REPORT

A Selective Sweep on a Deleterious Mutation in *CPT1A* in Arctic Populations

Florian J. Clemente,[1,19] Alexia Cardona,[1,19,*] Charlotte E. Inchley,[1] Benjamin M. Peter,[2] Guy Jacobs,[3,4] Luca Pagani,[1] Daniel J. Lawson,[5] Tiago Antão,[6] Mário Vicente,[1] Mario Mitt,[7] Michael DeGiorgio,[8] Zuzana Faltyskova,[1] Yali Xue,[9] Qasim Ayub,[9] Michal Szpak,[9] Reedik Mägi,[7] Anders Eriksson,[10,11] Andrea Manica,[10] Maanasa Raghavan,[12] Morten Rasmussen,[12] Simon Rasmussen,[13] Eske Willerslev,[12] Antonio Vidal-Puig,[9,14] Chris Tyler-Smith,[9] Richard Villems,[15,16,17] Rasmus Nielsen,[2] Mait Metspalu,[15,16] Boris Malyarchuk,[18] Miroslava Derenko,[18] and Toomas Kivisild[1,16,*]

Arctic populations live in an environment characterized by extreme cold and the absence of plant foods for much of the year and are likely to have undergone genetic adaptations to these environmental conditions in the time they have been living there. Genome-wide selection scans based on genotype data from native Siberians have previously highlighted a 3 Mb chromosome 11 region containing 79 protein-coding genes as the strongest candidates for positive selection in Northeast Siberians. However, it was not possible to determine which of the genes might be driving the selection signal. Here, using whole-genome high-coverage sequence data, we identified the most likely causative variant as a nonsynonymous G>A transition (rs80356779; c.1436C>T [p.Pro479Leu] on the reverse strand) in *CPT1A*, a key regulator of mitochondrial long-chain fatty-acid oxidation. Remarkably, the derived allele is associated with hypoketotic hypoglycemia and high infant mortality yet occurs at high frequency in Canadian and Greenland Inuits and was also found at 68% frequency in our Northeast Siberian sample. We provide evidence of one of the strongest selective sweeps reported in humans; this sweep has driven this variant to high frequency in circum-Arctic populations within the last 6–23 ka despite associated deleterious consequences, possibly as a result of the selective advantage it originally provided to either a high-fat diet or a cold environment.

Siberia, where local temperatures occasionally drop below −70°C in the winter and only animals are available for consumption for much of the year, is one of the most extreme habitats human populations have adapted to since their dispersal out of Africa. A high basal metabolic rate, low levels of serum lipids, and high blood pressure are among the characteristics considered to be consequences of adaptation in Siberian populations.[1,2] In a recent genome-wide SNP genotype study of 200 Siberian individuals,[3] the strongest signals of positive selection detected by tests for haplotype homozygosity and allele differentiation mapped to a 3 Mb region containing 79 protein-coding genes at chr11: 66–69 Mb in Northeast Siberian populations. Because of the limited density of markers in the SNP data, it was impossible to pinpoint the causative locus for the selection signal. Here, we sequenced the genomes of 25 unrelated individuals from the Chukchi, Eskimo, and Koryak populations (Figure S1, available online) with a mean coverage of >40× by using the Complete Genomics platform (Tables S1A and S1B). Informed consent was obtained from all human subjects, and the study was approved by the ethics committee of the Institute of Biological Problems of the North of the Russian Academy of Sciences in Magadan (statement no. 001/011 from January 21, 2011) and by the Cambridge Ethics Committee (HBREC.2011.01). We used these sequence data to search for derived variants that are common in Northeast Siberians and rare or absent elsewhere. We then applied sequence-diversity-, derived-allele-frequency-, and haplotype-homozygosity-based methods to identify the possible causative variant(s) driving the signal and to estimate their age and strength of selection.

