The Loss of Functional Caspase-12 in Europe Is a Pre-Neolithic Event

Montserrat Hervella1*, Theo S. Plantinga2, Santos Alonso1, Bart Ferwerda2,3, Neskuts Izagirre1, Lara Fontecha1, Rosa Fregel4, Jos W. M. van der Meer2,5, Concepcion de-la-Rúa1, Mihai G. Netea2,5

1 Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country, Leioa, Spain, 2 Department of Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, 3 Department of Genetics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, 4 Department of Genetics, University of La Laguna, La Laguna, Spain, 5 Nijmegen Center for Infection, Inflammation and Immunity (N4i), Nijmegen, The Netherlands

Abstract

Background: Caspase-12 (CASP12) modulates the susceptibility to sepsis. In humans, the “C” allele at CASP12 rs497116 has been associated with an increased risk of sepsis. Instead, the derived “T” allele encodes for an inactive caspase-12. Interestingly, Eurasians are practically fixed for the inactive variant, whereas in Sub-Saharan Africa the active variant is still common (~24%). This marked structure has been explained as a function of the selective advantage that the inactive caspase-12 confers by increasing resistance to infection. As regards to both when positive selection started acting and as to the speed with which fixation was achieved in Eurasia, estimates depend on the method and assumptions used, and can vary substantially. Using experimental evidence, we propose that, least in Eurasia, the increase in the frequency of the T allele might be related to the selective pressure exerted by the increase in zoonotic diseases transmission caused by the interplay between increased human population densities and a closer contact with animals during the Neolithic.

Methodological/Principal Findings: We genotyped CASP12 rs497116 in prehistoric individuals from 6 archaeological sites from the North of the Iberian Peninsula that date from Late Upper Paleolithic to Late Neolithic. DNA extraction was done from teeth lacking cavities or breakages using standard anti-contamination procedures, including processing of the samples in a positive pressure, ancient DNA-only chamber, quantitation of DNAs by qPCR, duplication, replication, genotyping of associated animals, or cloning of PCR products. Out of 50, 24 prehistoric individuals could finally be genotyped for rs497116. Only the inactive form of CASP12 was found.

Conclusions/Significance: We demonstrate that the loss of caspase-12 in Europe predates animal domestication and that consequently CASP12 loss is unlikely to be related to the impact of zoonotic infections transmitted by livestock.


Editor: David Caramelli, University of Florence, Italy

Received December 16, 2011; Accepted April 12, 2012; Published May 16, 2012

Copyright: © 2012 Hervella 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: This work was supported by a Vici grant 91810610 of the Netherlands Organisation for Scientific Research to MGN, and by the Spanish Ministry of Science and Innovation, CGL-2007-65515 and grants IT453-07 and IT542-10 from the Basque Government to Research Groups of the Basque University System to CR, and (UFI 11/09) from the University of the Basque Country, UPV/EHU. 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: montse.hervella@ehu.es

