



ARTICLE

Estimating the age of p.(Phe508del) with family studies of geographically distinct European populations and the early spread of cystic fibrosis

Philip Farrell¹ · Claude Férec^{2,3} · Milan Macek⁴ · Thomas Frischer⁵ · Sabine Renner⁵ · Katharina Riss⁵ · David Barton^{6,7} · Teresa Repetto⁸ · Maria Tzetzis⁹ · Karine Giteau³ · Morten Duno¹⁰ · Melissa Rogers⁶ · Hara Levy¹¹ · Mourad Sahbatou¹² · Yann Fichou² · Cédric Le Maréchal^{2,3} · Emmanuelle Génin²

Received: 1 May 2018 / Revised: 10 July 2018 / Accepted: 19 July 2018 / Published online: 8 August 2018
© European Society of Human Genetics 2018

Abstract

The high incidence of cystic fibrosis (CF) is due to the frequency of the c.1521_1523delCTT variant in the cystic fibrosis transmembrane conductance regulator (*CFTR*), but its age and origin are uncertain. This gap limits attempts to shed light on the presumed heterozygote selective advantage that accounts for the variant's high prevalence among Caucasian Europeans and Europe-derived populations. In addition, explaining the nature of heterozygosity to screened individuals with one c.1521_1523delCTT variant is challenging when families raise questions about these issues. To address this gap, we obtained DNA samples from 190 patients bearing c.1521_1523delCTT and their parents residing in geographically distinct European populations plus a Germany-derived population in the USA. We identified microsatellites spanning *CFTR* and reconstructed haplotypes at 10 loci to estimate the time/age of the most recent common ancestor (tMRCA) with the *EstiAge* program. We found that the age estimates differ between northwestern populations, where the mean tMRCA values vary between 4600 and 4725 years, and the southeastern populations where c.1521_1523delCTT seems to have been introduced only about 1000 years ago. The tMRCA values of Central Europeans were intermediate. Thus, our data resolve a controversy by establishing an early Bronze Age origin of the c.1521_1523delCTT allele and demonstrating its likely spread from northwest to southeast during ancient migrations. Moreover, taking the archeological record into account, our results introduce a novel concept by suggesting that Bell Beaker folk were the probable migrating population responsible for the early dissemination of c.1521_1523delCTT in prehistoric Europe.

Introduction

Cystic fibrosis (CF; OMIM 219700), the most common life-threatening autosomal recessive disorder among Caucasians,

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-018-0234-z>) contains supplementary material, which is available to authorized users.

✉ Philip Farrell
pmfarrell@wisc.edu

¹ Pediatrics and Population Health Sciences, University of Wisconsin, Madison, WI, USA

² UMR 1078 Génétique, Génomique fonctionnelle et Biotechnologies, Inserm, Université de Brest, EFS, CHU Brest, Brest, France

³ Laboratoire de Génétique, CHU Brest, Brest, France

⁴ Department of Biology and Medical Genetics, Charles University-2nd Faculty of Medicine and Motol University Hospital, Prague, Czech Republic

⁵ Paediatrics and Adolescent Medicine, Medical University Vienna, Vienna, Austria

⁶ School of Medicine, University College Dublin, Dublin, Ireland

⁷ Department of Clinical Genetics, Our Lady's Children's Hospital, Dublin, Ireland

⁸ Meyer Children Hospital, Cystic Fibrosis Center, Florence University, Florence, Italy

⁹ Department of Medical Genetics, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece

¹⁰ Department of Clinical Genetics, University Hospital, Copenhagen, Copenhagen, Denmark

¹¹ Northwestern University Feinberg School of Medicine and the Ann & Robert H. Lurie Children's Hospital, Chicago, IL, USA