We applied QIAGEN's Ingenuity Variant Analysis software to 21,105,873 variants detected in 25 Northeast Siberian samples by using a series of filtering steps. A total of 14,183,704 variants with call quality and depth higher than 20 were outside 0.2% of the most variable exonic 100 bp regions and outside 1% of the most variable exonic genes in the 1000 Genomes data.[4,5] We excluded variants with >1% frequency in public data (including

[1]Department of Archaeology and Anthropology, University of Cambridge, Cambridge CB2 3QG, UK; [2]Department of Integrative Biology, University of California, Berkeley, Berkeley, CA 94720-3140, USA; [3]Mathematical Sciences, University of Southampton, Southampton SO17 1BJ, UK; [4]Institute for Complex Systems Simulation, University of Southampton, Southampton SO17 1BJ, UK; [5]Heilbronn Institute, School of Mathematics, University of Bristol, Bristol BS8 1TH, UK; [6]Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK; [7]Estonian Genome Center, University of Tartu, Tartu 51010, Estonia; [8]Department of Biology, Pennsylvania State University, University Park, PA 16802-5301, USA; [9]Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK; [10]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK; [11]Integrative Systems Biology Laboratory, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia; [12]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen 1350, Denmark; [13]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kongens Lyngby 2800, Denmark; [14]Medical Research Council Metabolic Diseases Unit, Department of Clinical Biochemistry, University of Cambridge and Institute of Metabolic Science, Cambridge CB2 2QR, UK; [15]Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia; [16]Estonian Biocentre, Tartu 51010, Estonia; [17]Estonian Academy of Sciences, Tallinn 10130, Estonia; [18]Institute of Biological Problems of the North, Russian Academy of Sciences, Magadan 685000, Russia
[19]These authors contributed equally to this work
*Correspondence: ac812@cam.ac.uk (A.C.), tk331@cam.ac.uk (T.K.)
http://dx.doi.org/10.1016/j.ajhg.2014.09.016.

1000 Genomes data,[4,5] Complete Genomics public genomes,[6] and NHLBI Exome Sequencing Project data) and retained 8,278 variants with >50% carrier frequency in our Northeast Siberian sample. We applied additional filters, excluding indel and substitution variants, and retained 148 SNPs with nonreference allele frequency >50%. Among these 148 SNPs (Table S2), we detected three with a possible functional consequence predicted by their location in exonic, promoter, or enhancer regions or among phylogenetically conserved positions (phyloP value $< 10^{-3}$). Note that unrecognized regulatory regions that might be under selection would not be detected by these filters. Among these three mutations, two mapped to the same chromosomal region (chr11: 68.5–68.7 Mb), which lay within the previous candidate region of positive selection in Siberian populations:[3] one nonsynonymous (c.1436C>T) mutation in carnitine palmitoyltransferase I (*CPT1A* [MIM 600528]) and another in the promoter region of immunoglobulin mu binding protein 2 (*IGHMBP2* [MIM 600502]; position 68,670,984). To investigate whether one of these mutations could be the target of positive selection, we applied further neutrality tests and explored the genetic variation in this region.

First, we searched genome-wide in the sequence data for candidate regions under selection. We merged the 25 Northeast Siberian genomes with a control panel of 25 European and 11 East Asian publicly available, high-coverage Complete Genomics genomes (Personal Genomes Project)[6] (Table S3). The integrated haplotype score (iHS) test[7] confirmed our previous findings[3] in that the largest number of significant (here, top 1%) windows fell within the chr11: 66–69 Mb region (Figure S2; Table S4). The two top-ranking iHS windows mapped within the chr11: 66–69 Mb region, and the window containing *CPT1A* ranked 45th among 13,035 genomic windows, whereas the window containing *IGHMBP2* was not significant (Figure 1). In the control populations, neither of these windows was significant.

Second, we performed genome-wide Tajima's D[8] scans, which highlighted two windows with significantly negative D values in the chr11: 66–69 Mb region (Figure 1; Table S5). These two windows narrowed the 3 Mb selection candidate region to 400 kb (68.2–68.6 Mb), which contains five protein-coding genes: low-density lipoprotein receptor-related protein 5 (*LRP5* [MIM 603506]), protein phosphatase 6, regulatory subunit 3 (*PPP6R3* [MIM 610879]), galanin (*GAL* [MIM 137035]), metallothionein-like 5, testis-specific (*MTL5* [MIM 604374]), and *CPT1A*. As with iHS, the window containing *IGHMBP2* was not significant in the Tajima's D scans. Also, the Tajima's D statistic in the 400 kb region was not significantly different from the genomic average in the control populations (Figures 1 and S3). Thus, both iHS and Tajima's D tests provided evidence that the selection signal is restricted to Northeast Siberians and favored *CPT1A* over *IGHMBP2* as the target of selection.

