

ARTICLE

# The role of *GHR* and *IGF1* genes in the genetic determination of African pygmies' short stature

Noémie SA Becker<sup>\*1</sup>, Paul Verdu<sup>2</sup>, Myriam Georges<sup>1</sup>, Philippe Duquesnoy<sup>3</sup>, Alain Froment<sup>1,4</sup>, Serge Amselem<sup>3</sup>, Yves Le Bouc<sup>5</sup> and Evelyne Heyer<sup>1</sup>

African pygmies are at the lower extreme of human variation in adult stature and many evolutionary hypotheses have been proposed to explain this phenotype. We showed in a recent study that the difference in average stature of about 10 cm observed between contemporary pygmies and neighboring non-pygmies has a genetic component. Nevertheless, the genetic basis of African pygmies' short stature remains unknown. Using a candidate-gene approach, we show that intronic polymorphisms in GH receptor (*GHR*) and insulin-like growth factor 1 (*IGF1*) genes present outlying values of the genetic distance between Baka pygmies and their non-pygmy Nzimé neighbors. We further show that *GHR* and *IGF1* genes have experienced divergent natural selection pressures between pygmies and non-pygmies throughout evolution. In addition, these SNPs are associated with stature in a sample composed of 60 pygmies and 30 non-pygmies and this association remains significant when correcting for population structure for the *GHR* locus. We conclude that the *GHR* and *IGF1* genes may have a role in African pygmies' short stature. The use of phenotypically contrasted populations is a promising strategy to identify new variants associated with complex traits in humans.

European Journal of Human Genetics advance online publication, 10 October 2012; doi:10.1038/ejhg.2012.223

**Keywords:** pygmies; height; human evolution

## INTRODUCTION

There is great variation in stature among human populations, and African pygmies are known as the shortest human populations: the mean male stature varies from 141.7 (Mbuti from the Democratic Republic of Congo) to 160.9 cm (Twa from the Democratic Republic of Congo) across pygmy populations, compared to 166.8 cm on average across sub-Saharan non-pygmy populations.<sup>1</sup> Previous studies have shown that the divergence between ancestral pygmy populations and ancestral non-pygmy populations is quite ancient and occurred around 60 000 years ago.<sup>2,3</sup> Many evolutionary hypotheses have been proposed to explain this phenotype,<sup>4</sup> which could be due to an adaptation to food scarcity,<sup>5</sup> to the hot and humid climate of the rainforest,<sup>6</sup> to the density of this environment where mobility is uneasy,<sup>7</sup> to a life-history tradeoff in a context of high mortality<sup>8,9</sup> or to particular mate choices.<sup>10</sup> However, none of these hypotheses has proven to be true<sup>4,11</sup> and the evolution of the short stature of African pygmies remains largely unknown. In a recent paper, we showed that phenotypic plasticity alone cannot explain the differential phenotype between pygmies and neighboring non-pygmies and that unknown genetic factors are probably involved in the determination of pygmies' stature.<sup>12</sup>

In the search for the molecular mechanism underlying the pygmies' short stature, endocrinologists have shown that this phenotype is characterized by a normal serum level of the growth hormone (GH), but a lower level of insulin-like growth factor 1 (IGF1) and of GH binding protein (GHBP) as compared to controls.<sup>13–17</sup> Furthermore,

a recent study showed that the expression of the GH receptor gene (*GHR*, which is the same gene as *GHBP*) is 8 times lower in pygmy subjects, as compared with neighboring non-pygmies.<sup>18</sup> Small regions of the *GHR* and *IGF1* genes have been previously tested for association with the stature of African pygmies with negative results.<sup>18–21</sup> A recent genome-wide study has identified regions in the genome that might be involved in the difference in stature observed between pygmies from Cameroon and their non-pygmy neighbors, although this signal was not significant.<sup>22</sup> Some of these regions contain genes linked to the GH–IGF1 signaling pathway.

The aim of this study is to test for the implication of four candidate genes (*GHR*, *IGF1*, and the signal transducers and activators of transcription *STAT5A* and *STAT5B*) in the short stature of African pygmies, using a population genetics approach and phenotypic data. The genes under study have been chosen in view of the studies mentioned above and all the four participate in the GH–IGF1 signaling pathway (see Materials and methods).

## MATERIALS AND METHODS

### Study design and choice of the candidate genes

Several genes are known to be involved in height variation in humans, among which the genes of the GH response cascade have a major role in prenatal and postnatal growth.<sup>23,24</sup> According to endocrinological assays, pygmy individuals have normal levels of GH but lower serum levels of IGF1, and exhibit an eightfold underexpression of the *GHR* gene as compared with the controls.<sup>16,18,19</sup> Therefore, the *GHR* and *IGF1* genes represent good

<sup>1</sup>CNRS-MNHN-Université Paris 7, UMR 7206 Ecoanthropology and Ethnobiology, Paris, France; <sup>2</sup>Department of Biology, Stanford University, Stanford, CA, USA; <sup>3</sup>INSERM U.933 and Université Pierre et Marie Curie, Paris, France; <sup>4</sup>IRD-MNHN, UMR 208 'Patrimoines locaux', Paris, France; <sup>5</sup>Centre de Recherche UMRS.938 Inserm-UPMC, Paris, France

