



Published in final edited form as:

*Curr Biol.* 2021 May 24; 31(10): 2214–2219.e4. doi:10.1016/j.cub.2021.02.041.

## The genetic history of Greenlandic-European contact

Ryan K. Waples<sup>1,10</sup>, Aviaja L. Hauptmann<sup>2,10</sup>, Inge Seiding<sup>3</sup>, Emil Jørsboe<sup>1,4</sup>, Marit E. Jørgensen<sup>5,6,7</sup>, Niels Grarup<sup>7</sup>, Mette K. Andersen<sup>7</sup>, Christina V.L. Larsen<sup>5,6</sup>, Peter Bjerregaard<sup>5</sup>, Garrett Hellenthal<sup>8</sup>, Torben Hansen<sup>4,9</sup>, Anders Albrechtsen<sup>1,\*</sup>, Ida Moltke<sup>1,11,\*</sup>

<sup>1</sup>Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, Copenhagen, Denmark <sup>2</sup>Ilisimatusarfik - The University of Greenland, Nuuk, Greenland <sup>3</sup>Nunatta Katersugaasivia Allagaateqarfialu - Greenland National Museum and Archives, Nuuk, Greenland <sup>4</sup>Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark <sup>5</sup>National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark <sup>6</sup>Greenland Centre for Health Research, University of Greenland, Nuuk, Greenland <sup>7</sup>Steno Diabetes Center Copenhagen, Gentofte, Denmark <sup>8</sup>UCL Genetics Institute (UGI), Department of Genetics, Evolution and Environment, UCL, London, UK <sup>9</sup>Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark <sup>10</sup>These authors contributed equally <sup>11</sup>Lead contact

### SUMMARY

The Inuit ancestors of the Greenlandic people arrived in Greenland close to 1,000 years ago.<sup>1</sup> Since then, Europeans from many different countries have been present in Greenland. Consequently, the present-day Greenlandic population has ~25% of its genetic ancestry from Europe.<sup>2</sup> In this study, we investigated to what extent different European countries have contributed to this genetic ancestry. We combined dense SNP chip data from 3,972 Greenlanders and 8,275 Europeans from 14 countries and inferred the ancestry contribution from each of these 14 countries using haplotype-based methods. Due to the rapid increase in population size in Greenland over the past ~100 years, we hypothesized that earlier European interactions, such as pre-colonial Dutch whalers and early German and Danish-Norwegian missionaries, as well as the later Danish colonists and post-colonial immigrants, all contributed European genetic ancestry. However, we found that the European ancestry is almost entirely Danish and that a substantial fraction is from admixture that took place within the last few generations.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

\*Correspondence [albrecht@binf.ku.dk](mailto:albrecht@binf.ku.dk) (A.A.), [ida@binf.ku.dk](mailto:ida@binf.ku.dk) (I.M.).

#### AUTHOR CONTRIBUTIONS

A.A. and I.M. conceived and supervised the study. R.K.W., A.L.H., A.A., and I.M. finalized the specific research questions. R.K.W., A.A., and I.M. designed the statistical analyses with input from G.H. and E.J. R.K.W., E.J., A.A., and I.M. created and quality controlled the genetic datasets. R.K.W. performed the statistical analyses with input from I.M., A.A., and G.H. A.L.H. and I.S. provided the historical and contemporary context for the study. M.E.J., C.V.L.L., and P.B. collected and provided context for the Greenlandic samples. T.H., N.G., M.K.A., A.A., and I.M. designed the overall genetic study of the Greenlanders. R.K.W., A.L.H., I.S., A.A., and I.M. wrote the majority of the manuscript with input from all other co-authors.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

#### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cub.2021.02.041>.

## In brief

The Greenlandic Inuit have had extensive historical contact with Europeans, and the present-day Greenlandic population has substantial amounts of European ancestry. Waples et al. use genetic data to investigate the origin of this ancestry. They show that much of it is Danish and find little evidence of it being from pre-colonial European contact.

---

## RESULTS

### Background

The Greenlanders are mainly descendants of the Inuit of the Thule culture<sup>3</sup> that entered Northern Greenland from Canada around the 12<sup>th</sup> century.<sup>1,4</sup> At that time, the Norse had lived in the southern part of the island since 985 CE; they stayed in Greenland until approximately 1450 CE. Previous genetic research has not provided support for gene flow between the Norse and Inuit.<sup>2</sup> However, since the 16<sup>th</sup> century, thousands of Europeans from various countries either visited or moved to Greenland, and there has been substantial gene flow from Europe into the Greenlandic population.<sup>2,5-7</sup> This European contact with Greenland can be divided into three time periods: pre-colonial; colonial; and post-colonial (Figure 1). Pre-colonial contact was initially limited to exploration and trade, such as when a search for the Northwest Passage led English explorers to Greenland in the 1500s.<sup>8</sup> From the early 18<sup>th</sup> century, European whaling efforts off the west coast of Greenland brought whalers into contact with the Greenlandic Inuit. Initially, it was the Dutch who dominated European whaling, but in the latter half of the century, the whalers were also German, primarily Frisian, British, and Danish-Norwegian.<sup>8,9</sup> In 1721, the arrival of the Danish-Norwegian missionary priest Hans Egede marked the beginning of the colonial period, leading to a new and more permanent type of contact between Greenlandic Inuit and Europeans, although whaling was still a primary draw, with, e.g., 107 Dutch ships that year. In addition to Danish-Norwegian missionaries, the German Moravian Brethren established religious missions in the period 1733–1900, located in Nuuk and several other locations.<sup>8</sup> In 1751, Denmark-Norway expanded colonial activities and claimed a monopoly on trade,<sup>8</sup> and since then, the primary contact between Greenlanders and Europeans has been with the Kingdom of Denmark. Denmark-Norway remained a conglomerate state until 1814,<sup>10</sup> after which Greenland became an exclusively Danish colony, and in 1953, it became an equal part of the Kingdom of Denmark. The post-colonial period has seen a significant influx of mainly Danish workers but also seasonal fishers, primarily from Portugal and the Faroe Islands.<sup>8,11</sup>

Although there was a systematic documentation of marriages between Inuit and Scandinavians from the 1740s,<sup>8,10,12</sup> the degree of admixture prior to the colonial period is largely unknown. However, the extensive whaling and trading activities of the Dutch have led to a common belief in Greenland that admixture with Dutch whalers was relatively common.<sup>13</sup> And notably, any European contribution to the Greenlandic gene pool prior to the 20<sup>th</sup> century could have an outsized impact compared to more recent admixture, because the population of Greenland has recently greatly expanded, from less than 6,000 in 1789 to more than 55,000 today.<sup>14</sup> Thus, prior to performing this study, we hypothesized that especially Denmark but also the Netherlands, Germany, and Norway all made non-negligible

ancestry contributions. Here, we investigated this hypothesis by analyzing dense SNP array data from 3,972 Greenlanders from 15 different locations (Figure S1).