Introduction

Caspase-12 (CASP12) is a cysteine protease that exerts apoptotic functions, inhibits the inflammatory response and modulates susceptibility to sepsis [1–6]. In contrast to most mammals, humans show a nonsense C>T mutation at aminoacid 125 in exon 4 (rs497116), which defines two alleles. The ancestral “C” allele (long L-variant) encodes a full-length functional protein which has been associated with an increased risk of sepsis [7]. The derived “T” allele (short S-variant) encodes for a prematurely terminated, inactive caspase-12. Europeans and Asians are practically fixed for the inactive variant, whereas in Sub-Saharan Africa the active variant is still common (~24% on average, Table S1). This marked structure has been explained as a function of the selective advantage that the inactive caspase-12 confers by, supposedly, increasing resistance to infection. It is thought that the mutation originated in Africa before the Out-of-Africa expansion of modern humans [8,9], but as regards to both when positive selection started acting and as to the speed with which fixation was achieved in Eurasia, estimates depend on the method and assumptions used, and can vary substantially. Thus, according to [8], a) the age of the inactive allele can be 980KY if it is considered neutral (no SD provided); b) if instead, the allele is under positive selection (s~1%), the estimated age is 27,000 years (no SD provided); c) using an alternative phylogenetic method, the age can be 552+/−276 KYA or 397+/−223 KYA or 61+/−16 KYA, depending on the root haplotype considered; d) using a parametric model that predict the spatial pattern of nucleotide diversity and allele-frequency spectrum around the putative target of selection the estimated selection coefficient is 1.7% and the age for the mutation is ~19KYA (no SD provided); e) using a full-likelihood method then s~0.8% and the age is ~29KYA (no SD provided).
provided; f) they do not investigate selection from standing variation. On the other hand [9], estimate: a) using the information on non-coding region 4, the age of the allele is between 0 and 154 KYA, with an average at 74,250 years ago; b) using a deterministic selection model with parameters $w_{CC} = 0.991$, $w_{CT} = 0.999$ and $w_{TT} = 1$, with $h = 0.11$, then, the onset of the selective sweep for the T allele would have been between 51–55 KYA; c) under neutrality, the ages of the T allele is estimated as 943 KYA; d) They do not investigate selection from standing variation. To this, we have to add the persistence of the active CASP12 allele in Sub-Saharan Africa.

An assessment of the distribution of allele frequencies of rs497116 in Africa suggests an influence of farming on this locus. Thus, hunter-gatherer populations (Mbuti Pygmies and San) (Table S1) show a T-allele frequency of ~0.4, whereas for the rest of the populations, farmers, it rises to ~0.8. The case of the Biaka pygmies, with a frequency of ~0.8, would seem an exception, but it has been documented that Western Pygmies have had a substantial level of gene flow with neighboring farmers [10]. Therefore, we argue that, at least in Eurasia, the increase in the frequency of the T allele at CASP12 rs497116 might be related to the selective pressure exerted by husbandry development. Thus, the interplay between increased human population densities and a closer contact with animals, would lead to an increase in zoonotic diseases transmission [11]. In this scenario, the selective advantage of the inactive CASP12 allele is compatible with the increased resistance to infections that it would confer.

We decided to test this hypothesis by genotyping CASP12 rs497116 in prehistoric individuals from 6 archaeological sites from the North of the Iberian Peninsula (Figure 1) that date from Late Upper Paleolithic to Late Neolithic (Table 1). Faunistic domestic remains suggest that animal husbandry may have started, at least in the Basque Country, ~6,000-5,500 YBP [12]. Therefore, this sample seems ideal to explore whether the complete loss of caspase-12 function in humans was attained before or after the Neolithic transition. Note that herein we define Neolithic in terms of the development of animal husbandry, regardless of agriculture or other cultural industries. Out of 50, 24 prehistoric individuals could be genotyped for rs497116.
**Results and Discussion**

We first decided to explore if it is actually possible that Neolithic could have substantially shifted the frequencies of sub-Saharan Africans and Eurasians in a time span ranging from the onset of Neolithic to present time. For Africans, analytical formulae (see Material and methods) show that it is actually possible to increase a mean frequency from 0.4 (frequency before Neolithic; inferred from present-day sub-Saharan African hunter-gatherers) to 0.8 (present-day frequency in sub-Saharan African farmer populations) in ~200 generations (~4,000-5,000 years, an estimated time span from the onset of husbandry in sub-Saharan Africa to present time) under a selective regime with a selection coefficient ($s$) of 1%, and a heterozygous effect parameter ($h$) of 0.11, $s$ and $h$ taken from [8]. Further, our own simulations show that for a neutral allele whose origin is in Africa before the Out-of-Africa expansion and whose frequency in Africa is between 0.3–0.5, the probability that in Europe its frequency at present time is $\geq 0.95$ is 0.021, which rejects the hypothesis of a neutral evolution of CASP12 rs497116. If we assume that selection on rs497116 is 0.021, which rejects the hypothesis of a neutral evolution of CASP12 rs497116.