To corroborate this conclusion, we considered the derived site-frequency spectrum (SFS) in the region surrounding *CPT1A* against the background, allowing us to study deviations from neutrality without the assumption of mutation-drift equilibrium. In Northeast Siberians, the SFS in the 68–69 Mb region shows an excess of low- and high-frequency variants in comparison to the background SFS of the whole genome, consistent with the expectation under a selective sweep[9,10] (see Figure S4). Moreover, we investigated the pattern of linkage disequilibrium (LD) in the 3 Mb region (chr11: 66–69 Mb) and found two distinct LD blocks, one on each side of the c.1436C>T mutation, within the region of 68.2–68.8 Mb (Figure S5). Such a pattern is expected in the later stages of a selective sweep, when the beneficial allele has reached a frequency of >0.5.[11,12] Therefore, the LD pattern provides further evidence that the c.1436C>T mutation could be driving the signal. However, the c.1436C>T mutation, which is also frequent in other Arctic Inuits, has previously been associated with high infant mortality and hypoketotic hypoglycemia (carnitine palmitoyltransferase I deficiency [MIM 255120]) in Canadian Inuits, and its high frequency in these populations has been described as a "paradox" (e.g., Greenberg et al.[13]).

To further explore the global distribution of c.1436C>T, we measured its frequency in a global panel of modern and ancient genomes and created a median-joining network of the locus (Figure 2). In the modern samples, we found c.1436C>T at 68% frequency in the Northeast Siberian populations, but it was absent in other publicly available genomes.[4,5] In the ancient samples, the c.1436C>T mutation was absent in both the c. 24-ka-old Mal'ta boy[14] and the c. 12.6-ka-old Clovis sample[15] but was heterozygous in the genome of the c. 4-ka-old Saqqaq Palaeo-Eskimo.[17] Furthermore, either the c.1436C>T mutation or its associated haplotype was detected in a number of Canadian and Greenlandic Pre-, Middle-, and Late-Dorset ancient-DNA samples[16] dating to the last 4 ka and had a combined frequency of ~50% (Figure 2A; Table S7). The ancient-DNA evidence shows that the c.1436C>T mutation has a minimum age of c. 4 ka, whereas the ancient and modern samples together suggest that its spread was restricted to Eskimo-Aleutian and Northeast Siberian populations. The geographic distribution of the haplotype from which the c.1436C>T haplotype is derived was mainly restricted to East Asia (Figure 2A). Notably, in contrast to the recent finding that *CPT1A* was among the lipid-catabolism-associated genes enriched with Neandertal-like sites,[20] the c.1436C>T mutation occurs at the background of 27 other SNPs that are common in Siberians and rare or absent elsewhere. None of these 28 SNPs (Table S7) shared Neandertal or Denisovan ancestry.

Finally, we used different models to infer population parameters from the data. Based on the observed genetic diversity,[8,21] LD,[22] and multiple sequentially Markovian coalescent,[23] our estimates of the effective population size, N_e, yielded a long-term average N_e of 3,000–5,000 and supported a constant-size model (Figure S6; Table S8). In contrast, a bottleneck scenario, in which the c.1436C>T