### Inference of European admixture sources

Using the genetic data from the Greenlanders, we inferred admixture proportions (Figure S2; see also Data S1) and identified related individuals. Based on the results, we obtained a set of unrelated admixed Greenlanders with both Inuit and European ancestry ( $n = 1,582$ ) and a set of 181 unrelated, unadmixed Greenlanders with only Inuit ancestry. Genetic data from these individuals were then combined with SNP array data from 8,275 individuals from 14 different European countries, including Denmark, Norway, Sweden, Germany, the UK, and the Netherlands (Figure S1). After quality control and merging, we were left with a combined dataset with 135,702 SNPs, 1,582 admixed Greenlanders, and 8,456 reference individuals. Next, we applied the program ChromoPainter<sup>15</sup> to this dataset to reconstruct (“paint”) the genomes of the admixed Greenlandic individuals as mixtures of the haplotypes in the reference individuals (Table S1). The main outcome of this analysis was an estimate for each admixed Greenlander (and each reference individual) of over how much of their genome they are most closely related ancestrally to each of the reference individuals. These estimates were summarized in a so-called coancestry matrix (Figure S3).

We then estimated the genetic contribution from the unadmixed Greenlandic Inuit and each of the 14 different European countries to the ancestry of the admixed individuals in Greenland by applying the program SOURCEFIND<sup>16</sup> to a summary of the output from ChromoPainter. SOURCEFIND is a Markov chain Monte Carlo (MCMC) method that produces statistical samples from a posterior distribution of the ancestry contributions for a group of target individuals. In this case, the target individuals were the admixed Greenlanders and the possible ancestry sources were the reference countries. These statistical samples can then be summarized in different ways that are informative about the ancestral contribution of each of the reference countries. We applied this method to each admixed Greenlander separately to obtain individual-level resolution but also tried to analyze all the admixed Greenlanders jointly in one group in an attempt to reduce noise in our ancestry estimates.

When performing inference on each admixed individual separately, we first summarized the SOURCEFIND results using the posterior mean of the ancestry proportion from each country per individual to get a simple overview (Figure 2). Based on this, we estimated the 1,582 admixed Greenlanders to have an average of 65.6% Greenlandic Inuit ancestry and 34.4% European ancestry, with the far majority of the latter being Danish. To obtain a more detailed picture of the SOURCEFIND results, we also counted the number of individuals that were assigned at least 5% ancestry from any country with high probability (posterior probability  $> 0.99$ ). When doing so, we found 1,100 of the 1,582 admixed Greenlandic individuals (69.5%) were assigned at least 5% Danish ancestry, the most of any European country (Table 1). The only other European countries found to contribute more than 5% ancestry to five or more individuals are all Nordic countries: Norway with 98 (6.2%); Sweden with 20 (1.3%); and Finland with 5 (0.3%). A few other countries were inferred to contribute more than 5% ancestry for 1–5 individuals: UK (0.2%); Ireland (0.2%); Poland

(0.1%); Germany (0.1%); and the Netherlands (0.1%). The same overall pattern is observed with a lower 1% ancestry threshold (Table 1), with different posterior probability thresholds or a different prior (Table S2). Only four individuals with more than 20% European ancestry were inferred to have ancestry from countries other than Denmark or Norway (Table 1).

We obtained similar results in the group-based analysis, with an estimated Inuit ancestry fraction of 65.6% and European fraction of 34.4% (95% credible interval = 33.5%–36.0%) (Table 1). Please note these results pertain to the admixed Greenlandic individuals and do not reflect the Greenlandic population as a whole, which is estimated to have approximately 25% European ancestry.<sup>2</sup> The Danish ancestry fraction among admixed Greenlanders was 31%, with no other European country contributing more than 1%. This translates to Denmark making up 91% of the total estimated European ancestry, with the only other country contributing more than 1% of the European ancestry being Norway at 2.1% (Table 1). Notably, we performed a range of additional analyses to ensure the validity of these results (Data S2).

### Investigating European admixture in the last few generations

To further characterize the history of European admixture in Greenland, we performed an analysis to investigate the timing of admixture in Greenland. Specifically, we inferred local ancestry, Inuit or European, along the genome of each admixed Greenlander to estimate the proportions of the genome where each admixed individual has (1) inherited both alleles from Inuit ancestors, (2) inherited both alleles from European ancestors, or (3) inherited one allele from an Inuit ancestor and one allele from a European ancestor. These fractions are informative about the time of admixture because individuals with different admixture histories have different expected ternary fractions (see Figure 3A for some examples). We chose to estimate “ternary ancestry fractions” instead of using standard methods for timing of admixture based on admixture tract lengths (e.g., Pool and Nielsen<sup>17</sup> and Gravel<sup>18</sup>), because the number of phasing switch errors was large relative to the recombination rate since admixture, which we feared would markedly affect the timing estimates. In contrast, the ternary ancestry fractions are robust to phasing switch errors.

