybrid-isolated populations to reduce the cost.

## DISCUSSION

Inference of population history is a fundamental topic in the field of population genetics. A number of studies have developed various methods to identify ancestral chromosomal segments (or the number of recombination breakpoints between different ancestries) for inferring dates of admixture.[7–11] Recently, Moorjani et al.[34] developed ROLLOFF, which infers dates of admixture by calculating the decrease of admixture LD with genetics distance in the admixed population. Although ROLLOFF performs well on one-pulse admixture (eg, HI model), it is likely to underestimate the time of admixtures in multiple-pulses gene exchanges (eg, GA model). We provided the theoretical LACS distribution under both the HI and GA models, which could facilitate the data simulation and inference of population history. As more accurate approaches to identify ancestral chromosomal segments are being developed, the power to infer population history using our approach is expected to be improved in the future.

When both the admixture proportion and number of generations since admixture were identical in the HI and GA models, we found that the mean LACS in the GA model was twice of that in the HI model, which also advanced our understanding of the population admixture history. For example, our study showed that the number of generations since admixture of African and European ancestral components in African-Americans based on the GA model was 13, indicating that the mean LACS of African-Americans could be the same as that in the HI model with 6–7 generations. This could explain why the number of generations of African-Americans was estimated to be 6–8 in previous studies that did not consider the admixture process.[1,10,19,30,34] Based on the GA model, the number of generations for Amerindian and European ancestral components in Mexicans was estimated to be 21–22 in this study, which was almost twofold higher than those in previous studies that did not consider the admixture process,[37–40] which could also be easily explained by our theoretical framework.

Although it is contradicting to assume an infinite length of chromosomes in the theoretical framework, the theoretical LACS distribution was consistent with the simulated data when the number of generations was not very small. In summary, we found that the theoretical LACS distribution we deduced essentially fit the forward-time simulated distributions in each scenario, which supported and validated the theoretical framework developed in this study. As demonstrated in this study, the theoretical LACS distribution could greatly facilitate the inference of population admixture history in the future. For the implication of LACS distribution in admixture mapping, we suggest corresponding admixture mapping strategies for populations under different admixture models. In this study, we deduced only the theoretical distribution of ancestral chromosomal segments under two typical models. In future, identification of the theoretical distribution of ancestral chromosomal segments under more complex admixture models could be useful to infer more complex population history. Furthermore, we believe the theoretical distribution inferred in this study will have more extensive applications in the future.



**Figure 6** Signatures of association in admixture mapping in the HI and GA models. Vertical dashed green line represents the susceptibility locus. Heretical dashed black line represents the theoretical mean of genetic contribution of the given parental population.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Jin W, Wang S, Wang H, Jin L, Xu S: Exploring population admixture dynamics *via* empirical and simulated genome-wide distribution of ancestral chromosomal segments. *Am J Hum Genet* 2012; **91**: 849–862.

2 Seldin MF, Pasaniuc B, Price AL: New approaches to disease mapping in admixed populations. *Nat Rev Genet* 2011; **12**: 523–528.

3 Verdu P, Rosenberg NA: A general mechanistic model for admixture histories of hybrid populations. *Genetics* 2011; **189**: 1413–1426.

4 Pool JE, Nielsen R: Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* 2009; **181**: 711–719.

5 Gravel S: Population genetics models of local ancestry. *Genetics* 2012; **191**: 607–619.

6 Pugach I, Matveyev R, Wollstein A, Kayser M, Stoneking M: Dating the age of admixture *via* wavelet transform analysis of genome-wide data. *Genome Biol* 2011; **12**: R19.

7 Brisbin A, Bryc K, Byrnes J *et al*: PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum Biol* 2012; **84**: 343–364.

8 Lawson DJ, Hellenthal G, Myers S, Falush D: Inference of population structure using dense haplotype data. *PLoS Genet* 2012; **8**: e1002453.

9 Tang H, Coram M, Wang P, Zhu X, Risch N: Reconstructing genetic ancestry blocks in admixed individuals. *Am J Hum Genet* 2006; **79**: 1–12.