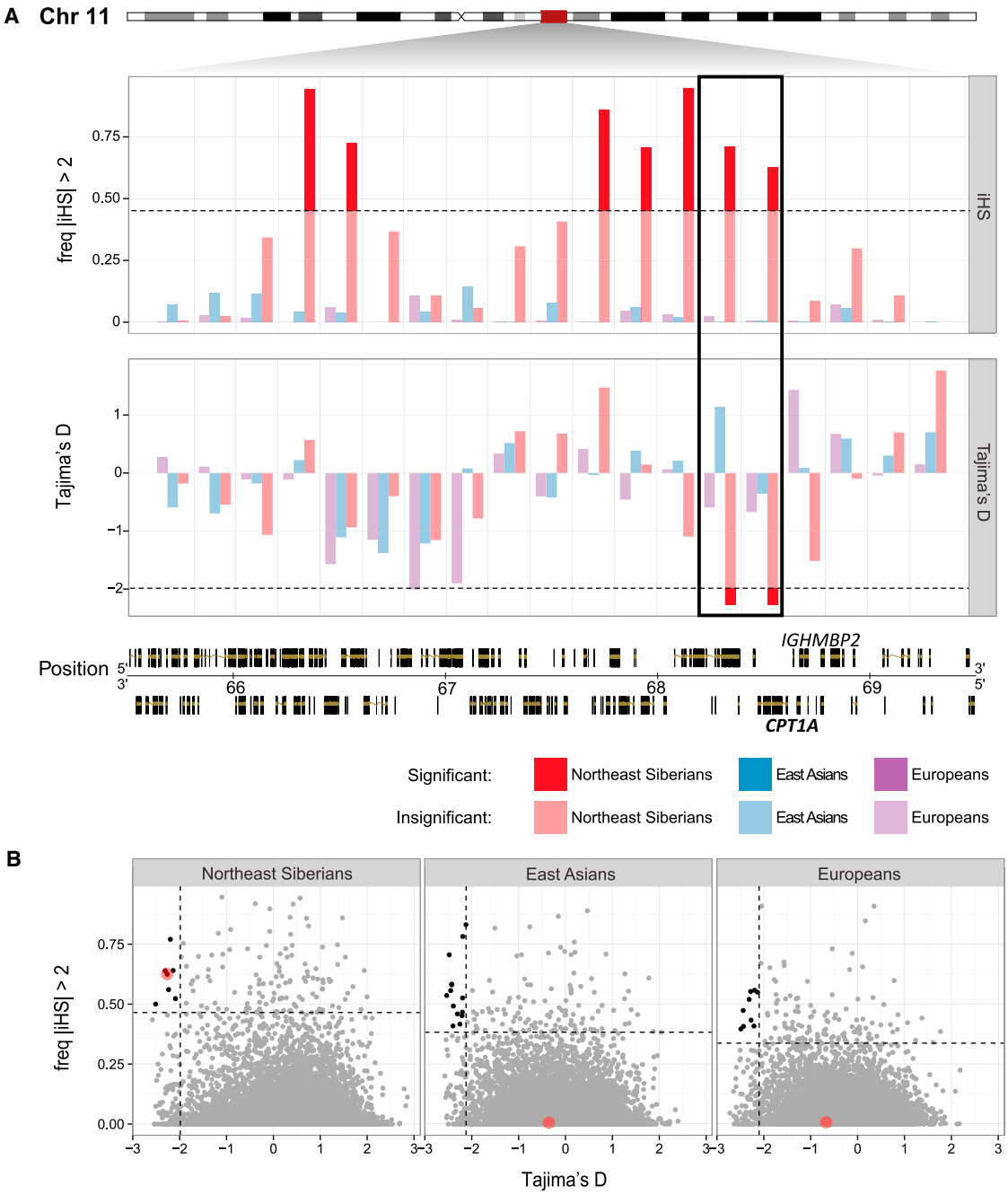

Figure 1. Localization of Positive-Selection Signals within a 3 Mb Region on Chromosome 11
(A) A high concentration of significant iHS (top) and Tajima's D (bottom) signals was found in a 3 Mb region (chr11: 66–69 Mb) in Northeast Siberians. The results are shown in the context of the East Asian and European control populations. Pale gray lines highlight the boundaries of each 200 kb window in the region. The horizontal black dotted line marks the threshold of 1% significance. The overlapping 400 kb region (chr11: 68.2–68.6 Mb) of significant results from both tests is highlighted by a bold black rectangle.
(B) The genome-wide selection scans show that the window containing *CPT1A* (chr11: 68.4–68.6 Mb), highlighted by the red dot in the three plots, is significant in the Northeast Siberian populations, but not in the control populations (Europeans and East Asians). The black dots are the significant data points in both iHS and Tajima's D tests. Dotted lines show the significance thresholds of the respective tests.

mutation would have reached high frequencies from standing variation by drift, was not supported because the c.1436C>T mutation is absent in other global populations and the derived A allele in Northeast Siberians occurs within a genomic region of unusually strong LD. Rather, our findings suggest a recent origin of the c.1436C>T mutation in Northeast Siberians.