\*Correspondence: Dr NSA Becker, CNRS-MNHN-Université Paris 7, UMR 7206 Ecoanthropology and Ethnobiology, 57, rue Cuvier, 75231 Paris, France. Tel: +33 1 40 79 30 51; Fax: +33 1 79 32 31; E-mail: nbecker@mnhn.fr

Received 5 April 2012; revised 18 July 2012; accepted 6 September 2012

candidates to explain pygmies' short stature. We develop here a candidate-gene approach focusing on these two genes and on the signal transducers and activators of transcription *STAT5A* and *STAT5B*. The latter genes code for two proteins mediating the signal from the activated GHR to the nucleus and inducing the synthesis of various proteins including IGF1. The four candidate genes are rather long (from 24.4 kb for *STAT5A* to 297.9 kb for *GHR*) and we had limited DNA quantities at our disposal for pygmies and non-pygmies. We therefore chose to re-sequence only portions of each one of these genes. Taking into account the fine genetic map of the genes (see Supplementary Figure S1), these portions were chosen in order to cover most of the genes' variability through linkage disequilibrium (Supplementary Methods and Supplementary Table S1).

### Sampling populations

The re-sequencing was performed for 90 unrelated individuals (Table 1) of known adult height from two pygmy populations: the Baka from South-Eastern Cameroon (16 men and 14 women; mean height in the population: 155.6 cm for men and 146.5 cm for women), the taller and more admixed Bongo from South-Eastern Gabon<sup>2,12</sup> (19 men and 11 women; 159.7 cm and 151.2 cm) and from the non-pygmy Nzimé (6 men and 24 women; 168.5 cm and 155.8 cm), who are the immediate neighbors of the Baka.

Blood (Baka and Nzimé) or saliva (Bongo) was sampled from each participant and stature was measured with a height gauge according to standard procedures.<sup>25</sup> Mean height in the populations and in our samples are presented in Table 1.

Oral and video-recorded informed consent was obtained from each donor and the sampling program was accepted by the Cameroonian and Gabonese ethics committees, as required by the French laws of bioethics.

### Amplification and analysis of the sequences

Two to three sequences in each candidate gene were chosen using information on the linkage disequilibrium in the region and on previously published sequences (see Supplementary Methods). Most of the primers were designed using Primer 3<sup>26</sup> and others were adapted from previous studies.<sup>21,27,28</sup> Primer sequences can be found in Supplementary Table S1. Amplifications were performed using standard PCR conditions (available upon request). PCR products were analyzed in an Applied Biosystems (Carlsbad, CA, USA) 3100 or 3130 XL automated sequencer. The sequences obtained were aligned with GenAlign 3.3 software.<sup>29</sup> All sequences are available upon request.

### Population genetics analyses

Haplotypes were reconstructed from unphased data using PHASE 2.1 software.<sup>30,31</sup>

Summary statistics were calculated using DNAsp 50.1 software<sup>32</sup> and Arlequin 3.11 software<sup>33</sup> (nucleotide diversity  $\pi$ <sup>34</sup> can be found in Supplementary Table S3). No significant departures from Hardy–Weinberg equilibrium were detected for any of the SNPs discovered in any of the three populations under study.

Genetic distances between pairs of populations (unilocus and multilocus  $F_{ST}$ <sup>35</sup> in Supplementary Table S2) were estimated using Arlequin 3.11

software.<sup>33</sup> Pairwise  $F_{ST}$   $P$ -values were estimated using a standard permutation procedure considering 10 000 haplotype permutations between individuals across populations.

To evaluate whether the  $F_{ST}$  values observed in our sequences between the two extreme populations (Baka pygmies and Nzimé non-pygmies) were significantly different from neutral expectations, we performed a test proposed by Beaumont and Nichols.<sup>36</sup> This test compares the  $F_{ST}$  values, conditional on heterozygosity ( $H_e$  estimated as the average pairwise difference between all possible pairs of genes between each sample<sup>36</sup>), obtained for each polymorphism, with a distribution of values obtained on data sets simulated under a neutral model. The test was performed using a version of the Dfdist software modified for bi-allelic co-dominant markers.<sup>37</sup> Following Segurel *et al.*,<sup>37</sup> we performed 500 000 coalescent simulations with a two-deme island model, the maximum frequency of the most common allele set to 0.99 and  $\theta = 2nN\mu = 0.01$ , where  $n = 2$  is the number of demes of size  $N$  and  $\mu$  is the mutation rate. This particular  $\theta$  value corresponds to  $N = 125$  and  $\mu = 2 \times 10^{-5}$ . Different values of  $\theta$  (0.1 and 0.001) were also tested and gave very similar results (Table 2). These neutral data sets had to be simulated with a mean  $F_{ST}$  value equal to that estimated on neutral sequences. No neutral sequence-based  $F_{ST}$  estimates were available between the Baka and the non-pygmy Nzimé here under study, but  $F_{ST}$  values between the Baka pygmies and two different non-pygmy populations from Western Central Africa have been published elsewhere<sup>25</sup> ( $F_{ST} = 0.033$  with the Gabonese Akele and 0.038 with the Cameroonian Ngumba). Moreover, all non-pygmy populations of this region, including the Nzimé, exhibit reduced pairwise genetic distances.<sup>2,3</sup> The mean  $F_{ST}$  value was thus set to 0.033, corresponding to the genetic distance between the Baka pygmies and the non-pygmy Gabonese Akele.<sup>3</sup> The same simulations were also performed using a mean  $F_{ST}$  value of 0.038 (corresponding to the distance between the Baka pygmies and the Cameroonian Ngumba) and gave similar results (Table 2). The  $P$ -value for the joint ( $F_{ST}$ ,  $H_e$ ) estimate (Table 2) at each SNP was calculated as the proportion of the bivariate probability distribution of the full set of simulations smaller than the ( $F_{ST}$ ,  $H_e$ ) estimate obtained for the observed polymorphism.<sup>37</sup>

