Extreme Population Differences in the Human Zinc Transporter ZIP4 (SLC39A4) Are Explained by Positive Selection in Sub-Saharan Africa

Johannes Engelken1,2,*, Elena Carnero-Montoro1,*, Marc Pybus1, Glen K. Andrews3, Carles Lalrea-Fox1, David Comas1, Israel Sekler4, Marco de la Rasilla5, Antonio Rosas6, Mark Stoneking2, Miguel A. Valverde7, Rubén Vicente2*, Elena Bosch1*

1 Institute of Evolutionary Biology (CSIC-UPF), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain, 2 Department of Evolutionary Genetics, Max-Planck Institute for Evolutionary Anthropology, Leipzig, Germany, 3 Department of Biochemistry and Molecular Biology, University of Kansas Medical Center, Kansas City, Kansas, United States of America, 4 Department of Physiology, Ben-Gurion University, Beer-Sheva, Israel, 5 Área de Prehistoria, Departamento de Historia, Universidad de Oviedo, Oviedo, Spain, 6 Group of Paleanthropology MNCN-CSIC, Department of Paleobiology, National Museum of Natural Sciences, CSIC, Madrid, Spain, 7 Laboratory of Molecular Physiology and Channelopathies, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain

Abstract

Extreme differences in allele frequency between West Africans and Eurasians were observed for a leucine-to-valine substitution (Leu372Val) in the human intestinal zinc uptake transporter, ZIP4, yet no further evidence was found for a selective sweep around the ZIP4 gene (SLC39A4). By interrogating allele frequencies in more than 100 diverse human populations and resequencing Neanderthal DNA, we confirmed the ancestral state of this locus and found a strong geographical gradient for the derived allele (Val372), with near fixation in West Africa. In extensive coalescent simulations, we show that the extreme differences in allele frequency, yet absence of a classical sweep signature, can be explained by the effect of a local recombination hotspot, together with directional selection favoring the Val372 allele in Sub-Saharan Africans. The possible functional effect of the Leu372Val substitution, together with two pathological mutations at the same codon (Leu372Pro and Leu372Arg) that cause acrodermatitis enteropathica (a disease phenotype characterized by extreme zinc deficiency), was investigated by transient overexpression of human ZIP4 protein in HeLa cells. Both acrodermatitis mutations cause absence of the ZIP4 transporter cell surface expression and nearly absent zinc uptake, while the Val372 variant displayed significantly reduced surface protein expression, reduced basal levels of intracellular zinc, and reduced zinc uptake in comparison with the Leu372 variant. We speculate that reduced zinc uptake by the ZIP4-derived Val372 isoform may act by starving certain pathogens of zinc, and hence may have been advantageous in Sub-Saharan Africa. Moreover, these functional results may indicate differences in zinc homeostasis among modern human populations with possible relevance for disease risk.


Editor: Joshua M. Akey, University of Washington, United States of America

Received July 24, 2013; Accepted December 5, 2013; Published February 20, 2014

Copyright: © 2014 Engelken et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: JE was supported through a Postdoc scholarship from the Volkswagenstiftung (Az: I/85 198). This work was partially funded by grants BFU2008-01046, SAF2011-29239, SAF2012-38140, CDI2012-36682 and SAF2010-16725 from the Spanish Ministry of Science and Competitiveness, Fondo de Investigación Sanitaria (Red HERACLES RD12/0042/0014), and FEDER Funds as well as by grants SAF2012-38140, SAF2011-29239, SAF2012-38140, CGL2012-36682 and SAF2010-16725 from the Spanish Ministry of Economy and Competitiveness, Fondo de Investigación Sanitaria (Red HERACLES RD12/0042/0014), and FEDER Funds as well as by grants 2009SGR-1101 and 2009SGR-1369 from Direcció General de Recerca, Generalitat de Catalunya. El Sídron research was also supported by Consejería de Cultura del Principado de Asturias. MAV is the recipient of an ICREA Academia Award. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ruben.vicente@upf.edu (RV); elena.bosch@upf.edu (EB)

† These authors contributed equally to this work.

‡ RV and EB are joint senior authors on this work.

Introduction

Zinc homeostasis is critically important for human health. Similarly to iron, zinc has manifold functions in the body, such as in the immune system [1], aging [2], DNA repair [3], signaling [4] and in diseases such as diabetes [5] and cancer [6]. On the molecular level, zinc acts as a co-factor in hundreds of metalloenzymes as well as in hundreds of DNA-binding proteins (e.g. zinc finger proteins). Zinc homeostasis is tightly regulated by 10 zinc efflux transporters and 14 zinc influx transporters (encoded by the SLC30A family genes, respectively). ZIP4 (SLC39A4) is the most important intestinal zinc uptake transporter and is expressed at the apical membrane of enterocytes [7,8]. Loss-of-function mutations in ZIP4 cause acrodermatitis enteropathica [9,10] (MIM 201100), a congenital disease characterized by extreme zinc deficiency if left untreated without supplemental zinc [11,12]. Fittingly, it was recently reported that the loss of expression of this gene in a ZIP4 intestine-specific knockout mouse caused systematic zinc deficiency, leading to disruption of the intestine stem cell niche and loss of intestine integrity [13]. The single nucleotide polymorphism (SNP) c.1114C>G (rs1871534) in the ZIP4 gene (SLC39A4; NM_130649.2) results
in the substitution of leucine for valine at amino acid 372 (Leu372Val) in the human ZIP4 transporter. This non-synonymous SNP is one of the most markedly differentiated genetic variants in the genome in terms of allele frequency differences between populations [14–16], according to data from HapMap [17], the Human Genome Diversity Panel (HGDP) [18] and the 1000 Genomes Project [16]. Extreme population differentiation is a signature of local positive selection [15,19–21], but genomic scans for targets of natural selection based on other criteria, such as extended long haplotypes [22–24] or selective signatures in the allele frequency spectrum [25], have failed to identify ZIP4 as a candidate gene for positive selection. To date, whether this variant has evolved under positive selection or neutrality, and its potential functional significance, has not been examined.