In conclusion, we demonstrate that the spread of the inactive allele of CASP12 in Europe predates animal domestication, and that consequently CASP12 loss in Eurasians is unlikely to be related to the impact of zoonotic infections transmitted by livestock. It does not look likely either that the ecological change brought about by Neolithic, i.e., anthropogenic modification of the environment affecting the pathogen or pathogen vector’s ecology [16], could have triggered the present day distribution of CASP12 rs497116 alleles. Then, which is the selective agent is responsible for the wide distribution of the CASP12 rs497116 alleles? Malaria immediately comes to mind, but previous investigations have also ruled out malaria as a selective agent responsible for the persistence of the active CASP12 allele in Africa [17]. At this point we cannot but speculate that perhaps the predatory/scavenging activity of humans could have triggered a selective response. It would be interesting to know the rs497116 genotypes of earlier humans. Unfortunately, none of the Neanderthal nuclear sequences available overlap this SNP. Perhaps future work analyzing a fair set of early Homo sapiens remains could help us bracketing the time point for the expansion of the inactive allele, and thus, we could be able to infer which environmental scenario was responsible for it.
Materials and Methods

Estimation of allele frequency after selection using analytical formulae

To assess if the development of agriculture could have driven a shift from a frequency of 0.4 to a present day frequency of 0.8 as a result of selection in Africa we used the following formula [18]:

\[ p^t = (p^2 w_{11} + pq w_{12}) / \bar{W} \]

where \( p \) and \( q \) are the allele frequencies of alleles \( A \) and \( a \) respectively, \( p^t \) is the frequency of allele \( A \) after 1 generation of selection (under constant population size), \( w_{11} \) and \( w_{12} \) are the viabilities of genotypes \( AA \) and \( Aa \) respectively, and \( \bar{W} \) is the mean fitness of the population. If the relative fitness of \( AA \) is 1, the relative fitness of \( Aa \) will be:

\[ 1 - s = w_{22} / w_{11} \]

and that of \( aa \), 1 - \( h \), where 1 - \( h \) = \( w_{12} / w_{11} \), \( h \) being the heterozygous effect.

Using this formula as a recursive algorithm, it can be shown that an allele \( A \) with an initial frequency of 0.4, under a selective regime with \( s = 1\% \) and \( h = 0.11 \), as suggested by the literature for \( CASP12 \) (see B) can reach a mean frequency \( p^t \) of 0.8 in 200 generations (~4,000 years).

Estimation of the probability of allele frequencies by means of simulations

We used SFS_CODE [19], a generalized Wright-Fisher style forward population genetic simulation program for finite-site mutation models with selection, recombination and demography, to infer the probabilities of obtaining an allele frequency greater than or equal to 0.95 in Europe, provided that that allele originated in Africa before the Out-of-Africa expansion of modern humans took place, and that its present day frequency in Africa is between 0.3 - 0.5 (as observed in African hunter-gatherer populations [Table S1]). Note that SFS_CODE does not allow simulating selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin. The command line, embedded within a Perl script, for the neutral allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.

We explored two scenarios. One in which the allele remains neutral along its evolutionary time, and a second one, in which the allele holds a selective coefficient \( s \) of 1% (additive) since its origin.

The command line, embedded within a Perl script, for the neutral along its evolutionary time, and a second one, in which the selection from standing variation.
For the standard curves, serial dilutions of plasmid pCR2.1-new (3.9 Kb) including an insert of 450 bp (Eurofins MWG/Oxford) containing the HVR-I region of interest, are included in each experiment to generate standard curves. Two different standard curves were performed. One with 3 points of 1.4*10^5, 1.4*10^4 and 1.4*10^3 molecules/µl for high concentration samples, and a second standard curve for low concentration samples, with 3 points at 1.12*10^4, 2.24*10^3 and 448 molecules/µl. Four replicates were used for each dilution point. Typical (%efficiency, r^2) values were respectively (102%, 0.99) for the high concentration curve, and (83.2%, 0.95) for the low concentration samples. Finally, at least three “no-template-controls” were included with each experiment. Quantification of results can be seen in Table S3.