Intra- and inter-specific tests for detecting natural selection (Tajima's  $D$ ,<sup>38</sup> Fu's  $F_s$ ,<sup>39</sup> Fu and Li's  $D$ ,  $D^*$ ,  $F$  and  $F^*$ <sup>40</sup> and Fay and Wu's  $H^*$ <sup>41</sup>) were performed on each sequence using DNAsp 50.1 software<sup>32</sup> and Arlequin 3.11 software<sup>33</sup> (Supplementary Table S3). For inter-specific tests, re-sequencing with the same primers and PCR conditions was performed on chimpanzee or bonobo samples. These sequences were also used to determine the ancestral state of each polymorphism.

### Phenotype–genotype correlations

To detect associations between stature and alleles at SNPs showing an outlying  $F_{ST}$  value, we performed Spearman's correlation tests using R 2.7.0 software<sup>42</sup> (Table 2). As the numbers of men and women sampled in the various populations differed and sample sizes were small, we chose to perform the correlations with individuals from both sexes grouped together and therefore transformed a woman's height into the height of a man of the same population who has the same  $Z$ -score.<sup>12</sup>  $Z$ -scores were calculated using the mean and variance of each population by sex (Table 1). Spearman's correlations were

**Table 1** Sample sizes and mean stature (cm) by sex in the present sample and in a larger sample<sup>a</sup> and their standard deviations

Population	Men				Women			
	Present sample		Larger sample <sup>a</sup>		Present sample		Larger sample <sup>a</sup>	
	Nb	Sample height; SD	Nb	Pop height; SD	Nb	Sample height; SD	Nb	Pop height; SD
Bongo Gabon <sup>b</sup>	19	159.6; 7.1	25	159.7; 7.6	11	152.1; 4.9	19	151.2; 4.8
Baka Cameroon <sup>c</sup>	16	151.4; 4.9	182	155.6; 5.7	14	145.7; 4.5	184	146.5; 4.9
Nzimé Cameroon <sup>c</sup>	6	169.8; 7.0	42	168.5; 6.5	24	155.3; 6.6	58	155.8; 6.8

<sup>a</sup>Mean stature estimated on larger samples of the same populations by Becker *et al.*<sup>12</sup>

Anthropometric data and DNA samples gathered by:

<sup>b</sup>P Verdu, 2006–2007.

<sup>c</sup>A Froment, 2002–2007.

**Table 2** Heterozygosities and genetic distances between Baka pygmies and Nzimé non-pygmies estimated for two SNPs from *GHR* and *IGF1* genes, associated *P*-values for the test of Beaumont and Nichols with different parameter values<sup>a</sup> and Spearman's estimates for the correlation between genotypes and height

Gene	Sequence	SNP	$H_e$	$F_{ST}$	<i>P</i> -value <sup>b</sup>	<i>P</i> -value <sup>c</sup>	<i>P</i> -value <sup>d</sup>	<i>P</i> -value <sup>e</sup>	$\rho$	<i>P</i> -value	$\rho$ with <i>ad</i> <sup>f</sup>	<i>P</i> -value
<i>GHR</i>	Intron 1b	rs2972419	0.255	0.232	0.011	0.011	0.011	0.018	0.35	<0.001	0.28	0.016
<i>IGF1</i>	Intron 2	rs11831436	0.427	0.238	0.020	0.019	0.018	0.026	-0.34	<0.001	-0.17	0.15

<sup>a</sup> $F_{ST}$  values were compared to the distribution of  $F_{ST}$  values calculated on 500 000 simulations using a modified version of the DFDIST software<sup>36,37</sup> (see Materials and methods for details) with the following parameter values:

<sup>b</sup>Mean  $F_{ST}$  set to 0.033 and  $\theta = 0.01$ .

<sup>c</sup>Mean  $F_{ST}$  set to 0.033 and  $\theta = 0.1$ .

<sup>d</sup>Mean  $F_{ST}$  set to 0.033 and  $\theta = 0.001$ .