In the work reported here, we had three main objectives: (i) to investigate evolutionary explanations for the extreme population differentiation of the ZIP4 Leu372Val polymorphism by use of coalescent simulations; (ii) to test for functional differences in cellular zinc transport between the alleles of the Leu372Val polymorphism using a heterologous expression system; and (iii) to discuss potential selective forces behind this possibly adaptive event and their implications for zinc homeostasis in modern humans. We have extensively characterized the extreme geographical differentiation of the Leu372Val substitution and provide evidence that it has been subject to a nearly complete selective sweep in Sub-Saharan Africa. Our simulations show how the extreme pattern of population differentiation, yet absence of other classical signatures of positive selection, can be explained by directional selection accompanied by the effects of a recombination hotspot near the polymorphic adaptive site. Additionally, our data demonstrate in vitro functional differences between the two human polymorphic alleles at codon 372 of the human ZIP4 transporter in surface protein expression, basal intracellular levels of zinc and zinc uptake. We hypothesize that the reduction in intracellular zinc levels mediated by the Val372 allele may have been advantageous in Sub-Saharan Africa, possibly by reducing access of a geographically restricted pathogen to this micronutrient.

In vitro functional differences between the two human polymorphic alleles at codon 372 of the human ZIP4 transporter in surface protein expression, basal intracellular levels of zinc and zinc uptake. We hypothesize that the reduction in intracellular zinc levels mediated by the Val372 allele may have been advantageous in Sub-Saharan Africa, possibly by reducing access of a geographically restricted pathogen to this micronutrient, and that other possible secondary consequences for disease risk and health may result from the differential activity of the ZIP4 alleles.

Results

Worldwide allele frequencies

Five common non-synonymous SNPs are known in the human ZIP4 gene (Table 1): Glu10Ala (rs2280839), Ala56Thr (rs2280838), Ala114Thr (rs12855765), Thr357Ala (rs2272662) and Leu372Val (rs1871534). However, only the latter two SNPs show elevated levels of population differentiation in the 1000 Genomes Phase1 sequencing data when comparing the Yoruba from Ibadan, Nigeria (YRI) with either the Han Chinese from Beijing, China (CHB) or Utah residents of Northern and Western European origin (CEU). As shown in Figure 1A and 1B, their FST values fall above the 99.999th percentile of the genome-wide FST distributions between CEU-YRI (with FST values for rs2272662 and rs1871534 of 0.48 and 0.98, respectively) and between CHB-YRI (with FST values of 0.51 and 0.98, respectively). We therefore verify that the Leu372Val substitution encoded by SNP rs1871534 is the non-synonymous polymorphism exhibiting the most extreme allele frequency differences in the human ZIP4 gene. Next, we genotyped the 51 populations from the Human Genome Diversity Panel (HGDP) and compiled additional allele frequencies for this position in worldwide populations from the Allred database [26,27]. Additionally, we obtained new data from a Pygmy population from Gabon and North African populations of Western Sahara, Morocco, and Libya. These new data confirm that the Leu372 variant is the most common allele outside of Africa, and provide a more detailed picture of the geographical allele frequency distributions of this non-synonymous polymorphism (Figure 1C and Table S1). Overall, the Val372 variant showed the highest frequencies in Sub-Saharan Africa, with populations such as the Ibo or the Yoruban people exhibiting the extreme derived allele frequencies worldwide (0.99 and 0.96, respectively). Interestingly, two presumably early-branching groups in Sub-Saharan Africa, the Pygmy and the San people, showed opposing trends in the derived allele frequency (0.94 and 0.0, respectively). Even though the small sample size from the San (only six individuals) means that a population frequency of up to 0.221 cannot be excluded (with p = 0.05 based on assuming Hardy-Weinberg equilibrium and a binomial approach), such divergent tendencies in these two Sub-Saharan populations are maintained. Given the elevated levels of population differentiation of the SNP rs2272662, we also genotyped the HGDP panel for the Thr357Ala polymorphism. However, compared with the Leu372-Val substitution, the derived allele at this non-synonymous SNP displayed intermediate frequencies worldwide (Figure S1 and Table S1) and less extreme allele frequency differences between populations.

Identification of Leu372 as the ancestral variant by resequencing in a Neanderthal

Given the allele frequency differences observed in the Leu372Val polymorphism between the two early human branches in Africa and the uncertainty associated with the low coverage of the Neanderthal genome draft sequence [28], we resequenced the corresponding orthologous positions for rs1871534 and rs2272662 in an additional Neanderthal sample, labeled SD1253 and excavated at El Sidrón site in Spain [29]. The two positions were amplified in a multiplexed reaction, along with a diagnostic Neanderthal mitochondrial DNA (mtDNA) fragment, to monitor contamination in the PCR reaction. For the L16250-H16262
diagnostic mtDNA fragment, 64 clones were generated (Figure S2), all of which show the Neanderthal-specific 16234T-16244A-16256A-16258G haplotype [28]. This again supports the very low level of contamination in this particular sample. For the orthologous positions of the human rs1871534 and rs2272662 SNPs, 19 and 14 sequences were successfully obtained, respectively. With the exception of one clone in the second position, all sequences showed the previously inferred ancestral alleles, in agreement with the reads present for the Vindija individuals 33.16 (one read for each position), 33.25 (two for rs1871534 and none for rs2272662) and 33.26 (two and one, respectively) (Figure 2). The successful resequencing of this Neanderthal individual, together with published reads from additional Neanderthals [28] and from the Denisovan individual [30], strongly suggests that the Leu372 variant (encoded by the C allele in rs1871534) is the ancestral human form, which is also in agreement with the chimpanzee state (Figure 2). Together with the extreme population differentiation pattern, these results suggest that a selective sweep may have taken place in Sub-Saharan Africa, where the derived variant is nearly fixed.