Under the assumption that the c.1436C>T mutation was the target of positive (directional) selection, we estimated the mode of selection, selection strength (s), and

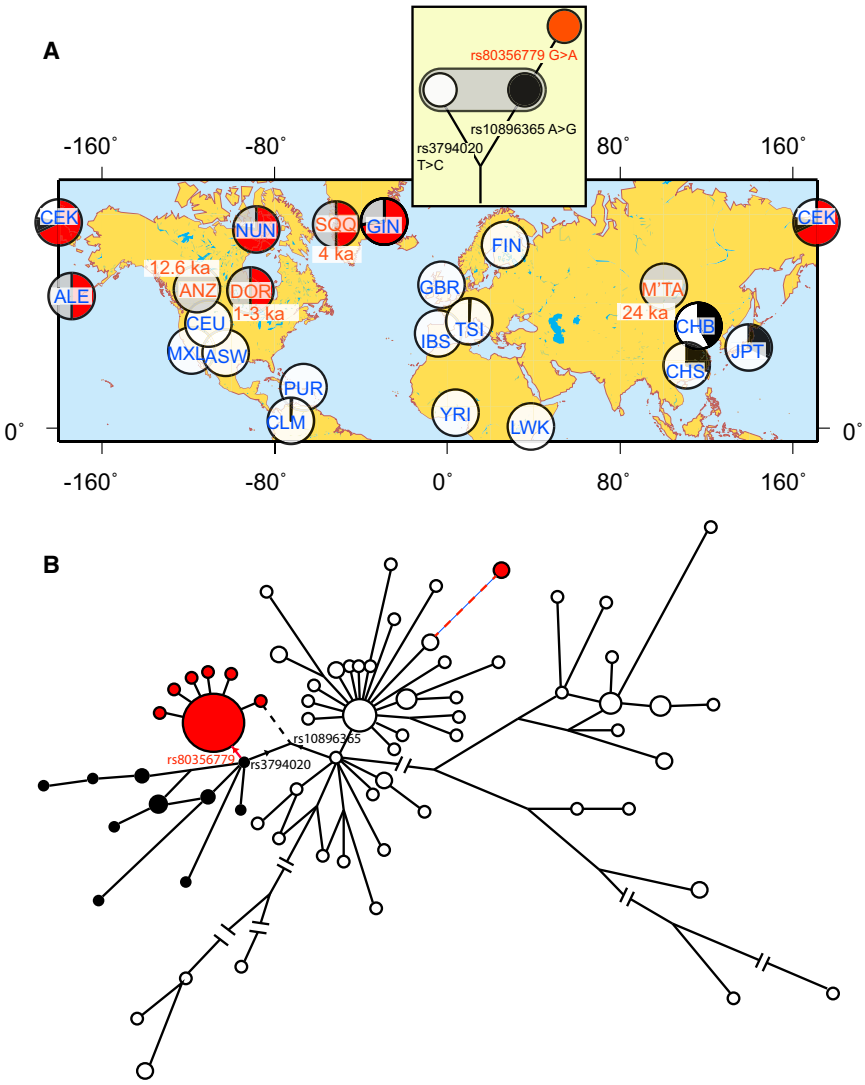

Figure 2. Geographic Distribution and Network of the *CPT1A* c.1436C>T Mutation and Its Associated Haplotype

(A) The c.1436C>T derived allele is defined by the rs80356779 G>A mutation and occurs with a frequency of 0.9, 0.875, and 0.5625 in Chukchi, Eskimo, and Koryaks, respectively, but is absent elsewhere (1000 Genomes Project,[4,5] CG public data,[6] and Personal Genome Project). We used three SNPs (rs10896365 A>G, rs80356779 G>A, and rs3794020 T>C) to define the haplotype GAT (red, see B). The haplotype ancestral to the c.1436C>T mutation (GGT) is shown in black. The white node represents all other haplotypes. Grey shading that encompasses both the white and black nodes refers to subjects for whom information was only available for rs80356779. The map shows the geographic distribution of these haplotypes. Haplotype data were drawn from modern-DNA (blue font) and ancient-DNA (red font) sources, including Chukchi (CEK), Eskimo, and Koryaks (present study); 1000 Genomes Project;[4,5] Nunavut Inuit (NUN);[13] Mal'ta (M'TA);[14] Clovis (ANZ);[15] Aleutian Islander (ALE); Early, Middle, and Late Dorset (DOR) (Table S10);[16] Saqqaq (SQQ);[17] and Greenland Inuit (GIN).[18]