<sup>e</sup>Mean  $F_{ST}$  set to 0.038 and  $\theta = 0.01$ .

<sup>f</sup>Partial correlation with admixture as a covariate (see Materials and methods).

calculated, on data from the three populations pooled together, between individual stature and number of copies of the derived allele for each locus.

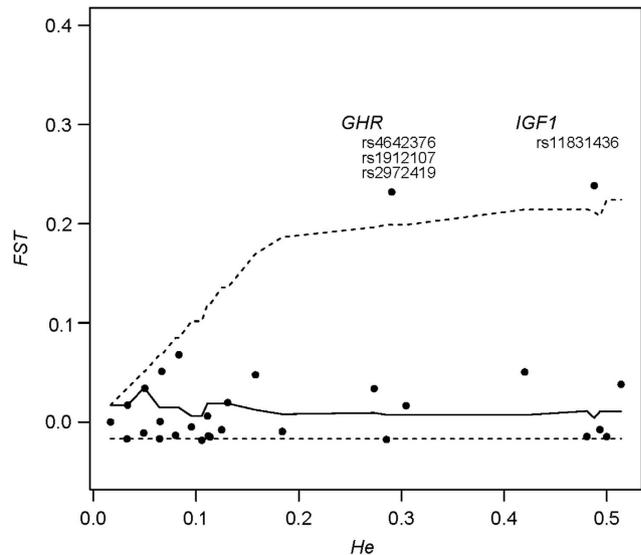
Spurious significant correlations could however be obtained because of population structure.<sup>43</sup> As pygmy populations have been shown to present a substantial level of admixture with the neighboring non-pygmy populations,<sup>2</sup> we also performed partial Spearman's correlations (Table 2) adding the individual level of admixture with the non-pygmy population as a covariate. For individual estimates of the admixture level, we used previously published<sup>3</sup> levels of genetic clustering in the non-pygmy cluster estimated using  $k = 2$  clusters with Structure v. 2.1.<sup>44,45</sup> In this previous publication, the genetic clustering was performed using data on 26 microsatellite polymorphisms from 604 pygmy and non-pygmy individuals as well as from 119 African individuals from the HGDP-CEPH panel.<sup>2,12,46</sup> We determined the individual levels of genotype membership proportions in the non-pygmy (blue) cluster for the 73 individuals of our study overlapping with the sample studied by Becker *et al.*<sup>12</sup> For each pygmy or non-pygmy individual, the admixture estimate was thus the proportion of genotypes assigned to the non-pygmy cluster. This value ranged from 0.23 to 0.73% for Baka pygmies, from 0.50 to 0.94% for Bongo pygmies and from 0.59 to 0.93% for Nzimé non-pygmies.

## RESULTS

In total, 44 SNPs were found in the 10 studied sequences (population frequencies in Supplementary Table S2), in addition to the *GHR* deletion of exon 3, which has previously been shown to be associated with a higher response to GH in GH-deficient children.<sup>47,48</sup> This latter polymorphism was not found to be significantly more frequent in pygmies than in non-pygmies.

For each polymorphism, we estimated the genetic distance between the Baka pygmies and their direct non-pygmy neighbors, the Nzimé ( $F_{ST}$ <sup>35</sup> in Supplementary Table S2). We found significant  $F_{ST}$  values between the Baka and the Nzimé for three SNPs located in the two sequences of the first intron of the *GHR* gene ( $F_{ST} = 0.232$ , FDR-corrected *P*-value < 0.001 for two SNPs of Intron 1a sequence and  $F_{ST} = 0.232$ , FDR-corrected *P*-value = 0.006 for one SNP of Intron 1b sequence) and for one SNP in the *IGF1* intron 2 ( $F_{ST} = 0.238$ , FDR-corrected *P*-value = 0.004). Multilocus  $F_{ST}$ <sup>35</sup> estimated on the *GHR* intron 1 sequences and on the *IGF1* intron 2 sequence were also found to be significant between Baka and Nzimé (results not shown). No significant genetic distances were found on the *STAT5A* and *STAT5B* genes.

$F_{ST}$  values can be influenced by demographic processes such as migrations and genetic drift, as well as natural selection forces.<sup>49,50</sup> To disentangle the respective effects of demography and natural selection on the genetic distances between the Baka pygmies and the Nzimé non-pygmies, we compared the  $F_{ST}$  estimates for each SNP with values obtained on 500 000 simulated data sets (see Materials and methods). These simulations were performed with a mean  $F_{ST}$  value



**Figure 1** Genetic distance ( $F_{ST}$ ) as a function of heterozygosity ( $H_e$ ) for the 42 SNPs present in the Baka and Nzimé samples. The lines represent the median value (solid line) and the 95% intervals (dashed lines) for 500 000 simulations performed under neutrality with a mean  $F_{ST}$  set to 0.033 and  $\theta$  value to 0.01.

equal to that obtained previously using neutral sequences between Baka and non-pygmy populations from Western Central Africa (0.033 with the Gabonese Akele and 0.038 with the Cameroonian Ngumba<sup>3</sup>). Two regions had significantly higher  $F_{ST}$  values as compared to the simulated data sets (Table 2, Figure 1). One of these regions was located on the first intron of the *GHR* gene and contained three SNPs in complete linkage disequilibrium (rs4642376, rs1912107 and rs2972419). For the rest of the analyses only SNP rs2972419 was kept as a representative of the *GHR* region (we chose this SNP because it presented no missing data). The other region was located on the *IGF1* intron 2 and contained one significant SNP: rs11831436. No regions in the two other candidate genes showed a significant difference in  $F_{ST}$  values between pygmies and non-pygmies.