**Extreme population differentiation explained by selection and a recombination hotspot**

Next we examined the complete genomic region around ZIP4 (Figure 3) in the 1000 Genomes sequencing data. Whereas we found a cluster of three strongly elevated $F_{ST}$ scores between CEU and YRI in the neighboring SNPs rs1871535 (intronic), rs1871534 and rs2272662 (further suggesting directional selection in a specific geographical region), in both populations there was a clear absence of extreme values in neutrality statistics such as Tajima’s D or Fay and Wu’s H (Figure S3). Notably, no other polymorphism in the flanking region of the human ZIP4 displays the high levels of population differentiation of the Leu372Val substitution. Interestingly, in both African and non-African populations there is a recombination hotspot in the ZIP4 gene, which could have reduced any signature of selection on the surrounding linked variation, thereby explaining the apparent lack of significant departures from neutrality. To further investigate this possibility, we carried out coalescence simulations under a variety of recombination and selection scenarios using a well-established demography [31]. As shown in Figure 3D, the observed values for $F_{ST}$ and most of the different neutrality statistics cannot be explained by neutral evolution or positive selection with a constant recombination rate. Instead, this atypical pattern of extreme population differentiation, yet seemingly neutral Tajima’s D and other neutrality statistics, showed a higher recovery in simulations with directional selection on the derived allele in Sub-Saharan African populations in the context of the observed recombination landscape, including the hotspot (Figure 3D and 3E, Figure S4). In a more formal evaluation of the results, we quantified the empirical probability for each scenario and neutrality test as well as for different combinations of tests by using composite scores encompassing at least three complementary signatures of positive selection: (i) site frequency spectrum, (ii) population differentiation, and (iii) haplotype structure. The scenario of “weak selection ($s = 0.005$) + hotspot” is the most likely among the different ones tested (Table S2). Moreover, all the empirical likelihoods calculated for the different composite scores indicate that the proposed scenario of “weak selection ($s = 0.005$) + hotspot” is more likely than the neutral scenario (Table 2). Therefore, our simulation results indicate that the atypical patterns of selection in the ZIP4 gene can indeed be explained by positive selection occurring upon the Val372 allele in Sub-Saharan African populations and that recombination has erased further accompanying signatures of the selective sweep. Selection coefficients lower than the ones tested (3.0%, 1.0%, 0.5%) further dilute the signal of selection in the site frequency spectrum based neutrality tests (results not shown), but require such long duration times of the sweep that would substantially predate the population split between African and Eurasian populations.

**Functional effect of Leu372Val**

We observed that the Leu372Val polymorphism affects a highly conserved amino acid (Figure 4) and that the same codon position has been altered in acrodermatitis patients carrying missense mutations Leu372Arg [32] and Leu372Pro [8]. Moreover, both PolyPhen [33] and SIFT [34] algorithms predict functional effects for the Leu372Val substitution (see Table 1). These observations led us to test the Leu372Val polymorphism for a possible functional change in the ZIP4 transporter, using transiently transfected HeLa cells. To be able to control for possible haplotypic effects between the two most highly differentiated non-synonymous SNPs in the ZIP4 transporter, we also considered variation at the Thr357A1a polymorphism in the functional analyses. Furthermore, we introduced the pathological mutations Leu372Arg and Leu372Pro in the Ala357 background of the human ZIP4 gene and analyzed them as well. The pathological impact of the Leu372Pro mutation on ZIP4 protein biology and function has already been evaluated in the mouse ZIP4 protein [10], but not the Leu372Arg mutation. Besides providing confirmation of their impact in the context of the human gene, the use of these pathological mutations provided us with an

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>DAF CEU</th>
<th>DAF CHB</th>
<th>DAF YRI</th>
<th>F_{ST} CEU-YRI</th>
<th>F_{ST} CHB-CEU</th>
<th>F_{ST} YRI-CHB</th>
<th>Amino Acid Replacement</th>
<th>Functional Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2280839</td>
<td>0.4812</td>
<td>0.4485</td>
<td>0.4545</td>
<td>-0.0042</td>
<td>-0.0030</td>
<td>-0.005</td>
<td>Glu10Ala</td>
<td>Benign</td>
</tr>
<tr>
<td>rs2280838</td>
<td>0.5235</td>
<td>0.4347</td>
<td>0.3803</td>
<td>0.0335</td>
<td>0.0080</td>
<td>0.0006</td>
<td>Ala58Thr</td>
<td>Benign</td>
</tr>
<tr>
<td>rs17855765</td>
<td>0.4812</td>
<td>0.4278</td>
<td>0.2601</td>
<td>0.0942</td>
<td>-0.0002</td>
<td>0.0535</td>
<td>Ala114Thr</td>
<td>Benign</td>
</tr>
<tr>
<td>rs2272662</td>
<td>0.4941</td>
<td>0.5401</td>
<td>0.0057</td>
<td>0.4823</td>
<td>-0.0005</td>
<td>0.5110</td>
<td>Thr357A1a</td>
<td>Benign</td>
</tr>
<tr>
<td>rs1871534</td>
<td>0.0059</td>
<td>0.0000</td>
<td>0.9787</td>
<td>0.9765</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

*Reported non-synonymous SNPs with minor allele frequencies (MAF) greater than 0.10 in any of the three 1000 Genomes Project populations. Values in bold are above the 99th percentile of the corresponding $F_{ST}$ genome-wide distribution among the two compared populations. Abbreviations: DAF, Derived Allele Frequency; PD, Probably Damaging.

doi:10.1371/journal.pgen.1004128.t001
Figure 1. Extreme population differentiation of the Leu372Val polymorphism in ZIP4 (SLC39A4). (A) Distribution of $F_{ST}$ pairwise scores between CEU and YRI in SNPs from the 1000 Genomes data, plotted on a linear and logarithmic scale. The $F_{ST}$ values and corresponding quantiles of the five common SNPs in ZIP4 are indicated. *The $F_{ST}$ quantile of rs1871534 (Leu372Val) was 0.99999977. (B) Distribution of $F_{ST}$ pairwise scores between CHB and YRI in SNPs from the 1000 Genomes data. The $F_{ST}$ values and corresponding quantiles of the five common SNPs in ZIP4 are indicated. **The $F_{ST}$ quantile of rs1871534 (Leu372Val) was 0.99999817. (C) Contour map of worldwide frequencies of the Val372 variant at the rs1871534 SNP. A complete list of populations and allele frequencies is available in Table S1.

doi:10.1371/journal.pgen.1004128.g001
extreme phenotype to which to compare the phenotype associated with the ZIP4 non-synonymous polymorphisms. In all cases, functional analyses were carried out to determine effects on expression, subcellular localization, and zinc transport.