Among the 1,582 admixed Greenlanders, 250 have ternary fractions that are consistent with having at least one fully European ancestry parent (Figure 3B, blue and yellow dots). Of these, 27 have two European alleles at nearly every genomic position (yellow dots on Figure 3B), suggesting they have two European parents. Together, these 277 ( $223 + 2 \times 27$ ) European ancestry parents account for >8% of the ancestors of the admixed individuals ( $277/[2 \times 1,582]$ ) and for almost 25% of the total European ancestry in Greenland. The ternary ancestry fractions of the remaining individuals are largely consistent with second and third generation admixture with Europeans, as shown in Figure 3A (dots near the right axis). However, it is important to emphasize that, due to variance in recombination and nonrandom mating, these fractions could also be the result of older admixture.

Among the group of admixed Greenlanders with at least one European-ancestry parent, Denmark was by far the largest European ancestry source, making up 98.4% of the European ancestry, with no other country contributing more than 1% (Table S3). In contrast, the group of Greenlanders without a European parent, i.e., a group for which the admixture

must have taken place less recently than for the group with at least one European parent, was inferred to have contributions from Norway (3.8%), Germany (2.1%), and Sweden (1.6%), with Denmark constituting 85.7%.

## DISCUSSION

Before discussing the results in a historical perspective, we should consider how the study design imposes limitations in the conclusions that can be drawn from the analysis. Briefly, we do not believe that the current study imposes major limitations that pertain to the conclusions presented here; for further details and discussion on this topic, please see Data S2.

The genetic analyses suggest that the European ancestry of the present-day Greenlanders is predominantly Danish and the result of very recent gene flow. This indicates that European activities prior to colonization did not have a significant impact on the current genetic composition of the population in Greenland, in contrast both to common beliefs in Greenland<sup>13</sup> and our own initial hypothesis.

The lack of genetic ancestry originating from the early exploration activities by the British is perhaps the least surprising, because these activities only involved a few ships. Similarly, the relatively small amount of ancestry originating from German Moravian missionaries, who stayed in Greenland for about 170 years until 1900, may be explained by the restrictions that the Moravian Brethren put on intermarriage with the Greenlandic population.<sup>19</sup>

However, the lack of ancestry from whaling countries, especially the Netherlands, is surprising given the common beliefs in Greenland as well as the historical records suggesting a high number of Dutch ships around Greenland's coasts for a substantial period of time.<sup>8,9</sup> This result may be explained by a number of factors. First, early European whalers often did not spend the winter in Greenland.<sup>13</sup> Second, Dutch, English, and other European whaling activities were reduced by the economic monopoly imposed in 1751 by Denmark-Norway. Finally, it has been postulated that first contact with Europeans was followed by severe epidemics and that the interaction with the Dutch around Disko Island led to some of the first incidents of epidemics in the region.<sup>13</sup> A well-documented example was a severe smallpox epidemic in Nuuk in the 1730s following the arrival of European ships.<sup>13</sup> It is possible that these epidemics could have impacted early patterns of European ancestry and reduced the impact of early admixture.

Our results suggest that most of the European ancestry is from after colonization was initiated. This result is consistent with the fact that most of the Greenlandic individuals without any European ancestry live in the very north as well as the east coast of Greenland,<sup>2</sup> because colonial activities were initiated later in the north (1909) and east (1894) than in the southwest (1721). Also, we found a higher fraction of Norwegian, Swedish, and German ancestry among the Greenlanders without at least one European-ancestry parent, which aligns well with the family registries from the colonial period. The large amount of inferred Danish ancestry, especially within the last generation, is consistent with historical records

showing that the influx of Danes to Greenland in the post-colonial period since the 1950s marked a substantial increase in the European immigration rate.

Hence, taken together, although at first perhaps surprising, the results of this study seem consistent with recent demographic trends in Greenland and with historical records of European contact.

## STAR★METHODS

### RESOURCE AVAILABILITY

**Lead contact**—Further information and requests should be directed to and will be fulfilled by the Lead Contact, Ida Moltke (ida@binf.ku.dk).

**Materials availability**—This study did not generate new unique reagents.

**Data and code availability**—The accession number for the Greenlandic genotype data from the Multi-ethnic genotyping array (MEGA) reported in this paper is EGA: EGAS00001004933. The Greenlandic genotype data on the MetaboChip are available from the European Genome-phenome Archive (<https://ega-archive.org>) under the accession EGA: EGAS00001002641. The European reference datasets are also available at the European Genome-phenome Archive with accessions EGAD00000000120, EGAD00010000124, EGAD00010000288, and EGAD00010000632. The 1000 Genomes data are publicly available from the 1000 Genomes Project (<https://www.internationalgenome.org/data>).

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study participants were Greenlandic individuals from two population surveys: the Inuit Health in Transition (IHIT,  $n = 3115$ ) and a survey consisting of Greenlanders living in Greenland (B99,  $n = 1401$ ), and Greenlanders living in Denmark (BBH,  $n = 547$ ).<sup>30,31</sup> The cohorts have participants from 15 different locations in Greenland from Qaanaaq in the northwest to Tasiilaq in the southeast, (Figure S1) as well as Greenlanders living in Denmark.

The participants gave oral and written consent to participate in the health surveys and subsequently they were mailed information about the population genetics analyses with an option to opt out at any time. The approval for population genetics analyses was given by the Commission for Scientific Research in Greenland (project 2014–08, 2014–098017).

To represent potential European source countries, we selected individuals from 14 different European countries ( $n = 14,385$ ): The UK, Sweden, Germany, Norway, Italy, Finland, Belgium, the Netherlands (Dutch), France, Ireland, Denmark, Spain, Northern Ireland, and Poland.

### METHOD DETAILS

**Greenlandic genotype data**—All Greenlandic participants were genotyped on two SNP arrays: the CardioMetaboChip (196,224 SNPs)<sup>2,22,32</sup> and the Multi-Ethnic Global Array (~1.5M SNPs).<sup>33</sup> Data from these two SNP arrays were merged on the plus strand and 3972

individuals with genotypes from both SNP arrays and a missing rate below 0.02 were retained. From these we removed singletons, sites not on an autosome, as well as sites with a significant ( $p < 1e-10$ ) deviation from Hardy-Weinberg equilibrium in a test that accounts for admixture.<sup>29</sup>

**European genotype data**—The European SNP array data are from the Wellcome Trust Case Control Consortium (EGAD00000000120, EGAD00010000124, EGAD00010000288, EGAD00010000632),<sup>20,21</sup> and were selected to represent a broad spectrum of potential European admixture sources in Greenland (Figure S1). The European datasets were lifted to hg19 and put on the plus strand, and sites with rates of missing data  $> 0.05$  were removed prior to merging. We also excluded sites within the MHC region and within the HsInv0501 inversion on chromosome 8, as well as sites with more than two alleles. Finally, we limited the number of individuals from each European country to at most 1000 and confirmed that there were no related individuals within each European country.