We further searched for evidence of natural selection at the intra-population level by performing seven tests that detect natural selection (see Materials and methods) at the sequence level. No sequence exhibited two significant test results for the same population (Supplementary Table S3).

As two regions, located respectively on *GHR* intron 1 and *IGF1* intron 2, showed a significantly higher genetic distance between Baka

pygmies and Nzimé non-pygmies compared to neutral expectations, we searched for an association with individual stature for these two regions. Both showed a significant Spearman's correlation (Table 2) between the number of copies of the derived allele and individual unisex height equivalent (see Materials and methods). For the *GHR* SNP rs2972419 the ancestral allele was associated with a shorter stature, whereas for the *IGF1* SNP the ancestral allele was associated with a taller stature (Figure 2). Population structure can have a strong confounding effect in association studies<sup>43</sup> and pygmy populations have been shown to present variable levels of admixture with the neighboring non-pygmy populations.<sup>2</sup> Furthermore, these variable levels of individual admixture have been shown to be positively correlated with adult height in pygmy populations.<sup>12</sup> In order to avoid false positives (ie, the Chinese chopstick effect<sup>51,52</sup>), we tested the correlation of the two previously identified regions, including a correction for individual admixture level (see Materials and methods). This partial correlation was significant for the *GHR* region only ( $\rho$  estimate = 0.278,  $P$ -value = 0.016) (Table 2).

The Bongo pygmies are taller and have a higher genetic similarity with the non-pygmy populations than the Baka pygmies.<sup>2,12</sup> Assuming an additive effect of height-associated SNPs, we expect that the Bongo would either have an intermediate frequency between the Baka and the Nzimé at the height-associated SNPs or share alleles with the Baka for some of these polymorphisms and with the Nzimé for others. The present data match this second hypothesis. Indeed, we found high but non-significant  $F_{ST}$  values between the Bongo and the Nzimé at the *GHR* region ( $F_{ST}$  = 0.11, FDR-corrected  $P$ -value = 0.125 for rs2972419) and significant  $F_{ST}$  values with the Baka at the *IGF1* region ( $F_{ST}$  = 0.17, FDR-corrected  $P$ -value = 0.032 for rs11831436). The Bongo were closer to the other pygmy group for the *GHR* gene but closer to the non-pygmy group for the *IGF1* gene, thus consistent with their intermediate height between the Baka and the Nzimé.

## DISCUSSION

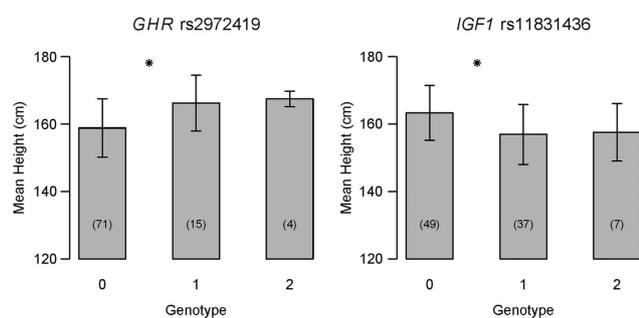
In this study, we have found one region composed by three linked polymorphisms: rs4642376, rs1912107 and rs2972419 in the *GHR* intron 1 showing outlying genetic differentiation between the Baka pygmies and the Nzimé non-pygmies as compared to the neutral expectation. For this region, pygmies have a higher proportion of the ancestral allele compared to non-pygmies, and its derived alleles are associated with a taller stature. The association with stature remained

significant even after correction for admixture levels, hence showing that the effect of these SNPs on stature may not be only due to population structure. An eightfold under-expression of the *GHR* gene was previously found in pygmies, but no variation in the exons of the gene was found in 13 pygmies as compared to eight non-pygmies.<sup>18</sup> The three polymorphisms found in our study are located in the first intron of the gene, which had not been re-sequenced in this previous study. Our three height-associated SNPs may have a role in the expression of the gene or may be linked to an exonic functional polymorphism that was not identified in the smaller sample used in the previous study.<sup>7</sup> Using the CEPH-HGDP genome-wide genotyping data for almost one million SNPs,<sup>53</sup> we calculated the  $r^2$  linkage disequilibrium value<sup>54</sup> of all SNPs within 1 Mb of rs2972419. Among the Biaka pygmies who are the closest to our Baka sample, there are 16 SNPs that showed a  $r^2 > 0.2$ . None of them are located in an exon, whereas two are in a *GHR* intron. However, functional polymorphisms absent in the HGDP-CEPH database could still be linked to the SNPs identified here.