As shown in Figure 5, human ZIP4 proteins carrying the Leu372Pro and Leu372Arg mutations showed an absence of surface protein expression ($P < 0.001$, one way ANOVA versus the Ala357-Leu372 isoform), consistent with the known causal role of these variants in the zinc deficiency disorder, acrodermatitis enteropathica. Interestingly, the derived Val372 variant also showed significantly decreased surface expression, but to a much lesser extent, and independently of the Thr357Ala substitution ($P < 0.05$ in both Ala357 and Thr357 backgrounds; one way ANOVA versus the Ala357-Leu372 isoform). Overall, the Leu372Val substitution had a highly significant effect on surface expression (ANOVA, $p = 0.00021$), while there was no effect ascribable to the Thr357Ala replacement ($p = 0.579$). Western blot analysis of all isoforms revealed a remarkable decrease in detection of the Ala357-Pro372 isoform (Figure S5A). However, the reduced expression of this isoform was not due to a defect in the construct sequence but to a higher protein degradation rate, as shown in Figure S5B. Further analysis showed that the Ala357-Leu372 and Ala357-Val372 isoforms do not differ in protein degradation rate. Therefore, the differences in the surface expression experiment must be due to a different trafficking pattern of these variants. In this sense, co-localization of ZIP4 with calnexin (a protein present in the lumen of the endoplasmic reticulum) indeed showed that those proteins presenting lower surface expression were partially retained in the endoplasmic reticulum (Figure S6).

Zinc transport analysis of the different ZIP4 isoforms was performed in two ways. First, we quantified basal zinc content with FluoZin-3 in HeLa cells overexpressing the various ZIP4 variants during a 24-hour period (Figure 6A), and second, we recorded intracellular zinc uptake upon perfusion with an external solution containing 200 $\mu$M $\text{Zn}^{2+}$ (Figure 6B). Our results show that basal zinc content in cells overexpressing pathological variants Pro372 and Arg372 did not differ from surrounding non-transfected HeLa cells. On the contrary, all common ZIP4 variants (Ala357, Thr357, Leu372 and Val372) promoted increased intracellular zinc levels. However, and in agreement with their reduced surface expression, Val372 variants (in both Ala357 and Thr357 backgrounds) presented lower basal zinc content compared to Leu372 ($P < 0.01$ and $P < 0.05$, respectively; one way ANOVA versus the Ala357-Leu372 isoform; Figure 6A). As shown in Figure 6B, cells overexpressing the pathological Leu372Arg and Leu372Pro mutations did not uptake zinc, consistent with their inability to traffic to the plasma membrane. Zinc uptake mediated by the Val372 variants was also consistent with their reduced membrane expression; i.e. the Val372 variants in both Ala357 and Thr357 backgrounds presented significantly lower maximum uptake.

Figure 2. Human ZIP4 sequence alignment with chimpanzee and archaic hominids. Archaic hominid sequences are shown from one Denisovan individual and four Neanderthal individuals (three from Vindija and one from El Sidrón, the latter resequenced for the present study). The orthologous positions of the human rs1871534 and rs227262 SNPs are shown in orange and black, respectively. Note that the reference sequence displayed at the bottom of the figure spans from position 145,639,648 to 145,639,756 in chr8 (hg19) and is the reverse complement of the SLC39A4 coding sequence. This analysis shows that the ZIP4 reference sequence carries the ancestral allele shared with archaic hominids and chimpanzee at these two locations. doi:10.1371/journal.pgen.1004128.g002
Figure 3. Genomic context and patterns of selection in a 10 kb region around ZIP4 (SLC39A4). (A) Structure of the human ZIP4 (SLC39A4) gene. (B) FST scores between YRI and CEU in a 10 kb window centered in the Leu372Val polymorphism (rs1871534). The three indicated SNPs show FST scores above the 99th percentile (indicated with a black line) of the corresponding genome-wide FST distribution. (C) Recombination landscape in YRI. (D) Distributions of diverse neutrality and population differentiation statistics based on coalescent simulations [31] carried out under neutrality and different selection scenarios. The presence of a recombination hotspot of moderate strength reduced all signals of positive selection in the neutrality tests except for population differentiation. Overall, the observed values (averaged in a 10 kb window around ZIP4; green line) are compatible with a moderate selection coefficient (0.5%) simulated under the observed recombination landscape. (E) Coalescent simulations
assuming different selection coefficients. Here, the recombination landscape was fixed to the observed landscape in YRI and we tested different selection coefficients (3%, 1%, 0.5%, and 0%, the latter corresponding to neutrality). As expected, stronger selection coefficients yielded increasingly stronger deviations from neutrality.

doi:10.1371/journal.pgen.1004128.g003

transport (T_{1/2}) compared to the Leu372 variant (P<0.01 in each case; Figure 6B). However, the time to reach half-maximal transport (t1/2) showed no significant difference, indicating that transport kinetics were not markedly different among the four common variants (Figure 6). Overall, these results suggest that the Val372 variant does not disturb the kinetics of the ZIP4 transporter but leads to lower zinc uptake transport due to reduced surface expression.

Discussion

Leu372Val as the target of an atypical selective sweep in Africa

Our study was triggered by the observation of extreme population differentiation between Sub-Saharan African and non-African populations involving the Leu372Val polymorphism in the \( \text{ZIP4} \) gene, unaccompanied by any other signals of a classic hard sweep, such as long extended haplotype homozygosity, in either population (Figures S3, S7 and S8). By interrogating and compiling allele frequencies in more than 100 worldwide human populations, we further characterized the extreme population differentiation of the Leu372Val polymorphism and confirmed that this result was not an artifact of allele switching [15]. Given the worldwide distribution of the human derived and ancestral alleles (confirmed by sequencing a Neanderthal and phylogenetic conservation), we conclude that this sweep must have taken place within Africa, probably in Sub-Saharan Africa, and not outside the African continent. Notably, the extreme population differentiation of the Leu372Val polymorphism represents the top fourth region of the worldwide distribution of the human derived and ancestral alleles (confirmed by sequencing a Neanderthal and phylogenetic conservation), we conclude that this sweep must have taken place within Africa, probably in Sub-Saharan Africa, and not outside the African continent. Notably, the extreme population differentiation of the Leu372Val polymorphism represents the top fourth region within the global genome-wide \( F_{ST} \) distribution between CEU-YRI obtained from the 1000 Genomes Project data. The only CEU-YRI \( F_{ST} \) values that are more extreme all involve well-known examples of local geographical adaptation in humans: the \( \text{SLC24A5} \) and \( \text{SLC45A2} \) genes (with an \( F_{ST} \) of 0.9826 and 0.9765, respectively), which have been associated with light skin pigmentation in Europeans; and the \( \text{Duffy} \) gene (with an \( F_{ST} \) of 0.9765), which provides resistance to the malaria pathogen \( \text{Plasmodium vivax} \). Moreover, with the notable exception of \( \text{Duffy} \) FY*O allele [35,36], most of the extreme \( F_{ST} \) values obtained when comparing Africans with non-Africans are usually attributed to local adaptation outside of Africa. Our detection of such a rare signature of natural selection in the African continent is therefore quite remarkable. Interestingly, it seems congruent with a recent study that has found only limited evidence for classical sweeps in African populations, which is likely due to a combination of limitations of the currently used methodology and specific characteristics of African population history [37].