**1000 Genomes Data**—For the ADMIXTURE analyses and local ancestry analyses with RFmix (see below) we selected the Han Chinese in Beijing (CHB), Yoruba in Ibadan (YRI), and Utah residents with Northern and Western European Ancestry (CEU) population samples from the Thousand Genomes Project (1000G),<sup>23</sup> for a total of 310 individuals. We used the phased genotypes from phase 3 aligned to GRCh37.

**Merged Greenlandic and European reference data**—For the haplotype-based analyses we worked on a dataset where the Greenlandic data and the European reference samples were merged. We kept all sites present in both datasets and excluded 52 sites with more than 2% missing data. The resulting merged dataset had 135,702 loci and 12,247 individuals with a total genotyping rate of 0.9995 and all loci with a minor allele count of at least 5.

The merged Greenlandic-European dataset was split by chromosome and phased without a reference panel using SHAPEIT<sup>34</sup> (v2.r904) with default settings, using the HapMap phase II recombination map for hg19.

After merging and phasing, we removed close relatives among all Greenlandic individuals by retaining at most one individual from each pair of individuals with a coefficient of relatedness  $> 0.2$ . Then we split the remaining Greenlanders into two sets based the results of a  $K = 2$  ADMIXTURE: 1) the un-admixed Greenlanders with  $>99\%$  inferred Inuit ancestry, and 2) the admixed Greenlanders with  $> \%$  inferred European ancestry, for additional details see Data S1. From the second set, we removed seventeen Greenlandic individuals estimated to have  $>5\%$  African or  $>7\%$  Asian ancestry in a  $K = 4$  ADMIXTURE analyses including 1000 genomes samples from China (CHB), Nigeria (YRI), the US (CEU). These thresholds were selected to exclude individuals that differed markedly from the majority of other Greenlandic individuals (data not shown) and to be able to avoid having to include any Asian and African reference samples in our fine-scale analyses. We also excluded admixed Greenlandic individuals living in Denmark as these individuals may be more likely to have Danish ancestry than other European ancestries. This left us with a dataset consisting of 1582 not closely related Greenlanders with European admixture

(admixed samples), 181 not closely related unadmixed Greenlanders (Inuit reference samples), and 8303 European reference samples.

Based on the results of a pilot ChromoPainter analysis, we subsequently excluded 28 of the European reference samples because they were significant outliers ( $z$ -score  $> 5$ ), based on comparing their total chunk counts to the rest of the individuals from their country (not shown). An atypically high number of chunks can be indicative of low data quality. This resulted in a final set of 8275 European reference samples (Figure S1) and thus  $8275 + 181 = 8456$  reference samples in total and 1582 not closely related Greenlanders with European admixture. These data were used to infer ancestry contributions, for details of this analysis see the Quantification and statistical analysis section and Data S2.

**Merged Greenlandic and 1000G data**—To construct a dataset for the ADMIXTURE and local ancestry analyses, we merged the Greenlandic genotype data with data from 310 individuals from three 1000G populations: Han Chinese in Beijing (CHB), Yoruba in Ibadan (YRI), and Utah residents with Northern and Western European Ancestry (CEU). We subsequently removed 46 sites with a greater than 0.25 frequency difference in the CEU individuals compared to the European admixture component in the  $K = 2$  analysis (see below), retaining 521,622 overlapping sites.

## QUANTIFICATION AND STATISTICAL ANALYSIS

**ADMIXTURE analyses**—We performed two different ADMIXTURE<sup>25</sup> analyses to facilitate the generation of input data for our main analyses: 1) an unsupervised  $K = 2$  ADMIXTURE analysis of all 3972 Greenlandic individuals assuming an Inuit and a European ancestry component, following a previous study<sup>2</sup> and 2) a supervised  $K = 4$  ADMIXTURE analysis of the Greenlandic individuals combined with individuals of European, Asian and African descent to investigate if there are any ancestry from Asian and African populations. We used the  $K = 2$  analysis (Figure S2) to create two sets of Greenlandic individuals: “unadmixed” and “admixed” with  $> 99\%$  or  $< 99\%$  Inuit ancestry respectively. These unadmixed individuals were used as a reference for the Greenlandic Inuit ancestry component, while the admixed individuals were the subject of the main analyses; for a more detailed discussion, see Data S1.

Before the unsupervised  $K = 2$  analysis, we applied a minor allele frequency (MAF) filter of 0.05 to the Greenlandic dataset described above ( $n = 3972$ ), resulting in a dataset with 538,514 sites. For the supervised  $K = 4$  admixture analysis, we selected the Han Chinese in Beijing (CHB), Yoruba in Ibadan (YRI), and Utah residents with Northern and Western European Ancestry (CEU) populations as proxies for Asian, African, and European ancestry, respectively.

For each analysis, we ran each ADMIXTURE (v1.3.0) ten times and selected the run with the maximum likelihood, checking convergence by ensuring multiple other runs within two log-likelihood units.

**Relatedness estimation**—To estimate relatedness coefficients for the Greenlandic individuals we used relateAdmix.<sup>28</sup> This method accounts for admixture by estimating



individual allele frequencies when estimating pairwise identity by descent (IBD) coefficients ( $k_1, k_2$ ) based on genome-wide ancestry proportions for each individual. We used the  $K = 2$  genotype data and ADMIXTURE estimates of these genome-wide ancestry proportions. To estimate relatedness for the Europeans we applied the IBD inference function (`-genome`) in PLINK2<sup>26,35</sup> to the genotype data from all the Europeans.