As in previous studies,<sup>20,21,55</sup> we could find no polymorphism in the 5' region of the *IGF1* gene associated with stature among Central African pygmies and non-pygmies. As a recombination hotspot separates most of the *IGF1* gene from its 5' region (see the recombination rate estimated from HapMap Phase 2<sup>56</sup> in Supplementary Figure S1), we re-sequenced other portions of this gene and found that a SNP (rs11831436) in the second intron presented an outlying value of the genetic distance between the Baka pygmies and the Nzimé non-pygmies. This SNP had a higher frequency of the derived allele in the Baka pygmies as compared to non-pygmies and we observed a suggestive association with height, although it was not significant after admixture correction. As this gene is known to be involved in growth, our results suggest that further study should be conducted to investigate the potential linkage between rs11831436 and yet undiscovered functional or regulatory polymorphisms in this gene.

In the two *STAT5* genes, we found common polymorphisms that all showed similar allele frequencies and no significant genetic distances between pygmies and non-pygmies. As these genes exhibit a low recombination rate (Supplementary Figure S1) we think that they can be considered as non-associated with the genetic determination of African pygmies' short stature.

Altogether, these results show that the GH-IGF1 axis is probably involved in the difference in stature observed between African pygmies and their non-pygmy neighbors. This confirms the suggestive associations found by a recent genome-wide study.<sup>22</sup> Interestingly, the SNPs identified here in the *GHR* and *IGF1* genes have not been shown to be associated with height in large-scale genome-wide association studies.<sup>57</sup> Indeed, their frequency is low in the European population (CEU) from HapMap<sup>56</sup> (*GHR* rs4642376: 10.17%, rs1912107: 10.0%, rs2972419: 11.67% and *IGF1* rs11831436: 5.83%). We can thus hypothesize that the genetic determination of the difference in stature between pygmies and non-pygmies is at least partly different from the genetic determination of stature variation in other populations worldwide. Height is known to be a highly multigenic trait;<sup>57</sup> it is therefore plausible that the genetic factors involved in the determination of height may differ across human populations, having experienced different evolutionary histories; variable natural selection pressures affecting human variation in stature could have triggered the evolution of different genes or mutations across populations. Functional assays of the polymorphisms identified in this study would be of considerable interest and could contribute to a better understanding of the genetic variation in human stature.



**Figure 2** Mean height by genotype (number of derived alleles) for the two loci showing an association with stature. Sample sizes for each genotype are given between parentheses. Female height was transformed into male equivalents (see Materials and methods). Genotype: number of copies of the derived allele identified by comparison with the chimpanzee or bonobo sequences. Bars represent SD. \*Tukey's *post hoc* test significant at 95% between these two genotypes.

The polymorphisms reported here in the *GHR* and *IGF1* genes show a higher  $F_{ST}$  level as compared to neutral expectations, suggesting that divergent selection probably occurred during the evolution of pygmies and non-pygmies for these genes. No other test of detection of natural selection was found significant on these sequences. The selection event may have occurred a long time ago and may thus not be detected by such tests. As our study was focused on short sequences, no LD-based tests to detect selection such as iHS values could be performed. The HGDP-CEPH<sup>53</sup> panel iHS values were available for the *GHR* SNP rs2972419 (the two others, *GHR* SNP and the *IGF1* SNP, are not in the HGDP-CEPH SNP database). The iHS value for this SNP was small in the Mbuti pygmies (1.356) and in the South African Bantu population (1.143) and was null for the other sub-Saharan populations (Biaka pygmies, Mandeka and Yoruba). No other CEPH population reached a value higher than 1, thus further supporting that this genetic region has not been strongly influenced by recent positive selection events.

Various evolutionary hypotheses have been proposed to explain African pygmies' short stature (see Perry and Dominy<sup>4</sup> for a review). As we show here, divergent selection may have occurred during the evolution of pygmies and non-pygmies at two genes related to growth. This supports an evolutionary hypothesis of direct selection on stature because of environmental pressures, for example, due to the climate of the equatorial forest (eg, hypotheses based on Bergman's rule that predicts the size of animals in relation to climate<sup>58–60</sup>). However, the SNPs identified here differ with respect to the effect of the ancestral allele, which is associated with shorter stature in *GHR* but with taller stature in *IGF1*. According to population genetic studies,<sup>2,3</sup> the common ancestral population of pygmies and non-pygmies lived about 50–70 000 years ago. Our results concerning the effect of the ancestral alleles do not allow us to make inferences on the stature of this ancestral population, re-emphasizing the fundamental question of whether pygmies evolved to shorter stature or non-pygmies evolved to taller stature. Thus, the divergent selection revealed here may also have resulted in the evolution of non-pygmies towards taller stature. However, little is known about the ancestral environment of either pygmy or non-pygmy Bantu populations that live in forest areas nowadays.<sup>61</sup> Our results open the way to further genetic studies of the *GHR* and *IGF1* genes in other populations of short stature, in other rainforest environments, that may help understand the evolution of stature.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We warmly thank all individuals who participated in the study, as well as Frédéric Austerlitz, Serge Bahuchet, Sylvie Le Bomin and Etienne Patin for useful comments and advice, Renaud Vitalis for help with the modified version of the Dfdist software and Friso Palstra for proofreading. This work was supported by Assistance Publique Hôpitaux de Paris, ATM-MNHN: 'Les relations Sociétés-Natures dans le long terme' 2009–2012, the Fondation pour la Recherche Médicale and the ANR-Blanc program 'GrowingAP'.