Notably, we observed a nearly complete but mild selective sweep for the Val372 variant in Africa, which involves three SNPs with extremely high population differentiation, whereas most other commonly used tests for selection show values not even close to genome-wide significance. Our coalescent simulations indicate that this unusual pattern might be explained by local positive selection in combination with an observed recombination hotspot of moderate strength. At approximately 7 cM/Mb, the recombination rate is only around 7-fold higher than the genomic background, but the hotspot is extended over 3–4 kb. Therefore, a similar number of recombination events may accumulate over time corresponding to a more typically sized hotspot of 1 kb and a recombination rate of around 25 cM/Mb. To our knowledge, this is the first example of a nearly complete selective sweep that is obscured by the effect of a recombination hotspot. It is compatible with earlier theoretical observations that instances of weaker selection in the presence of recombination may not always have an influence on polymorphism statistics [38] and with the observed effect of recombination on the partial sweep around the malaria-related \( \beta \)-globin gene [39]. Because of the unclear effects of the recombination hotspot, it was not possible to estimate the age of the sweep using linkage disequilibrium decay related methods (e.g. [40]). It is likely that a mild selection pressure would have needed a long time to reach the extreme population differentiation values observed, indicating this may be an ancient event. The fact that the high frequency of the Val372 allele is restricted to Sub-Saharan African populations suggests that the selection process started after the Out of Africa expansion of modern humans (i.e. sixty thousand years ago). Alternatively, it is also possible that the bottleneck in the Out of Africa expansion did not sample the Val372 allele, which in turn could explain its absence in most non-African populations. This implies that the Out of Africa event is not a hard upper limit for the age of the selection process.

Other more complex evolutionary scenarios cannot be entirely ruled out, and could warrant a more detailed investigation. For example: (i) selection acting on standing genetic variation, in the sense that the Val372 variant was already segregating when it came under the influence of local selection; (ii) additional directional selection against the Val372 allele in non-African populations; (iii) selection favoring the Leu372 variant on multiple, geographically independent origins mostly in non-African populations, in addition to positive selection on the Val372 variant in Africa; and (iv) ‘gene surfing’ of any of the two variants on the wave of a population range expansion [41]. However, we consider it is unnecessary to invoke such complex scenarios in preference to the simpler one we propose based on coalescent simulations. Moreover, back-and-forth migrations between Sub-Saharan African, Northern African and Middle Eastern populations after the first Out-of-Africa wave of migration [42] could easily explain the observed low-intermediate allele frequencies in Middle Eastern populations without invoking additional selection events.

Table 2. Likelihood of several combinations of empirical neutrality test values around \( \text{ZIP4} \) in different selection scenarios versus neutrality.

<table>
<thead>
<tr>
<th>Likelihood = P(Sweep)/P(Neutral)*</th>
<th>s = 0.005</th>
<th>s = 0.01</th>
<th>s = 0.03</th>
</tr>
</thead>
<tbody>
<tr>
<td>All statistics combined</td>
<td>20.105</td>
<td>0.524</td>
<td>0.003</td>
</tr>
<tr>
<td>( F_{ST} ) - XPEHH - Tajima’s D</td>
<td>18.518</td>
<td>3.321</td>
<td>0.271</td>
</tr>
<tr>
<td>( F_{ST} ) - XPEHH - Fu Li’s D</td>
<td>5.925</td>
<td>0.498</td>
<td>0.025</td>
</tr>
<tr>
<td>( F_{ST} ) - XPEHH - Fay Wu’s H</td>
<td>326.386</td>
<td>41.330</td>
<td>1.393</td>
</tr>
</tbody>
</table>

*Likelihoods were computed from the combined empirical probabilities obtained when considering the observed recombination landscape in \( \text{ZIP4} \) (see Table S3).

*Neutrality statistics include \( F_{ST} \) for population differentiation; XPEHH for extended linkage disequilibrium decay; and Tajima’s D, Fu Li’s D and Fay Wu’s H for site frequency spectrum.

doi:10.1371/journal.pgen.1004128.t002
In the absence of additional linked functional variants in the region, we infer that directional selection has acted on the ZIP4 gene. This conclusion is supported by: (i) the disease phenotype of acrodermatitis enteropathica, which involves extreme and potentially lethal zinc deficiency and is caused by, among others, diverse mutations at amino acid position 372 in ZIP4 [43]; (ii) the absence of cellular zinc transport in Leu372Arg and Leu372Pro acrodermatitis mutants; (iii) the finding that the Val372 variant leads to reduced zinc transport at the cellular level; and finally (iv) the conservation of this amino acid position across diverse species (Figure 4). Furthermore, we infer that the Leu372Val substitution was the functional site targeted by selection due to its location in the predicted center of selection (highest $F_{ST}$), and since it is the only putative functional polymorphism in the ZIP4 gene. Of the other two polymorphic variants with somewhat high allele frequency differences between populations, the Thr357Ala substitution (rs2272662) does not show any functional effect and the intronic rs1871535 cannot be associated with any known regulatory function (according to information on DNAse I hypersensitivity clusters, CpG Islands and transcription factor binding sites available from the ENCODE data [44]). Therefore, both rs1871355 and rs2272662 are likely to be neutral. Other non-synonymous polymorphisms with intermediate allele frequencies in the ZIP4 gene (Glu10Ala, Ala58Thr, and Ala114Thr) have very low $F_{ST}$ scores and are therefore not considered candidate variants for selection.