**Chromosome painting**—We characterized the coancestry between Greenlanders and Europeans with the haplotype-based method ChromoPainter.<sup>15</sup> This method is based on a Hidden Markov model (HMM) that statistically reconstructs (“paints”) a target haplotype as a mixture of a set of reference haplotypes while exploiting linkage disequilibrium among nearby SNPs. We combined the reference individuals from Greenland ( $n = 181$ ) and Europe ( $n = 8275$ ), with the unrelated admixed Greenlanders ( $n = 1582$ ) to construct the dataset for this analysis ( $n = 10038$ ). First, we painted each reference individual using all other reference individuals, then, we painted each admixed Greenlander using all reference individuals. We specified constant mismatch ( $\mu = 2.04 \times 10e-5$ ) and switch rate ( $N_e = 103.35$ ) parameters across all analyses, which we estimated as the weighted mean values using data from chromosomes 1, 4, 15, and 22 in a subset of 168 individuals chosen to represent all reference populations, using 10 iterations of the expectation-maximization (EM) algorithm implemented in ChromoPainter. For all these analyses, we used the same recombination map as during haplotype phasing.

ChromoPainter quantifies coancestry using two different measures, one based on the length of the genome copied from each donor in centiMorgans (cM), deemed “chunk lengths” by the program, and the second based on simple counting of the number of distinct ancestry chunks copied from each donor, deemed “chunk counts.” Unless otherwise noted, we used the chunk lengths measure in downstream analyses. For summaries of the ChromoPainter analysis, see Table S1 and Figure S3.

**Ancestry contributions from the European reference countries**—We estimated the ancestry contributions from each European reference country and Greenland with SOURCEFIND (v2),<sup>16</sup> based on summaries of the coancestry matrix estimated by ChromoPainter. We summarized the ChromoPainter output into a vector of length 15 for each admixed Greenlander and reference individual, with this vector containing the proportion of DNA by which that person is painted by individuals from each of the 15 reference populations. For each reference population, we averaged these vectors across individuals. We then applied SOURCEFIND to form the vector of each admixed Greenlander as a mixture of those from the reference populations. This is a Markov Chain Monte Carlo (MCMC) approach which puts a prior on the expected number of contributing reference populations and provides an estimate of the genetic contribution of the unadmixed Greenlandic Inuit and each of the 14 different European countries to the ancestry of the admixed individuals in Greenland, while accounting for sample size differences among reference populations. We conducted this analysis in two ways: 1) to each admixed Greenlander individually 2) to the entire set of admixed Greenlanders as a group. The individual-based analysis allowed us to investigate the range of individual-level patterns of European ancestry, while the group analysis considers a large number of individuals at once

and estimates the ancestry sources of the mathematically-average admixed Greenlander. The later was done to reduce the noise from averaging the estimates of the individual-based analysis. For additional analyses related to validating the ancestry inference, please see Data S2.

To ensure convergence was reached in the SOURCEFIND analyses, we ran 5 MCMC chains for each analysis and compared variance within and between separate chains with the Rhat diagnostic.<sup>36</sup> Each chain was run with 1M iterations, a 100K burn-in and a thinning factor of 1000. We tested that we discarded enough to burn-in by computing Rhat while discarding the first 500K iterations and compared these values to the shorter burn-in (data not shown). For each ancestry source in each individual the Rhat diagnostic was consistent with MCMC convergence; mean Rhat across all chains was 1.0001, and the max value was 1.0044. For most of our SOURCEFIND analyses we used default priors with eight eligible sources and a mean of four sources expected to contribute. However, we tested if the results were robust to choice of prior by also running additional analyses with a more sparse prior with eight eligible sources and two sources expected to contribute.

**Investigating European admixture in the last few generations**—To investigate the timing of European admixture, we assigned local ancestry, either Inuit or European, in each admixed Greenlandic individual using RFMix (v2).<sup>27</sup> In this analysis, we used the same Inuit reference individuals as in the ChromoPainter analysis, along with CEU individuals from 1000G to represent the European ancestry, this allowed us to utilize the larger number of overlapping loci with the 1000G dataset. RFMix was run with default parameters, except we specified two different admixture dates, either 3 or 8 generations ago, to ensure that our results were robust to this choice. We used genotype data from the merged Greenlandic and 1000 Genomes datasets with 521,622 sites, split by chromosome and phased without a reference panel. After phasing, the reference Inuit and CEU individuals were used as the ancestry references for local ancestry inference in the admixed Greenlanders.

We summarized the results for each individual by calculating the fraction of the genome, in cM, that has either two Inuit alleles, two European alleles, or one Inuit and one European allele. We found a few chromosomal regions, such as near the edge of chromosomes, with local ancestry fractions that were outliers relative to the rest of the genome, suggesting potential problems with the inference of local ancestry in these regions, or local genomic factors affecting ancestry. To address this, we removed 88 out of 26008 (0.3%) genomic windows of local ancestry calls with less than 62.5% Inuit ancestry or with more than 72.5% Inuit ancestry, for a total exclusion of 3.76 cm.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We gratefully acknowledge the participants in the Greenlandic health surveys. We thank Lucy van Dorp for help with ChromoPainter and Pongsakorn Wangkumhang for help with GLOBETROTTER. This project was funded by the Danish Council for Independent Research (DFR-4090-00244 from FNU to I.M.; DFR-4181-00383 to A.A.), the Lundbeck Foundation (R215-2015-4174 to A.A.), and the Novo Nordisk Foundation (NNF15OC0017918 to N.G.;

NNF16OC0019986 to N.G). G.H. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant number 098386/Z/12/Z). The Greenlandic health surveys (IHIT and B99) were supported by Karen Elise Jensen's Foundation, the Department of Health in Greenland, NunaFonden, the Medical Research Council of Denmark, the Medical Research Council of Greenland, and the Commission for Scientific Research in Greenland. The Novo Nordisk Foundation Center for Basic Metabolic Research (<https://cbmr.ku.dk>) is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation, grant number NNF18CC0034900.

