Greenlandic Inuit show genetic signatures of diet and climate adaptation

Matteo Fumagalli,1,2,3 Ida Mollke,3,4 Niels Grarup,4 Fernando Racimo,2 Peter Bjerregaard,5,6 Marit E. Jørgensen,5,7 Thorfinn S. Korneliussen,8 Pascale Gerbault,1,9 Line Skotte,3 Allan Linneberg,10,11,12 Cramer Christensen,13 Peter Bjerregaard,5,6 Marit E. Jørgensen,5,7 Thorfinn S. Korneliussen,8

The indigenous people of Greenland, the Inuit, have lived for a long time in the extreme conditions of the Arctic, including low annual temperatures, and with a specialized diet rich in protein and fatty acids, particularly omega-3 polyunsaturated fatty acids (PUFAs). A scan of Inuit genomes for signatures of adaptation revealed signals at several loci, with the strongest signal located in a cluster of fatty acid desaturases that determine PUFA levels. The selected alleles are associated with multiple metabolic and anthropometric phenotypes and have large effect sizes for weight and height, with the effect on height replicated in Europeans. By analyzing membrane lipids, we found that the selected alleles modulate fatty acid composition, which may affect the regulation of growth hormones. Thus, the Inuit have genetic and physiological adaptations to a diet rich in PUFAs.

To detect signals of positive selection, we used the population branch statistic (PBS) (D), which identifies alleles that have experienced strong changes in frequency in one population (GI) relative to two reference populations (CEU and CHB). (5). A sliding window analysis identified several SNP windows with high PBS values, indicative of selection (Fig. 1 and table S1). The strongest signal of selection is located within a region on chromosome 11 (Fig. 1A) and encompasses five genes: two open reading frames, C1orf91 (TEMEM258) and C1orf9 (MYRF); and three fatty acid desaturases, FADS1, FADS2, and FADS3. The SNP with the highest PBS value falls within FADS2. The function of FADS3 is not known; FADS1 and FADS2 encode delta-5 and delta-6 desaturases, which are the rate-limiting steps in the conversion of linoleic acid (omega-6) and α-linolenic acid (omega-3) to the longer, more unsaturated and biologically active eicosapentaenoic acid (EPA, omega-3), docosahexaenoic acid (DHA, omega-3), and arachidonic acid (omega-6).

1Department of Genetics, Evolution, and Environment, University College London, London WCIE 6BT, UK. 2Department of Integrative Biology, University of California—Berkeley, Berkeley, CA 94720, USA. 3The Bioinformatics Centre, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark. 4The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, 2100 Copenhagen, Denmark. 5National Institute of Public Health, University of Southern Denmark, 5353 Copenhagen, Denmark. 6Greenland Center for Health Research, University of Greenland, Nuuk, Greenland. 7Steno Diabetes Center, 2820 Gentofte, Denmark. 8Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark. 9Department of Anthropology, University College London, London WC1E 6BT, UK. 10Research Centre for Prevention and Health, Capital Region of Denmark, Copenhagen, Denmark. 11Department of Clinical Experimental Research, Rigshospitalet, Glostrup, Denmark. 12Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 13Department of Medicine, Lillebaelt Hospital, Veje, Denmark. 14Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark. 15Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 16Faculty of Medicine, University of Aalborg, Aalborg, Denmark. 17School of Natural Sciences, University of California—Merced, Merced, CA 95343, USA. 18Department of Cardiology, Aalborg University Hospital, 9100 Aalborg, Denmark. 19Department of Statistics, University of California—Berkeley, Berkeley, CA 94720, USA. *These authors contributed equally to this work. †Corresponding author. E-mail: tomas.hansen@sund.ku.dk (T.H.); albrechts@bINF.ku.dk (A.A.); rasmus.nielsen@berkeley.edu (R.N.)
Polymorphisms in FADS1 and FADS2 are associated with increased levels of plasma and erythrocyte delta-5 desaturases in Alaskan Inuit (12) as well as with levels of PUFA in blood and breast milk (13, 14).