- 4 Perry GH, Dominy NJ: Evolution of the human pygmy phenotype. *Trends Ecol Evol* 2009; **24**: 218–225.
- 5 Bailey RC, Head G, Tenike M *et al*: Hunting and gathering in tropical rainforest: is it possible? *Am Anthropol* 1989; **91**: 59–82.
- 6 Cavalli-Sforza LL: *African Pygmies*. London: Academic Press, 1986.
- 7 Diamond JM: Why are pygmies small? *Nature* 1991; **354**: 111–112.
- 8 Walker R, Gurven M, Hill K *et al*: Growth rates and life histories in twenty-two small-scale societies. *Am J Hum Biol* 2006; **18**: 295–311.
- 9 Migliano AB, Vinicius L, Lahr MM: Life history trade-offs explain the evolution of human pygmies. *Proc Natl Acad Sci USA* 2007; **104**: 216–219.
- 10 Becker NSA, Touraille P, Froment A *et al*: Short stature in African pygmies is not explained by sexual selection. *Evol Hum Behav* 2012; advanced publication online.
- 11 Becker NSA, Verdu P, Hewlett B: Can life history trade-offs explain the evolution of short stature in human pygmies? A response to Migliano *et al.* (2007). *Hum Biol* 2010; **82**: 17–27.
- 12 Becker NS, Verdu P, Froment A *et al*: Indirect evidence for the genetic determination of short stature in African pygmies. *Am J Phys Anthropol* 2011; **145**: 390–401.
- 13 Rimoin DL, Merimee TJ, Rabinowitz D *et al*: Peripheral subresponsiveness to human growth hormone in the African pygmies. *N Engl J Med* 1969; **281**: 1383–1388.
- 14 Merimee TJ, Rimoin DL, Cavalli-Sforza LL: Metabolic studies in the African pygmy. *J Clin Invest* 1972; **51**: 395–401.
- 15 Merimee TJ, Zapf J, Froesch ER: Dwarfism in the pygmy. An isolated deficiency of insulin-like growth factor I. *N Engl J Med* 1981; **305**: 965–968.
- 16 Merimee TJ, Zapf J, Hewlett B *et al*: Insulin-like growth factors in pygmies. The role of puberty in determining final stature. *N Engl J Med* 1987; **316**: 906–911.
- 17 Baumann G, Shaw MA, Merimee TJ: Low levels of highaffinity growth hormone-binding protein in African pygmies. *N Engl J Med* 1989; **320**: 1705–1709.
- 18 Bozzola M, Travaglino P, Marziliano N *et al*: The shortness of pygmies is associated with severe under-expression of the growth hormone receptor. *Mol Genet Metab* 2009; **98**: 310–313.
- 19 Hardouin S, Gourmelin M, Noguez P *et al*: Molecular forms of serum insulin-like growth factor (IGF)-binding proteins in man: relationships with growth hormone and IGFs and physiological significance. *J Clin Endocrinol Metab* 1989; **69**: 1291–1301.
- 20 Bowcock A, Sartorelli V: Polymorphism and mapping of the IGF1 gene, and absence of association with stature among African pygmies. *Hum Genet* 1990; **85**: 349–354.
- 21 Schneid H, Le Bouc Y, Seurin D *et al*: Insulin-like growth factor-I gene analysis in subjects with constitutionally variant stature. *Pediatr Res* 1990; **27**: 488–491.
- 22 Jarvis JP, Scheinfeldt LB, Soi S *et al*: Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. *PLoS Genet* 2012; **8**: e1002641.
- 23 Lopez-Bermejo A, Buckway CK, Rosenfeld RG: Genetic defects of the growth hormone-insulin-like growth factor axis. *Trends Endocrinol Metab* 2000; **11**: 39–49.
- 24 Walenkamp MJ, Wit JM: Genetic disorders in the growth hormone – insulin-like growth factor-I axis. *Horm Res* 2006; **66**: 221–230.
- 25 Weiner JS, Lourie JA: *Practical Human Biology*. London; New York: Academic Press, 1981, p 439.
- 26 Rozen S, Skaletsky H: Primer3 on the WWW for general users and for biologist programmers; in Krawetz S, Misener S (eds): *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Totowa: Humana Press, 2000; pp 365–386.
- 27 Amselem S, Duquesnoy P, Attree O *et al*: Laron dwarfism and mutations of the growth hormone-receptor gene. *N Engl J Med* 1989; **321**: 989–995.
- 28 Pantel J, Machinis K, Sobrier ML, Duquesnoy P, Goossens M, Amselem S: Species-specific alternative splice mimicry at the growth hormone receptor locus revealed by the lineage of retroelements during primate evolution. *J Biol Chem* 2000; **275**: 18664–18669.
- 29 Takahashi M, Matsuda F, Margetic N, Lathrop M: Automated identification of single nucleotide polymorphism from sequencing data. *J Bioinform Comput Biol* 2003; **1**: 253–265.
- 30 Stephens M, Scheet P: Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet* 2005; **76**: 449–462.
- 31 Stephens M, Smith NJ, Donnelly P: A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68**: 978–989.
- 32 Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R: DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 2003; **19**: 2496–2497.
- 33 Excoffier L, Laval G, Schneider S: Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* 2005; **1**: 47–50.
- 34 Nei M: *Molecular Evolutionary Genetics*. New York: Columbia University Press, 1987.
- 35 Weir BS, Cockerham CC: Estimating F-statistics for the analysis of population structure. *Evolution* 1984; **38**: 1358–1370.
- 36 Beaumont MA, Nichols RA: Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc Lon* 1996; **263**: 1619–1626.
- 37 Segurel L, Lafosse S, Heyer E, Vitalis R: Frequency of the AGT Pro11Leu polymorphism in humans: does diet matter? *Ann Hum Genet* 2010; **74**: 57–64.
- 38 Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; **123**: 585–595.
- 39 Fu YX: Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 1997; **147**: 915–925.
- 40 Fu YX, Li WH: Statistical tests of neutrality of mutations. *Genetics* 1993; **133**: 693–709.
- 41 Fay J, Wu C: Hitchhiking under positive Darwinian selection. *Genetics* 2000; **155**: 1405–1413.