¹² Fondation Jean Dausset – CEPH, Paris, France

is most frequently associated with the first variant discovered in the cystic fibrosis transmembrane conductance regulator gene (*CFTR*, OMIM 602421; reference sequence accession number NM_000492.3), the well characterized c.1521_1523delCTT disease causing variant, known commonly as F508del or p.Phe508del [1]. It accounts for about 70% of CF alleles in Europeans and Europe-derived populations such as Euro-Americans and explains the relatively high birth incidence and prevalence of CF among rare (Mendelian) genetic diseases affecting Caucasians [2]. CF undoubtedly originated in Europe, and there is a decreasing proportion of CF patients with p.(Phe508del) from north-western to southeastern Europe [3, 4]. Although this European geographic gradient is well established [2, 4], the age of the c.1521_1523delCTT variant has remained uncertain and somewhat controversial. Our direct studies of ancient DNA (aDNA) from Iron Age archeological specimens have shown that the c.1521_1523delCTT variant was definitely present at least 2300 years ago; more specifically, this variant was discovered in 3 of 32 individuals buried around 350 Before the Common Era (BCE) near the Danube River in the vicinity of present day Vienna, Austria [5]. Use of indirect strategies for the estimation of the age of this variant, however, have led to published age estimates that range from greater than 2100 generations or about 50,000 years to 3000 years ago [6, 7].

As preconception-, prenatal-, and neonatal screening for CF have proliferated during the past two decades [8, 9], the many thousands of individuals discovered to be heterozygous for the c.1521_1523delCTT allele have often raised questions about the origin and significance of carrying this mutation themselves or in their children identified through DNA-based neonatal screening tests [10]. It has not been possible to address their questions and concerns in this regard. Although a heterozygote selective advantage has been suspected [11, 12], and seems likely [13], efforts to identify it have been unsuccessful, despite many hypotheses such as protection from cholera [14] which was later refuted [15]. A major challenge in such research is the limited historical information that can be connected geographically as has been done convincingly for hemoglobin S carriers, i.e., those with “sickle cell trait” where an evidentiary strategy [16, 17] has convincingly confirmed the “malaria hypothesis” arising from the visionary 1949 report of Pauling et al. [18]. Similarly, a better understanding of *when* and *where* the c.1521_1523delCTT variant arose could aid in understanding *why* it became so frequent. Thus, the initial goal of our project is to gain more knowledge about the age of the c.1521_1523delCTT variant throughout Europe and thus potential insights about its dissemination to provide clues as to a probable heterozygote selective advantage. Consequently, we organized a study of patients and families with this CF-causing allele drawn from eight regions across

Europe comprising representative CF populations and also a Wisconsin, USA cohort with predominantly German ancestry. Recognizing that the principal CF-causing variant may have emerged in eastern Europe at a different time in history than in its western regions, we tested the hypothesis that the age of c.1521_1523delCTT varies among different European populations that are geographically dispersed.

Materials and methods

Specimens

To accomplish an age estimation of the principal CF-causing variant, we obtained blood specimens after informed consent from CF patients with the c.1521_1523delCTT variant and their parents. The blood was anticoagulated with EDTA and placed in plastic tubes labeled only with the family number and member (i.e., patient, mother, or father) but with no personal identifiers. In most of the countries, the DNA was extracted promptly in genetics laboratories and stored prior to analysis. In one (Austria), the blood was processed to prepare leukocyte-enriched samples in 2 ml microtubes after erythrocyte lysis and multiple washings. After being frozen, batched leukocyte specimens were shipped overnight on dry ice from Vienna to the Laboratoire de Génétique, Génomique fonctionnelle et Biotechnologies in Brest where DNA extractions were performed. Approvals from ethics committees were obtained at each institution, although some, such as the University of Wisconsin-Madison, designated the project as research exempt based on the investigators using deidentified samples and not returning results for any clinical practice purpose.

Selection of regional populations

We considered several factors in choosing the regions of Europe to investigate. Taking into account the well-established p.(Phe508del) variant gradient [3, 4] described above, our priority was to select countries that have a high proportion of patients with this allele [2] and a European CF center with an interested, cooperative director or geneticist, as well as knowledge of family ancestry and a willingness to provide specimens from the native population. The request to our collaborators was to select families who knew that their ancestors were native inhabitants in the region. However, it was not possible to carry out detailed genealogical studies on each family. Thus, we relied on family ancestry self-identification and the information drawn from collaborating CF centers. In addition, we selected one population of CF patients and their parents in the USA, or more specifically living in eastern Wisconsin, in which the ancestry was predominantly German. Massive immigration

from Germany to Wisconsin occurred during the second half of the 19th century for economic reasons and by 1890 led to 626,394 German-Americans residing in Milwaukee and the eastern region of the State to account for the majority of the population there.