CASP12 genotyping

CASP12 genotyping was done by direct sequencing of PCR products. Each extract was sequenced twice. The PCRs were performed in 25 µl of reaction mixture containing 10 mM Tris-HCl pH 8.3; 2 mM of MgCl2, 0.1 µM of each dNTP, 0.4 µM of each primer, 10 units of AmpliTaq Gold (Applied Biosystems) and 10 µl of diluted DNA (1 µl of DNA extract in 10 µl of 1 mg/ml BSA). Cycling parameter were 96°C 1 min for 1 cycle, followed by 45 cycles of 95°C 15 s, 48°C 30 s, 72°C 30 s and a final cycle of 72°C 10 min, using Forward primer 5’-CACTATGTCTCTATATCGAAA-3’ and Reverse primer 5’-TTTTATACCTACTGAGTATCCT-3’. These primers define amplicons of 91 bp (49 bp without the primer sequences). Both forward and/or reverse sequences were obtained using the above primers and Rhodamine chemistry in an ABI310 Genetic Analyzer (Applied Biosystems).

mtDNA HVR I sequencing

The sequencing of the HVR I (15,998-16,400) was carried out by sequencing 6 overlapping fragments of approximately 100 bp each, as described in [24], although in this case the amplification of each fragment was done in independent PCRs. The PCRs were performed in 25 µl of reaction mixture containing 10 mM Tris-HCl pH 8.3; 2 mM of MgCl2, 0.1 µM of each dNTP, 0.4 µM of each primer, 5 units of AmpliTaq Gold (Applied Biosystems) and 10 µl of diluted DNA (1 µl of DNA extract in 10 µl of 1 mg/ml BSA).

Cycling parameter were 95°C for 10 min; followed by 40 cycles of 95°C for 10 sec, annealing temperature for 30 sec, 72°C for 30 sec; and 72°C for 10 min. The annealing temperatures of the primers were as follows: 60°C for the A1/A1R primer pair, 58°C for 2F/2R and 4F/4R, 57°C for 1F/1R and 55°C for 3F/3R and 5F/5R (the primer sequences are listed in [24]). In the event of positive amplification and absence of contamination, the amplifications were purified by Exo SAP IT (USB corporation) and subsequent sequencing in an ABI310 automatic sequencer using chemistry based on dRhodamine. The results obtained were edited with the BioEdit software application (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Sequences were aligned by eye.

Reproducibility of the results

a) Duplication of CASP12 sequencing results

For each of the samples, one extract (one aliquot) was sequenced twice from independent reactions with BigDye 1.1 chemistry using the Reverse primer only. Both sequences were checked to be coincident at rs497116.

b) Cloning of CASP12 PCR products and sequencing of clones

For 13 samples, CASP12 PCR products (see above) comprising rs497116 were cloned (TOPO TA Cloning® Kit, Invitrogen) and from each cloning reaction at least 10 colonies were picked up, sequenced and checked for coincident results at rs497116. In this case most clones showed the same haplotype as the original PCR product as expected, although some clones did show some sequence variability outside rs497116. A mean of 1.8 mutations (1.2 A→G; 0.6 G→A) per sample cloned (~80 bp) were found in unique clones, which indicates limited post-mortem damage to the aDNA molecules (rather than contamination with exogenous DNA); (see Table S4).

c) CASP12 replication in independent laboratories

For the CASP12 locus, seven samples (five from the SJAPL site and one each from the Erralla and Marizulo sites) were analyzed in duplicate in two different labs (University of the Basque Country and University of La Laguna, Spain). For this, two different teeth from each sample, one for each lab, were extracted and genotyped independently in each lab. Six samples were coincident (TT) and for one sample from SJAPL no PCR product could be obtained in one of the labs.