We also found signals of selection in a region on chromosome 1 (Fig. 1A), which encompasses WARS2, a mitochondrial tryptophanyl-tRNA synthetase, and TBX15, a transcription factor member of the T-box family. Within this region, the SNP with the highest PBS value is located upstream of WARS2. Polymorphisms in or near WARS2 and TBX15 have been shown to be associated with numerous phenotypes among individuals of European descent, including waist-hip ratio (15). Based on linkage disequilibrium (LD) patterns in Greenlandic Inuit, the results from (15) suggest that the allele that occurs frequently in Greenlandic Inuit may decrease the waist-hip ratio. TBX15 plays a role in the differentiation of brown (subcutaneous) and brite (typically inguinal) adipocytes (16). The latter, upon stimulation by exposure to cold, can differentiate into cells capable of expressing UCP1 (uncoupling protein 1), which produces heat by lipid oxidation. Therefore, TBX15 may be associated with adaptation to cold in Inuit.

FN3KRP shows evidence of selection as well (Fig. 1A). FN3KRP encodes an enzyme that catalyzes fructosamines, psicosamines, and ribulosamines. This protein protects against nonenzymatic glycation, an oxidative process that is associated with various pathophysiological processes, including diabetes, cardiovascular disease, aging, and cancer (17). A high intake of PUFAs is associated with increased oxidative stress (18); it is possible that the alleles affected by selection in FN3KRP counteract the negative fitness caused by a PUFA-rich diet. A list of additional candidate regions under positive selection is presented in tables S2 and S3.

To corroborate our results from the SNP chip-based analysis described above, we also calculated PBS values (table S4) for exome sequencing data from 18 unrelated GI individuals (3), combined with data from 85 CEU individuals and 97 CHB individuals from the 1000 Genomes Project (fig. S1) (19).

These analyses identified two high-scoring genes (table S5): DSP, a gene associated with cardiomyopathy (20), and ANGPTL6, a gene that counteracts high-fat diet–induced obesity and related insulin resistance through increased energy expenditure (21). Gene ontology enrichment analyses of genes under selection revealed enriched muscle and heart-development categories, similar to those positively selected in polar bears (tables S6) (5, 22).

In addition, these analyses reproduced the strong signal observed in the FADS1-FADS2-FADS3 region, even though the SNPs with the highest PBS values are not detected by the system used for exome capture (Agilent SureSelect; fig. S2), and this region has the SNP with the strongest signal of selection (i.e., highest PBS value) in any of the data analyzed. We therefore focused on this region for the rest of this study. On the basis of an inferred demographic model (5), we estimated a divergence time between CHB and GI of 23,250 years before the present (yr B.P.), unidirectional gene flow from GI to CHB at some point in the history of these populations, and a reduced effective population size of GI (effective population size = 1550). The estimated model (fig. S3A) fits the observed joint site frequency spectrum (fig. S4), and the PBS value for the FADS region is a strong outlier, corroborating the idea that selection probably has affected this region (fig. S5).

Using an approximate Bayesian computation approach, we also estimated the starting time and intensity of selection, s (5). Because of the high LD within the region and the fact that our data were from SNP chip (fig. S6), we could not pinpoint the causative SNP(s) by means of population genetic analyses; we therefore used the SNP with the highest PBS value (reference SNP identification number rs7477917) as a proxy. This SNP has a derived allele frequency of 0.98 in GI, 0.025 in CEU, and 0.16 in CHB. Our analyses produced maximum a posteriori probability (MAP) estimates of the selection starting time, 19,751 yr B.P. [95% Bayesian credible interval (BCI): 2499 to 22,771 yr B.P.] (figs. S3B and S7), and of s, 3.13% (95% BCI: 0.98 to 19.49%) (fig. S3C). These results suggest that selection began to act on these genes long before the earliest settlement of Inuit in Greenland (4). In population samples from the HGDPCEPH (Human Genome Diversity Project–Centre d’Etude du Polymorphisme Humain) database, the selected allele of rs7477917 has much higher frequencies among Native Americans than it does among East Asians (fig. S8) (23), suggesting that selection began to act before the Inuit split from the Native Americans, when their common ancestors lived in or around Beringia (24).