(B) The Haplotype Median Joining Network was constructed from sequences of the Northeast Siberians and control populations (Europeans and East Asians) on the basis of 848 SNPs present in the 58,084 bp region (Table S6) surrounding c.1436C>T with the use of the Network 4.612 package.[19] The circles are proportional to the frequency of the shared haplotype. On the basis of this network analysis, we chose two SNPs (rs10896365 A>G and rs3794020 T>C) that defined the haplotype (black nodes) ancestral to the c.1436C>T mutation (red nodes). For visualization, some branches in the figure are shortened (marked with break). The dashed lines represent likely recombination events.

age of the c.1436C>T allele (time to the most recent common ancestor [T_{MRCA}]) by using an approximate Bayesian computation (ABC) method.[24] We found strong evidence in favor of selection from a de novo mutation P(SDN) = 0.98, as opposed to selection on standing variation. Assuming an additive model of selection, constant population size of N_e = 3,000, and a generation time of 29 years, we estimated the age of c.1436C>T to be 3 ka (highest posterior density [HPD] = 1–23) with a strong selection coefficient of s = 0.14 (HPD = 0.02–0.30) (Figure S7). Selection coefficients of such magnitude have rarely been reported in humans before.[24] To further explore the impact of N_e on our estimates, we repeated the analysis with N_e = 5,000 and N_e = 7,000, yielding qualitatively unchanged results (Table S9). Note, however, that despite the fact that weak selection coefficients can be ruled out by the ABC approach (Figure S7), the method fails to constrain the upper bound of selection strength, limiting the accuracy of the age estimation. It is thus likely that the times are underestimates of the true age of the mutation. Under the assumption of a Poisson model and a star-like genealogy, a maximum-likelihood (ML) estimate for the T_{MRCA} for c.1436C>T resulted in slightly older age estimates, but they were within the HPD interval of the ABC method. Based on high and low mutation rates,[25] these T_{MRCA} estimates were 6.7 ka (confidence interval [CI] = 3–13 ka) and 13.3 ka (CI = 6–26 ka), respectively, for a 58,084 bp region surrounding c.1436C>T (Table S10). These independent estimates of T_{MRCA} further support strong selection with the use of the diffusion approximation to the fixation time of a selective sweep[26] and are consistent with the presence of the allele in the 4-ka-old Saqqaq genome and its absence from older genomes. Overall, when ancient-DNA evidence is combined with the ABC- and ML-based estimates and a slower mutation rate is used, a T_{MRCA} of 6–23 ka ago seems most plausible.

Our results provide evidence that the c.1436C>T mutation was the likely target of selection and drove the sweep signal over the surrounding genomic region. Together with the variant's high frequency in the coastal populations in Northeast Siberia, North America, and Greenland[18,27] (Figure 2A), this finding suggests that the mutation might have historically conveyed a selective advantage to populations in these regions. With agriculture being unsustainable in this part of the world as a result of its extremely cold environment, these coastal populations mostly fed on marine mammals for a high-fat diet rich in n-3 polyenoic fatty acids.[13,28] Such a diet would have led the populations to be in a permanent state of ketosis,[13,29] where metabolism is mainly "lipocentric" (ketone bodies, fatty acids) rather than "glucocentric" (glucose), as found in a high-carbohydrate diet.[30] A lipocentric metabolism provides an efficient means of maintaining energy, which is similar to the state experienced during starvation.[13]

CPT1A imports long-chain fatty acids into mitochondria for use in fatty-acid oxidation. This helps to maintain energy homeostasis and normoglycemia when carbohydrate intake is low.[27] The extent to which the c.1436C>T mutation contributes to disorders associated with CPT1 deficiency, such as hypoketotic hypoglycemia and sudden infant death syndrome, is still unclear. The derived allele has been reported as being deleterious in both the homozygous and the heterozygous state. Yet, its phenotypic effect might depend upon many environmental factors, e.g., feeding history, infection, and climate.[13,27] It is known that the mutation decreases fatty-acid oxidation and ketogenesis, explaining its role in hypoketotic hypoglycemia.[13,28] However, there is also evidence that the mutation decreases the inhibitory effect of malonyl-CoA on fatty-acid β-oxidation in mitochondria, thereby partially compensating for the drop in ketogenesis associated with reduced CPT1A activity.[13,28] A study on Alaskan Yup'iks also suggests that the c.1436C>T mutation might exert a cardioprotective role through its association with elevated levels of high-density lipoprotein cholesterol and reduced adiposity.[28] Moreover, the large amounts of n-3 polyenoic fatty acids in the traditional diet of these aboriginal peoples are known to increase the activity of CPT1A.[13,28] In this context, the CPT1A-activity decrease due to the c.1436C>T mutation could be protective against overproduction of ketone bodies.[13] These important metabolic effects of CPT1A provide the basis of our hypothesis that the c.1436C>T mutation might have conferred a metabolic advantage for the Northeast Siberian populations in dealing with their traditional high-fat diet. The deleterious effect of the mutation might be explained by a change from the traditional diet to a more carbohydrate-based one or by recent cultural shifts and environmental stressors such as fasting and pathogens.