Possible consequences at the cellular and organ level

Our functional results in transfected HeLa cells indicate that the Val372 form of the ZIP4 receptor has lower relative cell surface expression, despite no expected differences in mRNA expression and protein synthesis. Interestingly, we found that this decreased expression translated into reduced zinc transport of the derived Val372 variant at the cellular level. That is, we observed differences in the maximal transport ($T_{max}$) with no significant differences in the transport kinetics ($T_{1/2}$) between Leu372 and Val372. The functional results observed in transfected HeLa cells are likely to be transferable to other epithelial cells, in accordance with independent experiments showing an effect of acrodermatitis variants at position 372 on surface expression (in CHO cells) and on zinc transport (in HEK293 cells) when using mouse cDNA [10]. However, the critical function of ZIP4 in knockout studies has been shown to primarily affect intestinal zinc uptake [13].

In contrast to the Leu372Pro and Leu372Arg acrodermatitis mutations, which served as controls and showed an almost complete absence of zinc transport, both the Leu372 and Val372 variants must be capable of carrying out zinc transport in the normal range of concentrations, given their high frequency in the healthy population. The consequences of this difference in zinc transport at the organ and organismal level are currently unclear, although there is a strong indication that this variant may indeed be phenotypically relevant. For example, a similar non-synonymous mutation in the porcine homologue of ZIP4 leads to non-pathogenic reduced tissue concentrations of zinc in piglets [45].
Nutritional immunity as a putative selective force

Could the concept of "nutritional immunity" [46,47] involving zinc explain a putative selective force in Sub-Saharan Africa? According to this hypothesis, the human host restricts access to certain micronutrients, so that pathogens become less virulent. This is a well-known mechanism of immune defense mediated by iron metabolism [48], and there are indications that zinc metabolism could have a similar function [47,49]. For example, hypoferrernia and hypozincemia are both part of the acute phase response to infection and both seem to be influenced by a different zinc transporter from the same family, ZIP14 [50]. We speculate that the selective force behind the extreme FST pattern of the Leu372Val substitution may be related to pathogens or infectious diseases. It is known that decreased zinc uptake mediated by ZIP4 leads to decreased zinc concentrations in the major organs, as shown in a mouse knockout model [13]. While the phenotypic effect of the Val372 allele in humans is currently unknown, we conjecture that the in vitro difference may indeed translate into physiological differences, possibly leading to a slightly decreased uptake of dietary zinc. Fittingly, there is suggestive evidence that...
African genetic ancestry may involve lower serum levels of zinc [51], as African-American children have a fourfold risk of zinc deficiency compared to Hispanic children. This result would suggest that African ancestry may be associated with lower serum zinc levels, although these results may be biased due to differences in lifestyle, socio-economic status etc., and this observation would need to be confirmed by controlled studies. Alternatively, lower zinc concentrations mediated by the Leu372Val substitution in the enterocyte cells could facilitate early diarrhoeal episodes during a digestive infection in order to reduce the pathogen load on the luminal surface [52,53]. Similarly, the lower level of expression of the ZIP4 isoform carrying the Val372 variant could also be advantageous if any parasite uses the ZIP4 receptor to enter enterocytes. Furthermore, the selective force may be related to pre-historic differences in dietary zinc due to lifestyle or to local levels of zinc concentrations in soil and the food chain.

Potential implications – towards a phenotype

No large-scale ethnic comparisons related to serum or tissue zinc concentrations are available. To our knowledge, rs1871534 has not been tested in case-control studies in African populations related to one of the numerous existing infectious diseases like malaria, trypanosomias or Lhassa fever. It is therefore possible that important evidence for a possible selective force has been missed. In future research, the inclusion of additional cell lines, and genotype-phenotype association studies in diverse ethnic populations may help to clarify further phenotypic consequences of this non-synonymous polymorphism. Genotype-phenotype association studies should involve African-American or East African populations in which the Val372 allele is segregating at intermediate frequencies. Candidate phenotypes and traits to interrogate could be serum zinc concentrations, zinc content in hair and nails, serum zinc concentrations after controlled zinc supplementation, and a range of disease traits, especially diseases with an elevated risk in different populations, for example, diverse types of cancer in African Americans. As this SNP was not included in the commonly used Affymetrix and Illumina SNP arrays with up to one million variants (although it is included in several of the latest arrays), potential clinically relevant associations may have been missed. Interestingly, common polymorphisms in other zinc transporters show genome-wide associations with disease traits, such as a non-synonymous variant in the zinc efflux transporter ZnT8 (SLC39A8) and diabetes incidence [54], as well as a regulatory variant in the zinc influx transporter ZIP6 (SLC39A6) and survival in esophageal cancer [55].

Conclusions

The identification of a high-frequency derived allele polymorphism in the ZIP4 zinc transporter gene (SLC39A4), combined with a more complete picture of worldwide allele frequencies and in-depth coalescent simulations, is consistent with a long lasting selective event in Sub-Saharan Africa driven by a moderate selection coefficient. This event did not leave the typical footprint of a selective sweep with long haplotypes or detectable neutral deviations in the allele frequency spectrum of the surrounding region, most likely because of the presence of a moderate recombination hotspot. Through functional experiments we have verified the Leu372Val substitution as the likely causal site. Given that two functionally different alleles of this key component of cellular zinc uptake are distributed so divergently across worldwide populations, our results may point to functional differences in zinc homeostasis among modern human populations with possible broader relevance for health and disease.

Materials and Methods

Samples and genotyping

The G and C alleles at rs1871534 (Leu372Val) have been swapped in various public sources such as HapMap (http://www.hapmap.org) or dbSNP (http://www.ncbi.nlm.nih.gov/SNP) that report conflicting allele frequencies in populations with a similar geographical origin. This situation led us to repeat the genotyping of this SNP in the Human Genome Diversity Panel (HGDP-CEPH) [18]. We also genotyped rs2272662 (which causes the Thr357Ala substitution) because, within the ZIP4 gene, it shows the second highest allele frequency differences between CEU and YRI HapMap populations and allele frequencies were not available at the worldwide level. The rs1871534 and rs2272662 loci were genotyped in the H971 subset [56] of the HGDP-CEPH [18], representing 51 worldwide populations, and in an additional population from Africa: Pygmies from Gabon (N = 39)[57]. We also genotyped rs1871534 in North African populations from Western Sahara (Saharawi, N = 50), Morocco (Casablanca, N = 30; Rabat, N = 30; Nadir, N = 30) and Libya (Libyans, N = 50). Genotyping was performed using Taqman assays C_11446716_10 and C_26034235_10 on an Applied Biosystems Light Cycler (7900HR), according to standard protocols. Additional genotypes for rs1871534 were obtained from the Alfed database (http://alfred.med.yale.edu) [26,27].