Six SNPs in the FADS region (Table 1) have PBS values above 2, suggesting that they have been subjected to strong selection. One of these SNPs, rs174570, is associated with circulating high-density lipoprotein, low-density lipoprotein (LDL), and total cholesterol levels in Europeans (25). We therefore tested for associations between the top six SNPs and 13 metabolic and anthropometric phenotypes in Greenlanders by analyzing data from the Greenlandic cohorts IHT (Inuit Health in Transition) and B99 (Greenland Population Study 1999), which include 2733 and 1331 genotyped individuals, respectively (5). We analyzed the cohorts separately, combined the results in a meta-analysis (5), and found marginally significant associations with multiple phenotypes, including body-mass index, fasting serum insulin, and fasting serum LDL cholesterol (tables S7 to S12). In all cases, the derived (selected)
allele was associated with a reduction in the phenotypic value. The strongest association was with body weight ($P = 1.1 \times 10^{-7}$; rs7115739) and height ($P = 0.00012$; rs7115739) (table S10). Both of these associations remained significant after Bonferroni correction for testing for association between 13 phenotypes and six SNPs. To further validate the association with height, we genotyped an additional Greenlandic cohort, known as BBH, consisting of 541 Greenlandic individuals who live in Denmark and for whom height information is available. When we added these data to the meta-analysis of height, the association signal for rs7115739 became even stronger ($P = 4.6 \times 10^{-7}$). Moreover, the per-allele effect size estimates for the derived allele for height and weight are $-0.66$ cm and $-2.2$ kg in IHIT and $-1.2$ cm and $-2.4$ kg in B99 (Fig. 2, A and B, and table S10). As mentioned, the statistical method that we used accounts for admixture. Furthermore, we observed an effect both in Greenlanders with little or no European ancestry and in Greenlanders with more than 40% European ancestry when we stratified the data on the basis of ancestry proportions, which we would not expect if the association signal was caused by admixture in our data (fig. S9). These observations indicate that our association results are not caused by insufficient correction for admixture.

The six SNPs with the highest PBS values are also polymorphic in Europeans (Table 1). However,

Table 1. Annotation for the top six SNPs under positive selection in Greenlandic Inuit. DAFs for each population (CEU, CHB, and GI) and PBS values are reported, along with the genomic position for each SNP.

<table>
<thead>
<tr>
<th>Position*</th>
<th>Reference SNP identification number</th>
<th>Alleles†</th>
<th>DAF</th>
<th>PBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>chr14:21627960</td>
<td>rs74771917 C/T</td>
<td>0.025</td>
<td>0.16</td>
<td>0.98</td>
</tr>
<tr>
<td>chr14:61631510</td>
<td>rs3168072 A/G</td>
<td>0.017</td>
<td>0.18</td>
<td>0.98</td>
</tr>
<tr>
<td>chr14:61632310</td>
<td>rs12577276 A/G</td>
<td>0.017</td>
<td>0.18</td>
<td>0.98</td>
</tr>
<tr>
<td>chr14:61641717</td>
<td>rs7115739 G/T</td>
<td>0.017</td>
<td>0.22</td>
<td>0.98</td>
</tr>
<tr>
<td>chr14:61624414</td>
<td>rs174602 C/T</td>
<td>0.80</td>
<td>0.73</td>
<td>0.01</td>
</tr>
<tr>
<td>chr14:61597212</td>
<td>rs174570 C/T</td>
<td>0.16</td>
<td>0.34</td>
<td>0.99</td>
</tr>
</tbody>
</table>

*Positions refer to human genome assembly hg19. †Alleles are coded as ancestral/derived states.
because most of the identified SNPs have low allele frequencies in Europeans, they may have been missed by GWAS studies. When combining seven European cohorts, including GIANT (Genetic Investigation of Anthropometric Traits; 26), we found associations with lower height in carriers of the derived T-allele for rs7115739 (n = 207,300; \( P = 0.0007 \pm 1 \)) and rs174570 (n = 263,451; \( P = 1.24 \times 10^{-6} \)) (Fig. 2, C and D, and table S13). The meta-analysis–based effect sizes are equivalent to −0.35 and −0.12 cm for rs7115739 and rs174570, respectively. In contrast, we found no evidence that the six SNPs are associated with weight in Europeans. These results are consistent with results that we obtained when we explicitly tested for differences in effect sizes between Europeans and Greenlandic Inuit (table S14): We found no evidence of a difference in effect size for height for rs7115739 (\( P = 0.44 \)), but we found significant evidence for a difference in effect size for weight (\( P = 0.025 \) and \( P = 0.012 \)) for rs7115739 and rs174570, respectively, with little or no effect on weight in Europeans. The associations with height in Europeans are unexpected, because this locus was not found to be significant genome-wide in the recent GIANT study of the height of more than 170,000 Europeans (26). In addition to the associations with height, we also found known associations with low fasting serum levels of insulin, total cholesterol, and LDL cholesterol for European carriers of low-frequency–derived alleles of FADS2 variation, suggesting that there may be a protective effect of these variants on cardiometabolic phenotypes (table S13).