In conclusion, *CPT1A* c.1436C>T joins the short list of known human variants where ill health in present-day populations is a likely consequence of the same variant's being selectively advantageous in the past. Compared with the sickle cell allele rs334 (associated with sickle cell disease [MIM 603903] and malaria resistance [MIM 611162])[31] or rs73885319, rs60910145, and rs71785313 in apolipoprotein L-I (*APOL1* [MIM 603743], associated with kidney disease [FSGS4 (MIM 612551)] and sleeping-sickness resistance[32]), the c.1436C>T allele shares the property of altering a protein sequence. However, unlike the sickle cell allele, it does not represent an example of heterozygous advantage but instead provides an advantageous or disadvantageous effect dependent on the environment. In this way, it extends the range of selective forces contributing to current ill health beyond infectious diseases. It illustrates the medical relevance of an evolutionary understanding of our past and suggests that evolutionary impacts on health might be more prevalent than currently appreciated.

Supplemental Data

Supplemental Data include eight figures and ten tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2014.09.016.

Acknowledgments

This research was supported by European Research Council Starting Investigator grant FP7-261213 to T.K. C.T.-S., Y.X., Q.A., and M.S. were supported by Wellcome Trust grant 098051, and T.A. was supported by Wellcome Trust grant WT100066MA. M. Metspalu and R.V. received supported from the European Union European Regional Development Fund Centre of Excellence in Genomics to the Estonian Biocentre. T.K, M. Metspalu, and R.V. were supported by Estonian Institutional Research grant IUT24-1, and M. Metspalu received Estonian Science Foundation grant 8973.

Received: August 1, 2014
Accepted: September 29, 2014
Published: October 23, 2014

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://browser.1000genomes.org
Complete Genomics, http://www.completegenomics.com/
ERC: European Research Council, http://erc.europa.eu/
Estonian Biocentre, http://www.ebc.ee/free_data
European Nucleotide Archive, http://www.ebi.ac.uk/ena
iHS software, http://hgdp.uchicago.edu/Software/
Ingenuity Variant Analysis, http://www.ingenuity.com/products/variant-analysis
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
Personal Genome Project, https://my.pgp-hms.org/public_genetic_data

Accession Numbers

The raw read data on the whole-genome sequences presented in the current study have been deposited in the European Nucleotide

Archive under accession number PRJEB7258. The data are also available at the data repository of the Estonian Biocentre.