d) mtDNA sequencing

Because SNP rs497116 is monomorphic in European populations, to increase the chances of detecting possible contamination mitochondrial DNA sequences were obtained from 33 samples by direct sequencing as depicted in Table S5.

e) Duplication of mtDNA sequencing

In addition, for 9 samples the same extract was analyzed twice for their mtDNA (either in the University of the Basque Country, Spain, or in the Radboud University, The Netherlands). MtDNA was chosen because its high variability increases the chances to genetically individuate each sample, and therefore to increase the chances to detect (and track) contamination with exogenous DNA.

f) Replication of mtDNA sequencing

We assessed further the reproducibility of the results obtained in the Univ. of the Basque Country by replicating the analysis of the six overlapping fragments of mtDNA HVRI of a second tooth from each of 2 prehistoric individuals in the University of La Laguna (Spain) (Table S5).

g) Cloning and sequencing of mtDNA PCR products

To further corroborate the results, PCR products corresponding to one of these fragments (~110 bp) of mtDNA HVRI of 24 different samples (from the La Pasiega, La Chora, Erralla, SJAPL, Longar and Marizulo sites) were cloned (using the TOPO TA Cloning® Kit, Invitrogen). From each cloning reaction at least 10 colonies were picked up, sequenced and checked for coincident results. In this case most clones showed the same haplotype as the original PCR product as expected, although some clones did show some sequence variability outside key mutations. A mean of 5.8 mutations (1.8 A→G, 2.2G→A, 1.5 insertion of A, 0.04 insertion of G, 0.2 C→T, 0.08 T→G) per sample fragment cloned (~100 bp) were found in unique clones, possibly as a consequence of post-mortem damage to the aDNA molecules (rather than contamination with exogenous DNA) (see Table S4).

h) Genotyping of individuals involved in the study

Finally, the mtDNA types of most of the personnel that was in contact with the samples was also obtained (see Table S3) in order to minimize the possibility of contamination from the manipulators.

i) Genotyping of animal remains found at the sites

In addition, two dog (Canis familiaris) samples, found at the sites SJAPL and Marizulo respectively, and one chamois (Rupicapra rupicapra) from the Erralla site could be analyzed. None of them produced PCR products after amplification with mammalian-specific cytochrome b primers [25]. One dog did not produce positive results with mammalian-specific cytochrome b primers.
Supporting Information

Table S1  Distribution of CASP12 alleles at aminoacid position 125 (rs497116) in the populations studied and other world populations. The cytosine (C) from the ancient allele has been substituted in European populations by a thymidine (T), leading to a non-sense mutation and an inactive truncated molecule.

(DOC)

Table S2  Number of samples analyzed and proportion of which were positive.

(DOC)

Table S3 Mitochondrial haplotypes (HVR-I) and CASP12 rs497116 genotypes of researchers in this study.

(DOC)

Table S4 Mitochondrial DNA and CASP12 sequences of the clones from the samples analyzed in the present study (xls).

(XLS)

References


Table S5 Summary of quantification, replication and cloning.

(DOC)

Acknowledgments

We are grateful to the archaeologists and to the institutions that granted permission for human remains, including the Government of Navarra, the Basque Government (Cultural Heritage Dept.), the Archaeology and Palaeontology Heritage Centre of Gipuzkoa, the Museum of Archaeology of Alava and the Museum of Prehistory and Archaeology of Cantabria.

Author Contributions

Conceived and designed the experiments: SA JWMM CR MGN. Performed the experiments: MH TSP SA BF NI LF RF. Analyzed the data: MH TSP SA BF NI JWMM CR MGN. Contributed reagents/materials/analysis tools: CR. Wrote the paper: MH SA CR MGN. Revised the manuscript critically for important intellectual content: BF LF RF. Approved final version to be published: MH TSP SA BF NI MH LF RF JWMM CR MGN.

Pre-Neolithic Loss of CASP12