Population meeting criteria

A total of 190 CF patients and their parents were included in this study. Of this group, 185 were trios with DNA from both parents, while five were pairs with DNA available from only one parent. All the patients were shown to have the *CFTR* c.1521_1523delCTT variant; 166 were homozygous and 24 were compound heterozygotes with one c.1521_1523delCTT and one other CF-causing variant. These individuals were sampled from Albanian, Austrian, Czech, Danish, French, Greek, Irish, and Ukrainian populations as listed in Table 1.

Estimation of tMRCA

All individuals were assessed for the same 10 microsatellite markers that were selected by Fichou et al. [7], around the *CFTR* gene (Table 2). Evaluation of the 10 informative microsatellite DNA sequences amplified by multiplex PCR was performed using Universal Florescent Labeling. Microsatellites were selected by software freely available (zeon.well.ox.ac.uk/git-bin/microsatellite.cgi). Contiguous sequences around the *CFTR* gene (localized at chromosome 7q31) were successively screened for micro-satellite regions. Haplotypes were reconstructed using version 3.3.2 of the Beagle program [19] on both the trios and the pairs with 100 reconstructions. Only the unambiguous haplotypes that were the same over the 100 reconstructions were kept for the analysis. This stringent requirement was established to avoid the inclusion of uncertain haplotypes and that could have biased our estimates. There were 272 such non-ambiguous haplotypes coming from 148

Table 2 The genetic markers in the region surrounding the *CFTR* gene used in this study

Marker	Motif	Genomic position (Human GRCh38)	Distance from <i>CFTR</i> ^a
M01	ca	chr7:112397465–112397507	–5
M02	ca	chr7:114807866–114807902	–2.7
M03	ct	chr7:115424035–115424073	–2.1
M04	ctat	chr7:115831302–115831359	–1.7
M05	ca	chr7:116450632–116450669	–1.1
M09	gt	chr7:117499430–117499462	^b
M10	gt	chr7:118351456–118351505	0.7
M11	ca	chr7:119028265–119028306	1.4
M12	agat	chr7:119623651–119623707	2
M13	ca	chr7:120739799–120739832	3.1

^aDistances from the *CFTR* gene are given in Mb (–centromeric, +telomeric)

^bThis marker is located in the first intron of the *CFTR* gene

independent trios or pairs; 24 haplotypes were from heterozygous p.Phe508del carriers and 248 from homozygote CF patients and their parents.

The age of the most recent common ancestor (tMRCA) of the c.1521_1523delCTT carriers was estimated in each population from the length of the haplotypes shared by the carriers using the *EstiAge* program [20] under a stepwise model at the different markers (assuming a variant rate of 10^{-3} per meiosis). This program provides maximum likelihood estimates of the number of generations since the most recent common ancestor using the multilocus marker data information on patients. It assumes that all affected patients in the sample descended from a common ancestor who introduced c.1521_1523delCTT n generations ago. An estimate of n is obtained from the size of the haplotype shared by individuals on both sides of the disease locus by finding the most likely positions of recombinations on the ancestral haplotype in the different patients; then, the value of n is converted to age in years by multiplying by the assumed 25 years per generation. Chronologic dates were determined using 2017 for the computations. The archeological periods for Europe were assigned by traditional criteria in consultation with archeologists aware of the European regions we selected as sources of DNA [21, 22].

To avoid a possible underestimation of the age associated with ancestral consanguinity, only one of the two haplotypes were considered in homozygous c.1521_1523delCTT carriers. Allele frequencies at the different markers were obtained by considering the parental haplotypes that were not transmitted to the patients in each population separately. Different hypotheses regarding the ancestral haplotype that carries the c.1521_1523delCTT allele were considered. First, it was reconstructed independently in each population

Table 1 Origins and characteristics of the different patient populations

Population	#Patients	#Trios	#Homozygous
Albania	17	16	17
Austria	27	24	27
Czech Republic	17	17	0
Denmark	32	32	32
France	30	30	30
Greece	25	22	18
Ireland	20	20	20
Ukrania	5	5	5
United States ^a	19	19	19