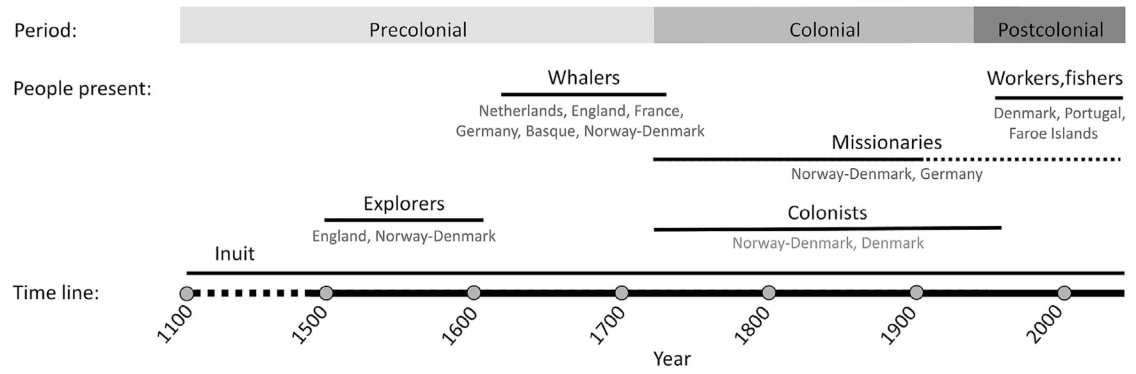
## REFERENCES

- Gulløv HC (2008). The nature of contact between native Greenlanders and Norse. *J. North Atlantic* 1, 16–24.
- Moltke I, Fumagalli M, Korneliussen TS, Crawford JE, Bjerregaard P, Jørgensen ME, Grarup N, Gulløv HC, Linneberg A, Pedersen O, et al. (2015). Uncovering the genetic history of the present-day Greenlandic population. *Am. J. Hum. Genet.* 96, 54–69. [PubMed: 25557782]
- Raghavan M, DeGiorgio M, Albrechtsen A, Moltke I, Skoglund P, Korneliussen TS, Grønnow B, Appelt M, Gulløv HC, Friesen TM, et al. (2014). The genetic prehistory of the New World Arctic. *Science* 345, 1255832. [PubMed: 25170159]
- Friesen TM, and Arnold CD (2008). The timing of the Thule migration: new dates from the Western Canadian Arctic. *Am. Antiq.* 73, 527–538.
- Bosch E, Calafell F, Rosser ZH, Nørby S, Lynnerup N, Hurler ME, and Jobling MA (2003). High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. *Hum. Genet.* 112, 353–363. [PubMed: 12594533]
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, 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. [PubMed: 20148029]
- Pereira V, Tomas C, Sanchez JJ, Syndercombe-Court D, Amorim A, Gusmão L, Prata MJ, and Morling N (2015). The peopling of Greenland: further insights from the analysis of genetic diversity using autosomal and X-chromosomal markers. *Eur. J. Hum. Genet.* 23, 245–251. [PubMed: 24801759]
- Frandsen NH, Gulløv Hans C, Jens H, Jensen EL, Marquardt O, Rud S, Seiding I, Toft PA, and Thuesen S (2017). Grønland – den arktiske koloni. In *Danmark og Kolonierne*, Gulløv HC, ed. (Gads Forlag), pp. 46–107.
- Avango D, Hacquebord L, and Wråkberg U (2014). Industrial extraction of Arctic natural resources since the sixteenth century: technoscience and geo-economics in the history of northern whaling and mining. *J. Hist. Geogr.* 44, 15–30.
- Ostermann H (1940). *Nordmænd på Grønland 1721–1814* (Gyldendal Norsk Forlag).
- Hansen P (1955). De fremmede fiskere i Vestgrønlands farvande. *Tidsskriftet Grønland* 5, 190–200.
- Schultz BJ, and Frandsen NH (2010). *Nordgrønland 1790–96: Inspektør B.J. Schultz' Indberetninger til Direktionen for den Kongelige Grønlandske Handel (Selskabet for Udgivelse af Kilder til Dansk Historie)*.
- Gad F (1969). *Grønlands Historie. 2, 1700–1782: Halvbd. 2* (Nyt Nordisk Forl.).
- Hamilton LC, and Rasmussen RO (2010). Population, sex ratios and development in Greenland. *Arctic* 63, 43–52.
- Lawson DJ, Hellenthal G, Myers S, and Falush D (2012). Inference of population structure using dense haplotype data. *PLoS Genet.* 8, e1002453. [PubMed: 22291602]
- Chacón-Duque J-C, Adhikari K, Fuentes-Guajardo M, Mendoza-Revilla J, Acuña-Alonzo V, Barquera R, Quinto-Sánchez M, Gómez-Valdés J, Everardo Martínez P, Villamil-Ramírez H, et al. (2018). Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. *Nat. Commun.* 9, 5388. [PubMed: 30568240]
- Pool JE, and Nielsen R (2009). Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* 181, 711–719. [PubMed: 19087958]
- Gravel S (2012). Population genetics models of local ancestry. *Genetics* 191, 607–619. [PubMed: 22491189]