Ethics statement

Informed consent was obtained for all human samples analysed and genotyping analyses were performed anonymously. The project obtained the ethics approval from the Institutional Review Board of the local institution (Comite´ Etic d’Investigacio´ Clı´nica - Institut Municipal d’Assisti`ncia Sanita`ria (CEIC-IMAS) in Barcelona, Spain.

Neanderthal resequencing

The El Sidrón Neanderthal sample SD1253 has been used in many paleogenomic studies due to its high endogenous DNA content and low contamination levels [28,58–62], attributable in part to having been extracted using an anti-contamination protocol [63]. In addition, it has the advantage of having been dated to 49,000 years ago [64], prior to the arrival of modern humans to Europe. The two orthologous positions for rs1871534 and rs2272662 were amplified using a two-step PCR protocol [59] in a multiplexed reaction along with a diagnostic Neanderthal mitochondrial DNA (mtDNA) fragment. After visualizing the PCR products in a low-melting temperature agarose gel, the bands were excised, purified and cloned using the TOPO-TA cloning kit (Invitrogen). Inserts of the correct size were sequenced on an ABI3730 XL capillary sequencer (Applied Biosystems).

Simulations

Simultaneous coalescent simulation of recombination hotspots and selection were carried out using Cosi v.1.2 [31,65]. As the underlying neutral demography, we used the best-fit model of Shaffner et al. [31,65] with slight modifications (Table S3), similar to a previously used approach [66]. In particular, the migration frequencies were set to zero and the time points of the European and African population bottlenecks were moved back to 3,300 generations before present to accommodate the long sweep times resulting from the lowest selection coefficient we used (0.5%). The sweep was shifted back 350 generations to retain the final population expansions with the advantage of (i) a better approximation to the fitted model, and (ii) the generation of sufficient singletons when compared to the 1000 Genomes Phase1
Selection on a Human Zinc Transporter in Africa

Neutrality tests on simulated and the 1000 Genomes data

Neutrality tests on simulated and the 1000 Genomes population data were performed as described by Pybus et al. [68] and using the 1000 Genomes Selection Browser (http://hsb.upf.edu). Briefly, Tajima’s D, Fu and Li’s D and Fay and Wu’s H were calculated using a sliding window approach with 30 kb windows and approximately 3 kb offset. \( F_{ST} \) [69] and \( XP-EHH \) [70] between CEU and YRI were calculated for each polymorphic position.

Cells and reagents

Human \( ZIP4 \) cDNA encoding the long isoform of the protein and the Ala357 and Leu372 variants was cloned into pcDNA 3.1 (+) expression vector together with a hemagglutinin (HA) tag at the carboxyl terminus as discussed in the main text [71]. The Leu372Pro and Leu372Arg mutants, as well as the Thr357Ala and Leu372Val polymorphisms, were introduced via site-directed mutagenesis following standard conditions (QuickChange II XL, Stratagene; see Table S4 for complete human cDNA and primers used in the mutagenesis). The six human \( ZIP4 \) isoforms obtained (i.e. Ala357-Leu372, Ala357-Val372, Thr357-Leu372, Thr357-Val372 as well as Ala357-Pro372 and Ala357-Arg372) were confirmed by sequencing with the ABI Prism 3.1 BigDye kit before their use in transfection experiments. HeLa cells were cultured in DMEM plus 10% FBS and, subsequently, each of the various \( ZIP4 \) forms were transiently transfected using polyethylenimine as the transfection reagent (PolySciences).

Immunodetection

For the cell surface expression experiments, live cells were incubated with anti HA (1:1000) in DMEM without serum for 1h at 37° before fixation with 4% paraformaldehyde. After blocking for 30 min (1% BSA, 2% PBS in PBS), cells were incubated with a secondary antibody (1:2000) for 45 min in the blocking solution. For the total cell expression experiments, cells were permeabilized with 0.1% Triton in PBS for 10 min after fixation. Following blocking for 30 min (1% BSA, 2% PBS in PBS), cells were incubated in the blocking solution with anti HA (1:1000) for 1 h 30 min, washed with PBS and incubated with the secondary antibody (1:2000) for 45 min. Images were acquired using an inverted Leica SP2 confocal microscope with a 40 x 1.32 Oil Ph3 CS objective. Expression was quantified by measuring chemiluminescence with a plate reader (24-well plates) using peroxidase-linked anti-mouse antibody (GE Healthcare) as a secondary antibody and SuperSignal West Femto reagent as a substrate (Thermo scientific). Data are presented as the ratio between surface expression and total expression of the transporter. Statistical significance was tested using standard ANOVA.

Zinc measurements

Cells were transiently transfected with the various \( ZIP4 \) isoforms plus empty ECFP vector for 24–36 h. Cytosolic Zn2+ signal was determined in CFP-positive cells loaded with FhuOxZin3 2.5 μM (Invitrogen) in a solution containing 140 mM NaCl, 5 mM KCl, 1.2 mM CaCl₂, 0.5 mM MgCl₂, 5 mM glucose, 10 mM HEPES,
300 mosmol/l, pH 7.4 for 20 min. Cytosolic [Zn2+] increases are presented as the difference with respect to the basal signal of emitted fluorescence (510 nm) after adding 200 μM ZnSO4 in a continuous perfusion bath. The kinetics of the various isoforms were calculated using a sigmoidal non-linear regression. In the same set of experiments, basal cellular Zn2+ content was estimated as the difference in Fluozin intensity between transfected cells and non-transfected cells before adding Zn2+ to the bath. Fluorescence intensity was measured using an Olympus IX70 inverted fluorescence microscope, controlled by Aquacosmos software (Hamamatsu).