10 Price AL, Tandon A, Patterson N *et al*: Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet* 2009; **5**: e1000519.

11 Sankararaman S, Kimmel G, Halperin E, Jordan MI: On the inference of ancestries in admixed populations. *Genome Res* 2008; **18**: 668–675.

12 Zakharia F, Basu A, Absher D *et al*: Characterizing the admixed African ancestry of African Americans. *Genome Biol* 2009; **10**: R141.

13 Kidd JM, Gravel S, Byrnes J *et al*: Population genetic inference from personal genome data: impact of ancestry and admixture on human genomic variation. *Am J Hum Genet* 2012; **91**: 660–671.

14 Chakraborty R, Weiss KM: Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci USA* 1988; **85**: 9119–9123.

15 McKeigue PM: Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. *Am J Hum Genet* 1997; **60**: 188–196.

16 McKeigue PM: Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. *Am J Hum Genet* 1998; **63**: 241–251.

17 Montana G, Pritchard JK: Statistical tests for admixture mapping with case-control and cases-only data. *Am J Hum Genet* 2004; **75**: 771–789.

18 Patterson N, Hattangadi N, Lane B *et al*: Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* 2004; **74**: 979–1000.

19 Seldin MF, Morii T, Collins-Schramm HE *et al*: Putative ancestral origins of chromosomal segments in individual African Americans: implications for admixture mapping. *Genome Res* 2004; **14**: 1076–1084.

20 Pfaff CL, Parra EJ, Bonilla C *et al*: Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am J Hum Genet* 2001; **68**: 198–207.

21 Stephens JC, Briscoe D, O'Brien SJ: Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *Am J Hum Genet* 1994; **55**: 809–824.

22 Long JC: The genetic structure of admixed populations. *Genetics* 1991; **127**: 417–428.

23 Ewens WJ, Spielman RS: The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet* 1995; **57**: 455–464.

24 Guo W, Fung WK: The admixture linkage disequilibrium and genetic linkage inference on the gradual admixture population. *Yi Chuan Xue Bao* 2006; **33**: 12–18.

25 Jin W, Xu S, Wang H *et al*: Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Res* 2012; **22**: 519–527.

26 Lilliefo HW: On Kolmogorov–Smirnov test for normality with mean and variance unknown. *J Am Stat Assoc* 1967; **62**: 399–402.

27 Zhu X, Luke A, Cooper RS *et al*: Admixture mapping for hypertension loci with genome-scan markers. *Nat Genet* 2005; **37**: 177–181.

28 Freedman ML, Haiman CA, Patterson N *et al*: Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci USA* 2006; **103**: 14068–14073.

29 Smith MW, Patterson N, Lautenberger JA *et al*: A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet* 2004; **74**: 1001–1013.

30 Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF: A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. *Am J Hum Genet* 2006; **79**: 640–649.

31 Falush D, Stephens M, Pritchard JK: Inference of population structure using multi-locus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003; **164**: 1567–1587.

32 Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000; **155**: 945–959.

33 Shriver MD, Parra EJ, Dios S *et al*: Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet* 2003; **112**: 387–399.

34 Moorjani P, Patterson N, Hirschhorn JN *et al*: The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet* 2011; **7**: e1001373.

35 Gallay A: *The Indian Slave Trade: The Rise of the English Empire in the American South 1670–1717*. New Haven, CT, USA: Yale University Press, 2002.

36 Seybert T: Slavery and Native Americans in British North America and the United States: 1600 to 1865. New York Life, 2004.

37 Tian C, Hinds DA, Shigeta R *et al*: A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. *Am J Hum Genet* 2007; **80**: 1014–1023.

38 Wang S, Ray N, Rojas W *et al*: Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet* 2008; **4**: e1000037.

39 Price AL, Patterson N, Yu F *et al*: A genomewide admixture map for Latino populations. *Am J Hum Genet* 2007; **80**: 1024–1036.

40 Johnson NA, Coram MA, Shriver MD *et al*: Ancestral components of admixed genomes in a Mexican cohort. *PLoS Genet* 2011; **7**: e1002410.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)