To further elucidate the possible functional effects of the alleles of rs7115739 and rs174570, we investigated associations with red blood cell–membrane lipid composition, which reflects fatty-acid intake from the preceding 2 to 4 months and which has previously been measured in IHIT, the largest of our Greenlandic cohorts (27). We found significant associations with multiple different fatty acids (fig. S10 and tables S15 and S16). Particularly, we found that the selected alleles are significantly associated with an increase in the concentration of eicosatetraenoic acid (ETA, 20:4n-3) and other omega-3 fatty acids upstream in the omega-3 synthesis pathway, before conversion to EPA (20:5n-3), but a decrease in the concentration of both EPA and omega-3 docosapentaenoic acid (DPA, 22:5n-3), with no significant effect on DHA (22:6n-3) (Fig. 3). These results are consistent with previous observations of linked alleles in Europeans (28). The conversion of ETA to EPA is catalyzed by delta-5 desaturases encoded by FADS1 and EPA is a major dietary omega-3 fatty acid in the traditional Inuit diet (18). Hence, these results suggest that selection affecting the fatty acid desaturases may have compensated for a high dietary intake of EPA.

The changes in the concentration of omega-6 fatty acids mirror those of omega-3 fatty acids (Fig. 3). This might be expected, given that the same enzymes (encoded by FADS1 and FADS2) are involved in both the omega-3 and omega-6 biosynthesis pathways. The similar changes in concentration could therefore be a side effect of selection, driven by a omega-3 PUFAs–rich diet. However, selection may also have worked directly on omega-6 fatty acid concentrations early in the ancestral history of Inuit and Native Americans, in the context of a late Paleolithic diet rich in meat from land mammals.

Both rs7115739 and rs174570 show strongly significant associations in conditional analyses where we adjusted for the effects of the other SNP and of rs714602. The remaining three highest–PBS SNPs are in strong LD with rs7115739 in IHIT and would produce similar results. This suggests that there are either multiple causative SNPs or that both rs7115739 and rs174570 are in strong LD with the causal SNPs.

The challenging environmental conditions of the Arctic have probably imposed strong selective pressures on the Inuit and their ancestors. In all the data that we analyzed, the most pronounced allele–frequency difference between Inuit and other populations was found in a cluster of fatty acid desaturases—FADS1, FADS2, and FADS3—although it is possible that even more extreme differences are present in noncoding regions not covered by our exome data. The FADS region has probably been under selection, driven by a diet high in PUFAs. The FADS genes have previously been hypothesized to be under selection in other populations in response to dietary changes (28, 29), suggesting that these genes in general play an important role in human adaptation to dietary regimes. Our results also show that genetic variants in fatty acid desaturases have a strong effect on height, probably because of the effect of fatty acid composition and concentration on the regulation of growth hormones (30). Previous studies (31) have shown that fish oil supplementation is associated with increased concentrations of plasma insulin-like growth factor–1. This study illustrates the utility of evolutionary studies of locally adapted populations for understanding the genetic basis of phenotypic variation among humans.

REFERENCES AND NOTES

4. H. C. Gulle, Granlands Forhistorie (Gyldendal, Copenhagen, 2004).
5. Materials and methods are available as supplementary materials on Science Online.
CRYO-EM shows the polymerase structures and a nonspoiled genome within a dsRNA virus

Hongrong Liu*† and Lingpeng Cheng*‡

Double-stranded RNA (dsRNA) viruses possess a segmented dsRNA genome and a number of RNA-dependent RNA polymerases (RdRps) enclosed in a capsid. Until now, the precise structures of genomes and RdRps within the capsids have been unknown. Here we report the structures of RdRps and associated RNAs within nontranscribing and transcribing cypoviruses (NCPV and TCPV, respectively), using a combination of cryo–electron microscopy (cryo-EM) and a symmetry-mismatch reconstruction method. The RdRps and associated RNAs appear to exhibit a pseudo-D₃ symmetric organization in both NCPV and TCPV. However, the molecular interactions between RdRps and the genomic RNA were found to differ in these states. Our work provides insight into the mechanisms of the replication and transcription in dsRNA viruses and paves a way for structural determination of lower-symmetry complexes enclosed in higher-symmetry structures.