Supporting Information

**Figure S1** Worldwide allele frequencies for the Leu372Val (rs1871534, top) and Thr357Ala (rs2272662, bottom) polymorphisms. Circles are not proportional to sample sizes. Maps were generated with MapViewer. Complete list of population and singletons positions improves the fit of simulated data based on site frequency spectrum. Singleton positions improves the fit of simulated data based on site frequency spectrum. (TIF)

**Figure S2** Neandertal mt-DNA control for contamination. (TIF)

**Figure S3** Patterns of selection in a genomic region of 100 kb around the ZIP4 (SLC39A4) gene for three human populations. Gene context and summary of tests for positive selection obtained from the 1000 Genomes data for three populations: Yoruba from Ibadan, Nigeria (YRI), Han Chinese from Beijing, China (CHB) and Utah residents with Northern and Western European origin (CEU). With the exception of population differentiation (here: the FST statistic), those statistics based on site frequency (Tajima's D, Fu and Li's D, Fay and Wu's H) and haplotype structure (XPEHH) do not reach genome-wide significance (not shown) in any of the three populations. (TIF)

**Figure S4** Patterns of selection in a genomic region of 100 kb around the ZIP4 gene (SLC39A4). (A) Gene context and summary of tests for positive selection obtained in the Yoruba population from the 1000 Genomes data. Those statistics based on the site frequency spectrum (Fay and Wu's H, Fu and Li's D and Tajima's D) show weakly negative scores near ZIP4 that do not approach genome-wide significance (not shown), so they should not be regarded as indicative of positive selection. Those statistics based on population differentiation (here: FST) show three SNPs (see Figure 1) with elevated values between CEU and YRI. One of them, rs1871534 (Leu372Val), is among the most highly differentiated SNPs in the genome. (B) Fine-scale recombination rate from the Yoruba population plotted on a linear scale reveals a moderate recombination hotspot near SLC39A4. (C) Detailed view of simulated values along the 100 kb region for different statistical tests of positive selection assuming different scenarios comparable to Figure 1: (i) no selection and considering the observed recombination landscape from the Yoruba population (black lines); (ii) a selective sweep in the West African population and a constant recombination rate (orange lines); and (iii) a selective sweep in the West African population and the observed recombination landscape including the hotspot (red lines). Statistics were calculated in a sliding window approach with 30 kb windows and approximately 3 kb offset. For FST only the maximum score for each window was considered. Solid lines indicate median values and dashed lines indicate the 5th and the 95th percentiles of 500 replicated simulations. (TIF)

**Figure S5** Detection of ZIP4 isoforms by western blot. (A) Gel was loaded with 80 μg of total protein extracts from HeLa cells transiently transfected with the different ZIP4 isoforms. Anti-HA antibody (1:1000) was used to detect the transporters and anti-actin (1:3000) as a loading control. (B) HeLa cells transfected with the Ala357-Leu372, Ala357-Val372, and Ala357-Pro372 isoforms were treated with 10 μg/ml cyclohexamide for different time periods (1 h, 3 h, 6 h and 8 h). Total protein extracts were obtained and western blotting was performed. A representative experiment for each isoform is shown (left). The quantification analysis normalized the band intensity to the initial amount of protein before the treatment (time 0) (right). This experiment was performed three times per isoform (n = 3). (TIF)

**Figure S6** Retention of ZIP4 in the endoplasmatic reticulum. Immunostaining under permeabilizing conditions on cells expressing different ZIP4 variants using anti-HA (1:1000) for ZIP4 detection and anti-calnexin (1:1000) (Abcam) as an endogenous endoplasmatic reticulum marker protein. (TIF)

**Figure S7** Linkage disequilibrium plot for the YRI population in a 50 kb window around the ZIP4 (SLC39A4) gene. The plot was generated with Haploview and using HapMap 2 data (release 21). (TIF)

**Figure S8** Haplotype visualization in a 40 kb window around ZIP4 (SLC39A4) gene. Plots from the HapMap browser (http://hapmap.ncbi.nlm.nih.gov) are shown for the Yoruba, Han Chinese and French populations. There is no indication of extended haplotype patterns that could indicate a classical selective sweep in any of the three populations. (TIF)

**Figure S9** Demographic models versus empirical data. Empirical results based on the 1000 genomes data (only chromosome 1) are compared to an established demographic model [31] and against two demographic models adapted for capturing selective sweeps (Grossman et al. [66] and the present study) using neutral simulations of 500 kb length (500 replicates; roughly matching the length of chromosome 1). (A) Derived allele frequency distributions based on data or simulations reflecting African and European genetic origin. (B) FST distributions in a pair-wise population approach. As indicated in the text, the thinning of random singleton positions improves the fit of simulated data based on site frequency spectra. (TIF)

**Table S1** Worldwide allele frequencies for the Leu372Val (rs1871534) and Thr357Ala (rs2272662) polymorphisms. (PDF)

**Table S2** Empirical probabilities under neutrality and different selection scenarios. (PDF)

**Table S3** Simulation parameters similar to best-fit model Schaffner et al. [31]. (PDF)

**Table S4** Description of primers and hcDNA used in mutagenesis. (PDF)

Acknowledgments

We thank Jean-Marie Hombert and Lolie van der Veen (Dynamiq du Langage, Institut des Sciences De l'Homme, Lyon France) as well as Patrick Mouguiama Daouda (University Omar Bongo, Libreville, Gabon).
and the Centre International des Recherches Médicales de Franceville (CIRMF, Gabon) for sharing the Gabon samples. We also acknowledge Karima Fadhalou-Ziel from the University Tunis El Manar (Tunisia) for providing the Libyan samples and Francesc Calafell and Pierre Luisi for discussion as well as Nino Spataro and the National Institute of Bioinformatics (http://www.inab.org) for computational support.

Author Contributions
Conceived and designed the experiments: JE GKA CLF MS MAV RV EB. Performed the experiments: JE ECG CLF RV. Analyzed the data: JE ECM MP RV EB. Contributed reagents/materials/analysis tools: GKA CLF DC MdR AR MAV RV EB. Wrote the paper: JE ECG CLF MS MAV RV EB.

References

PLOS Genetics | www.plosgenetics.org
13 February 2014 | Volume 10 | Issue 2 | e1004128

Selection on a Human Zinc Transporter in Africa