19. Wilhjelmsen H (2001). Af Tilbøielighed er jeg Grønlandsk. Om Samuel Kleinschmidts Liv og Værk (Det Grønlandske Selskab).
20. Dubois PCA, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, Zhernakova A, Heap GAR, Adány R, Aromaa A, et al. (2010). Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* 42, 295–302. [PubMed: 20190752]
21. Strange A, Capon F, Spencer CCA, Knight J, Weale ME, Allen MH, Barton A, Band G, Bellenguez C, Bergboer JGM, et al.; Genetic Analysis of Psoriasis Consortium & the Wellcome Trust Case Control Consortium 2 (2010). A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. *Nat. Genet.* 42, 985–990. [PubMed: 20953190]
22. Moltke I, Grarup N, Jørgensen ME, Bjerregaard P, Treebak JT, Fumagalli M, Korneliussen TS, Andersen MA, Nielsen TS, Krarup NT, et al. (2014). A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature* 512, 190–193. [PubMed: 25043022]
23. The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526, 68–74. [PubMed: 26432245]
24. Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, and Myers S (2014). A genetic atlas of human admixture history. *Science* 343, 747–751. [PubMed: 24531965]
25. Alexander DH, Novembre J, and Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. [PubMed: 19648217]
26. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, and Lee JJ (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. [PubMed: 25722852]
27. Maples BK, Gravel S, Kenny EE, and Bustamante CD (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* 93, 278–288. [PubMed: 23910464]
28. Moltke I, and Albrechtsen A (2014). RelateAdmix: a software tool for estimating relatedness between admixed individuals. *Bioinformatics* 30, 1027–1028. [PubMed: 24215025]
29. Meisner J, and Albrechtsen A (2019). Testing for Hardy-Weinberg equilibrium in structured populations using genotype or low-depth next generation sequencing data. *Mol. Ecol. Resour.* 19, 1144–1152. [PubMed: 30977299]
30. Bjerregaard P, Curtis T, Borch-Johnsen K, Mulvad G, Becker U, Andersen S, and Backer V (2003). Inuit health in Greenland: a population survey of life style and disease in Greenland and among Inuit living in Denmark. *Int. J. Circumpolar Health* 62 (Suppl 1), 3–79. [PubMed: 14527126]
31. Bjerregaard P (2011). Inuit Health in Transition: Greenland Survey 2005–2010: Population Sample and Survey Methods (Statens Institut for Folkesundhed).
32. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, Burt NP, Fuchsberger C, Li Y, Erdmann J, et al. (2012). The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* 8, e1002793. [PubMed: 22876189]
33. Bien SA, Wojcik GL, Zubair N, Gignoux CR, Martin AR, Kocarnik JM, Martin LW, Buyske S, Haessler J, Walker RW, et al.; PAGE Study (2016). Strategies for enriching variant coverage in candidate disease loci on a multiethnic genotyping array. *PLoS ONE* 11, e0167758. [PubMed: 27973554]
34. Delaneau O, Zagury J-F, and Marchini J (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* 10, 5–6. [PubMed: 23269371]
35. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, and Sham PC (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. [PubMed: 17701901]
36. Gelman A, and Rubin DB (1992). Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457–472.

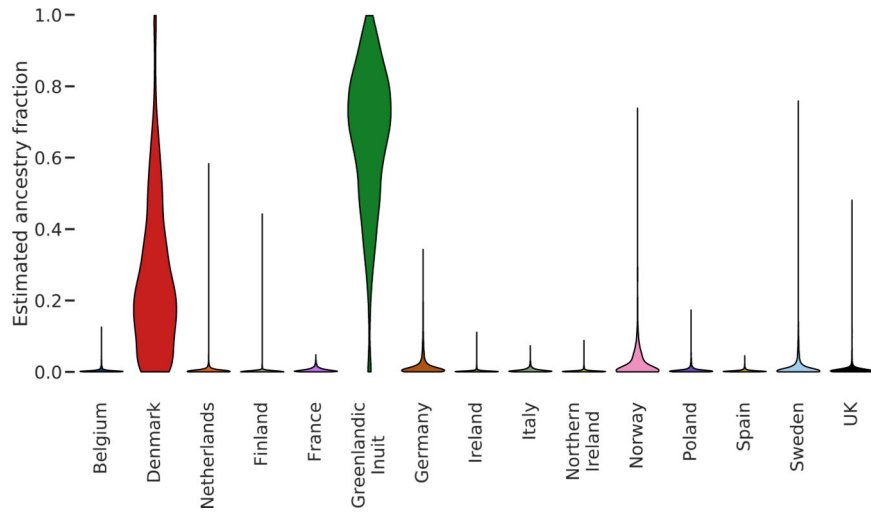
### Highlights

- The present-day Greenlandic population has substantial amounts of European ancestry
- Denmark is the main source of this European ancestry
- There is little evidence of European ancestry from pre-colonial European contact
- The timing of much of the European admixture is very recent



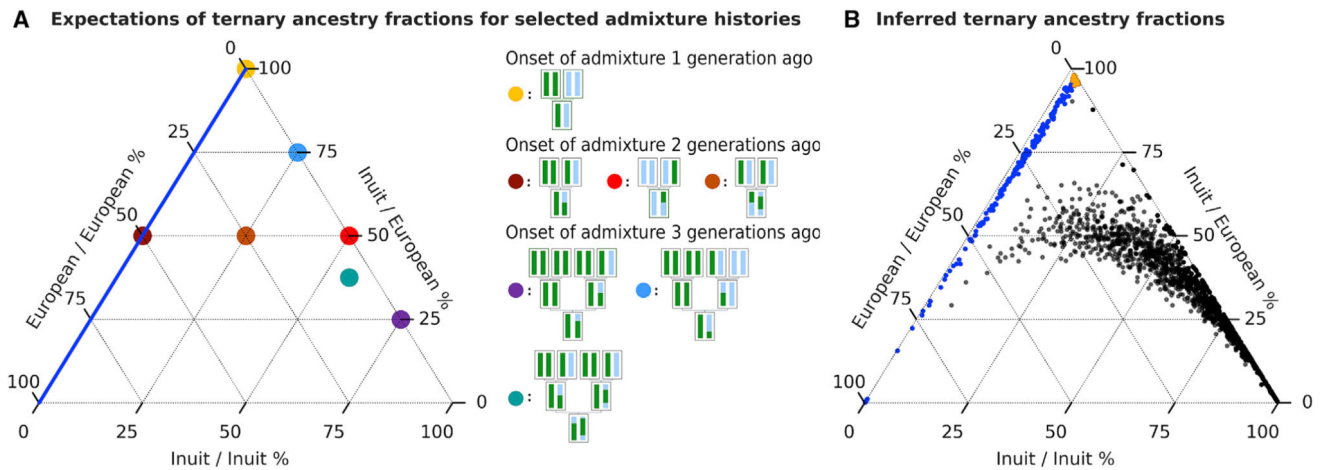
**Figure 1. Timeline of significant Greenlandic-European contact**

The timeline covers the time since the arrival of the Inuit in Greenland. The presence of different groups of people are shown by horizontal bars. Times are approximate. The most relevant European countries are listed under each group.