The family Reoviridae causes disease in humans, livestock, insects, and plants. The virions have 10 to 12 segments of dsRNA enclosed in a single-, double-, or triple-layered capsid. The inner capsids (cores) remain intact after the viruses are delivered into the host cell's cytoplasm, and the RNA-dependent RNA polymerases (RdRps) repeatedly transcribe RNA from the minus-strand RNA genome within the core (1, 2). Assembly of the reovirus cores requires encapsidation of the genomic RNA plus strands, along with a roughly equal number of RdRps. The maturation of the reoviruses is accompanied by RdRps-driven synthesis of RNA minus strands complementary to the plus strands, in turn forming genomic double-stranded RNA (dsRNA) segments within the mature virions (3, 4).

Although structures of viral capsids and isolated RdRp complexes have been studied extensively for more than two decades (3–13), the structures of genomes and RdRps within viral capsids have thus far evaded determination. In this study, we used cryo–electron microscopy (cryo-EM), in combination with our symmetry-mismatch reconstruction method, to report the structures of RdRps and associated RNAs for nontranscribing and transcribing cypoviruses (NCPV and TCPV, respectively) in the family Reoviridae.

Cypovirus particles were isolated and purified, and viral transcription was assayed (14, 15). We reconstructed the structures of the NCPV and TCPV without imposing any symmetry (see supplementary materials and methods). Our analysis of the NCPV showed that the genomic RNAs and RdRps are located inside the capsid within a region of 50 Å radius. The structure of the genomic RNAs is of spherical outline; is composed of regularly distributed layers that are formed by discontinuous dsRNA fragments running in parallel, and is associated with RdRps (Fig. 1A to C; fig. S1; movie S1). Each RdRp is anchored at the inner surface of the capsid and surrounded by multiple layers of dsRNA (Fig. 1B and C). The distance between two adjacent dsRNA fragments within the same layer is fixed at ~25 Å, whereas two adjacent layers are ~30 Å apart. The double helices of both dsRNA fragments located close to the inner capsid surface and interacting with the RdRps have a measured helix pitch of ~28 Å (Fig. 1C). The dsRNA fragment structures located closer to the spherical center are not as well resolved as the ones at the periphery (fig. S1). Each RdRp density anchors to the inner surface of the capsid, slightly off-center from the fivefold axis (Fig. 1B) (16). These RdRps and the associated dsRNA fragments appear to exhibit a pseudo-D₃ symmetric organization (Fig. 1A and figs. S1 to S3), allowing for 12 distinct locations of RdRps inside a viral capsid: Two groups containing three RdRps (threefold RdRps) each approach and are symmetrically arranged about the threefold axes on opposite sides of the virion, and three groups containing two RdRps (twofold RdRps) each approach and are symmetrically arranged about the twofold axes that encircle the center of the virion (fig. S4). Within the three-dimensional density maps, the average density value of the twofold RdRps amounts to approximately two-thirds of the average density value of the threefold RdRps. In contrast, the dsRNA densities surrounding the twofold and the threefold RdRps are all of similar intensity. We reason that this reflects six RdRps occupying the six positions of the threefold RdRps and only four RdRps occupying the six positions of the twofold RdRPs (thus, two-thirds of the average density). Therefore, the total number of RdRps within the capsid is 10, in tentative agreement with the observation that each cypovirus genome contains only 10 RNA segments, with each genome segment being specifically associated with one RdRp (17). Our structural analysis also revealed that TCPV and NCPV have almost identical genome structures (figs. S2 and S5), except for those genome regions that interact with RdRps. Given the great variations of size and the encoded genes of the 10 different genomic RNA segments in each cypovirus, it is likely that the observed D₃ symmetry in the dsRNA organization does not reflect the true organization of the RNA genome. The layers of the dsRNA fragment resemble the organization of the cholesteric liquid crystal (18) (fig. S1 and movie S2), which is consistent with earlier evidence that the dsRNA genome forms liquid crystalline arrays within the highly condensed capsid (5). The liquid crystalline model of genome...
Greenlandic Inuit show genetic signatures of diet and climate adaptation
Matteo Fumagalli et al.
Science 349, 1343 (2015); DOI: 10.1126/science.aab2319