- 1 Froment A: Biological variation and adaptation in humans: the case of African pygmies. *Bull Mém Soc Anthropol Paris* 1993; **5**: 417–448.
- 2 Verdu P, Austerlitz F, Estoup A *et al*: Origins and genetic diversity of pygmy hunter-gatherers from Western Central Africa. *Curr Biol* 2009; **19**: 312–318.
- 3 Patin E, Laval G, Barreiro LB *et al*: Inferring the demographic history of African farmers and pygmy hunter-gatherers using a multilocus resequencing data set. *PLoS Genet* 2009; **5**: e1000448.

- 42 R Development Core Team: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2008.
- 43 Thomas DC, Witte JS: Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev* 2002; **11**: 505–512.
- 44 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.
- 45 Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000; **155**: 945–959.
- 46 Rosenberg NA, Pritchard JK, Weber JL *et al*: Genetic structure of human populations. *Science* 2002; **298**: 2381–2385.
- 47 Dos Santos C, Essioux L, Teinturier C, Tauber M, Goffin V, Bougneres P: A common polymorphism of the growth hormone receptor is associated with increased responsiveness to growth hormone. *Nat Genet* 2004; **36**: 720–724.
- 48 Wassenaar MJ, Dekkers OM, Pereira AM *et al*: Impact of the exon 3-deleted growth hormone (GH) receptor polymorphism on baseline height and the growth response to recombinant human GH therapy in GH-deficient (GHD) and non-GHD children with short stature: a systematic review and meta-analysis. *J Clin Endocrinol Metab* 2009; **94**: 3721–3730.
- 49 Cavalli-Sforza L: Population structure and human evolution. *Proc R Soc Lond B Biol Sci* 1966; **164**: 362–379.
- 50 Nielsen R: Molecular signatures of natural selection. *Annu Rev Genet* 2005; **39**: 197–218.
- 51 Cardon LR, Palmer LJ: Population stratification and spurious allelic association. *Lancet* 2003; **361**: 598–604.
- 52 Hamer D, Sirota L: Beware the chopsticks gene. *Mol Psychiatry* 2000; **5**: 11–13.
- 53 Cann HM, de Toma C, Cazes L *et al*: A human genome diversity cell line panel. *Science* 2002; **296**: 261–262.
- 54 Hill WG, Robertson A: Linkage disequilibrium in finite populations. *Theor Appl Genet* 1968; **38**: 226–231.
- 55 Schneid H, Seurin D, Le Bouc Y: Pathology of the growth and abnormalities of IGF genes. *Ann Endocrinol-Paris* 1991; **52**: 479–482.
- 56 Frazer K, Ballinger D, Cox D *et al*: A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007; **449**: 851–861.
- 57 Lango Allen H, Estrada K, Lettre G *et al*: Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010; **467**: 832–838.
- 58 Hiernaux J: Long-term biological effects of human migration from the African savanna to the equatorial forest: a case study of human adaptation to a hot and wet climate; in Harrison GA (ed.): *Population Structure and Human Variation*. Cambridge: Cambridge University Press, 1977.
- 59 Ruff CB: Climatic adaptation and hominid evolution: the thermoregulatory imperative. *Evol Anthropol* 1993; **2**: 53–60.
- 60 Schreider E: Morphological variations and climatic differences. *J Hum Evol* 1975; **4**: 529–539.
- 61 Eggert MKH: The Bantu problem and African archaeology; in Stahl AB (ed.): *African Archaeology*. Oxford: Blackwell, 2005; pp 301–326.

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