**Figure 2. Violin plot of per-country ancestry estimates**

Results shown are produced by SOURCEFIND from the analysis where inference was performed on each admixed individual separately. Each source country has a violin showing the distribution of the estimated mean ancestry fraction from that country, across all admixed individuals. Each admixed individual appears in the distribution for each country. Violins are scaled to all have the same max width.



**Figure 3. Ternary ancestry fractions, i.e., the fraction of the genome where (1) both alleles have Inuit ancestry, (2) both alleles have European ancestry, or (3) one allele has Inuit ancestry and one has European ancestry**

The three corners of the plots represent genomes with all loci having two European alleles (bottom left), two Inuit alleles (bottom right), or one Inuit and one European alleles (top).

(A) Expected ternary ancestry fractions. Colored dots show the expected ternary fractions for individuals with 7 selected admixture histories illustrated in the legend by pedigrees, where green indicates Inuit ancestry and light blue indicates European ancestry. The admixture histories include Greenlanders with admixture from one European parent (yellow); one European grandparent (dark brown); two European grandparents, one on each parental side (red); three European grandparents (light brown); one European great-grandparent (blue), two European great-grandparents, one on each parental side (blue-green); and, finally, three European great-grandparents, all on the same parental side (purple). The left axis in blue indicates fractions that are expected for individuals with at least one European parent because it has no sites with two Inuit alleles.

(B) Inferred ternary ancestry fractions. Colored dots show the inferred ternary ancestry fractions for each of the 1,582 admixed Greenlanders. The colors convey the way we have categorized the individuals: individuals inferred to have one Greenlandic parent and one European parent are yellow; the remaining individuals inferred to have a European parent shown in blue; and other individuals are shown in black.

See also Table S3.



**Table 1.**

Inferred ancestry across 1,582 admixed Greenlandic individuals

	No. of reference individuals	<u>Individual-based</u>			<u>Group-based</u>
		1%	5%	20%	
Belgium	537	0.0% (0)	0.0% (0)	0.0% (0)	0.1%
Denmark	327	76.4% (1,208)	69.5% (1,100)	35.8% (567)	31.6%
Finland	580	0.6% (9)	0.3% (5)	0.1% (1)	0.0%
France	478	0.0% (0)	0.0% (0)	0.0% (0)	0.2%
Greenlandic Inuit	181	98.3% (1,555)	98.3% (1,555)	97.4% (1,541)	65.6%
Germany	1,000	0.1% (1)	0.1% (1)	0.0% (0)	0.3%
Ireland	344	0.2% (3)	0.2% (3)	0.0% (0)	0.1%
Italy	745	0.0% (0)	0.0% (0)	0.0% (0)	0.2%
The Netherlands	1,000	0.1% (1)	0.1% (1)	0.1% (1)	0.1%
Northern Ireland	61	0.0% (0)	0.0% (0)	0.0% (0)	0.1%
Norway	942	17.8% (281)	6.2% (98)	1.1% (18)	0.7%
Poland	57	0.1% (2)	0.1% (1)	0.0% (0)	0.2%
Spain	204	0.0% (0)	0.0% (0)	0.0% (0)	0.2%
Sweden	1,000	3.5% (56)	1.3% (20)	0.1% (1)	0.3%
UK	1,000	0.3% (5)	0.2% (3)	0.1% (1)	0.2%

Results from both individual-based and group-based analyses are shown. The first column gives the number of reference individuals from each source. For individual-based, the values shown are the assignment to country at 1%, 5%, and 20% ancestry thresholds. The percentage values are the percentages of admixed Greenlandic individuals inferred to have at least 1%, 5%, or 20% ancestry from each source country. The numbers in parentheses are the number of individuals in each category. To be counted here, an individual must have had at least 1%, 5%, or 20% ancestry with a posterior probability above 99%. For group-based, the values shown are the percentage of the ancestry of the group of 1,582 admixed Greenlanders inferred to come from each country. All the results were inferred using SOURCEFIND. See also Table S2.

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
European genotype data	20,21	EGAD0000000120, EGAD00010000124, EGAD00010000288, EGAS00001002641
Greenlandic genotype data (Metabochip)	22	EGAD00010001427, EGAD00010001428
1000 Genomes data	23	CHB, YRI, CEU
Deposited data		
Greenlandic genotype data (MEGA)	this paper, available at <a href="https://www.ebi.ac.uk/ega/studies/EGAS00001004933">https://www.ebi.ac.uk/ega/studies/EGAS00001004933</a>	EGAS00001004933
Software and algorithms		
CHROMOPAINTER (v2)	15	<a href="https://people.maths.bris.ac.uk/~madj/finestructure-old/chromopainter_info.html">https://people.maths.bris.ac.uk/~madj/finestructure-old/chromopainter_info.html</a>
GLOBETROTTER (v Dec.30.2016)	24	<a href="https://people.maths.bris.ac.uk/~madj/finestructure/globetrotter.html">https://people.maths.bris.ac.uk/~madj/finestructure/globetrotter.html</a>
fineSTRUCTURE (v2, v4)	15	<a href="https://people.maths.bris.ac.uk/~madj/finestructure/finestructure.html">https://people.maths.bris.ac.uk/~madj/finestructure/finestructure.html</a>
SOURCEFIND (v2)	16	Contact Garrett Helleenthal at <a href="mailto:ghelleenthal@gmail.com">ghelleenthal@gmail.com</a>
ADMIXTURE (v1.3.0)	25	<a href="http://dalexander.github.io/admixture/">http://dalexander.github.io/admixture/</a>
PLINK (v1.9)	26	<a href="https://www.cog-genomics.org/plink2">https://www.cog-genomics.org/plink2</a>
RFMix (v2)	27	<a href="https://github.com/slowkoni/rfmix">https://github.com/slowkoni/rfmix</a>
relateAdmix	28	<a href="https://github.com/aalbrechtsen/relateAdmix">https://github.com/aalbrechtsen/relateAdmix</a>
PCAngsd	29	<a href="https://github.com/Rosemeis/pcangsd">https://github.com/Rosemeis/pcangsd</a>