References

1. Leonard, W.R., Snodgrass, J.J., and Sorensen, M.V. (2005). Metabolic adaptations in indigenous Siberian populations. Annu. Rev. Anthropol. *34*, 451–471.
2. Snodgrass, J.J., Leonard, W.R., Sorensen, M.V., Tarskaia, L.A., and Mosher, M.J. (2008). The influence of basal metabolic rate on blood pressure among indigenous Siberians. Am. J. Phys. Anthropol. *137*, 145–155.
3. Cardona, A., Pagani, L., Antao, T., Lawson, D.J., Eichstaedt, C.A., Yngvadottir, B., Shwe, M.T.T., Wee, J., Romero, I.G., Raj, S., et al. (2014). Genome-wide analysis of cold adaptation in indigenous Siberian populations. PLoS ONE *9*, e98076.
4. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A.; 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.
5. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
6. Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., et al. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science *327*, 78–81.
7. Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biol. *4*, e72.
8. Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585–595.
9. Kim, Y., and Stephan, W. (2000). Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics *155*, 1415–1427.
10. Fay, J.C., and Wu, C.I. (2000). Hitchhiking under positive Darwinian selection. Genetics *155*, 1405–1413.
11. McVean, G. (2007). The structure of linkage disequilibrium around a selective sweep. Genetics *175*, 1395–1406.
12. Stephan, W., Song, Y.S., and Langley, C.H. (2006). The hitchhiking effect on linkage disequilibrium between linked neutral loci. Genetics *172*, 2647–2663.
13. Greenberg, C.R., Dilling, L.A., Thompson, G.R., Seargeant, L.E., Haworth, J.C., Phillips, S., Chan, A., Vallance, H.D., Waters, P.J., Sinclair, G., et al. (2009). The paradox of the carnitine palmitoyltransferase type Ia P479L variant in Canadian Aboriginal populations. Mol. Genet. Metab. *96*, 201–207.
14. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.
15. Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., DeGiorgio, M., Stafford, T.W., Jr., Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S.M., et al. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature *506*, 225–229.
16. Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliussen, T.S., Grønnow, B., Appelt, M., Gulløv, H.C., Friesen, T.M., et al. (2014). The genetic prehistory of the New World Arctic. Science *345*, 1255832.
17. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature *463*, 757–762.
18. Rajakumar, C., Ban, M.R., Cao, H., Young, T.K., Bjerregaard, P., and Hegele, R.A. (2009). Carnitine palmitoyltransferase IA polymorphism P479L is common in Greenland Inuit and is associated with elevated plasma apolipoprotein A-I. J. Lipid Res. *50*, 1223–1228.
19. Bandelt, H.J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. *16*, 37–48.
20. Khrameeva, E.E., Bozek, K., He, L., Yan, Z., Jiang, X., Wei, Y., Tang, K., Gelfand, M.S., Prufer, K., Kelso, J., et al. (2014). Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. Nat Commun *5*, 3584.
21. Watterson, G.A. (1975). On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. *7*, 256–276.
22. McEvoy, B.P., Powell, J.E., Goddard, M.E., and Visscher, P.M. (2011). Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. Genome Res. *21*, 821–829.
23. Schiffels, S., and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nat. Genet. *46*, 919–925.
24. Peter, B.M., Huerta-Sanchez, E., and Nielsen, R. (2012). Distinguishing between selective sweeps from standing variation and from a de novo mutation. PLoS Genet. *8*, e1003011.
25. Scally, A., and Durbin, R. (2012). Revising the human mutation rate: implications for understanding human evolution. Nat. Rev. Genet. *13*, 745–753.
26. Ewens, W.J. (2004). Mathematical Population Genetics: Theoretical Introduction (Berlin: Springer-Verlag).
27. Collins, S.A., Sinclair, G., McIntosh, S., Bamforth, F., Thompson, R., Sobol, I., Osborne, G., Corriveau, A., Santos, M., Hanley, B., et al. (2010). Carnitine palmitoyltransferase 1A (CPT1A) P479L prevalence in live newborns in Yukon, Northwest Territories, and Nunavut. Mol. Genet. Metab. *101*, 200–204.
28. Lemas, D.J., Wiener, H.W., O'Brien, D.M., Hopkins, S., Stanhope, K.L., Havel, P.J., Allison, D.B., Fernandez, J.R., Tiwari, H.K., and Boyer, B.B. (2012). Genetic polymorphisms in carnitine palmitoyltransferase 1A gene are associated with variation in body composition and fasting lipid traits in Yup'ik Eskimos. J. Lipid Res. *53*, 175–184.
29. Phinney, S.D. (2004). Ketogenic diets and physical performance. Nutr. Metab. (Lond) *1*, 2.
30. Westman, E.C., Feinman, R.D., Mavropoulos, J.C., Vernon, M.C., Volek, J.S., Wortman, J.A., Yancy, W.S., and Phinney, S.D. (2007). Low-carbohydrate nutrition and metabolism. Am. J. Clin. Nutr. *86*, 276–284.
31. Allison, A.C. (1954). Protection afforded by sickle-cell trait against subtertian malareal infection. BMJ *1*, 290–294.
32. Genovese, G., Friedman, D.J., Ross, M.D., Lecordier, L., Uzureau, P., Freedman, B.I., Bowden, D.W., Langefeld, C.D., Oleksyk, T.K., Uscinski Knob, A.L., et al. (2010). Association of trypanolytic ApoL1 variants with kidney disease in African Americans. Science *329*, 841–845.