^aPredominately German ancestry

Table 3 Age estimates of the CFTR p.Phe508del in the different populations^a

Population	#Haplotypes	#Generations	#Years	95% CI in years		Ancestral haplotype ^b
				Low	High	
Albania	16	52	1300	850	2050	206_114_199_276_300_253_F508del_221_319_279_240
Austria	21	143	3575	2575	5075	200_120_199_276_302_253_F508del_227_305_287_240
Czech Republic	14	128	3200	2125	4950	208_124_197_280_302_253_F508del_223_305_287_240
Denmark	24	184	4600	3350	6400	200_122_197_276_300_253_F508del_227_305_283_240
France	22	184	4600	3300	6575	210_122_199_276_300_253_F508del_223_305_283_244
Greece	20	47	1175	800	1750	206_114_199_276_300_253_F508del_221_319_279_240
Ireland	10	189	4725	2925	8125	206_108_195_276_300_253_F508del_223_305_279_246
Ukraine	5	86	2150	1050	4625	208_120_199_276_306_253_F508del_227_323_271_240
USA(Germany-derived)	16	185	4625	3150	7025	200_118_199_276_300_253_F508del_227_305_283_240

^aIn each population, the total number of haplotypes included in the analysis (only haplotypes with no ambiguity were kept and one haplotype was randomly chosen for homozygous carriers) and age estimates in number of generations and in years (assuming 1 generation = 25 years) with the lower and upper bound of the 95% confidence interval are listed

^bThe last column gives the imputed ancestral haplotype in each population

by considering the most frequent allele at each position. Second, it was constrained to be the same in every population, considering the most frequent allele in the entire sample.

Results

Haplotypes studies

A total of 148 haplotypes carrying c.1521_1523delCTT were considered for age estimation. Only one haplotype was kept per family. Among these 148 haplotypes, 137 were different, but most (95.9%) of them share a common allele (allele 253) at marker M09 located in intron 1 of the *CFTR* gene (Supplementary Table 1). This allele is the same as the arbitrarily termed “allele 256” in the previous study reported by Fichou et al. [7], for the Breton population. In each population, a different ancestral haplotype was imputed based on the data (Table 3), except in Greece and Albania where the same ancestral haplotype is found.

Time to the most recent ancestor

Table 3 shows tMCRA values estimated with calculation of the 95% confidence intervals for each population group. It was found that the age estimates were quite different between western and eastern populations. In the former, the mean age estimates of the most recent common ancestors vary between 4725 (Ireland) and 4600 years ago (both France and Denmark). It was of special interest to find that the predominantly ancestral German population sampled in the USA was close to the northwestern European

populations at 4625 years ago. In the two populations from the southeast of Europe (Greece and Albania), p.Phe508del must have been introduced much more recently based on our data revealing mean tMRCA values of 1175–1300 years ago with 95% CI ranges that do not overlap with the results obtained in western European populations. Compared with these two groups, the tMRCA values obtained in families residing in the two central European countries, the Czech Republic and Austria, were intermediate at 3200 and 3575 years ago, respectively, but the 95% CIs overlapped. In the Ukrainian group, our estimate is closest to the southeastern countries at 2150 years ago, but the sample size is comparatively small because only five trios met our criteria for inclusion, i.e., were kept for the *Estiage* analysis.

The c.1521_1523delCTT variants found in the various populations investigated were most often carried on the same haplotype, namely microsatellite haplotype 206-114-199-276-300-253- p.Phe508del -227-305-287-240. Assuming there is common ancestral haplotype, as other data also suggest [23], and constraining the analysis to this same haplotype in every population, we repeated the *Estiage* analysis with the most frequent allele. Supplementary Table 2 lists the estimates for each population and reveals the same trends described above but with wider 95% confidence limits.

Conversion of tMRCA values to chronologic archeological periods

Table 4 converts the tMRCA values of Table 3 to calendar years on an absolute chronological scale and also shows the archeological period. Three categories were evident. (1) Ireland = 2708 BCE, derivative predominantly German

Table 4 Year of the most recent common ancestor with 95% CI values and probable archeological era of distribution^a

Population	Year ^b	Archeological period 3	95% CI years	Archeological range
Ireland	2708 BCE	Early Bronze Age	6108–908 BCE	Late Neolithic-Early Iron Age
United States	2608 BCE	Early Bronze Age	5008–1133 BCE	Late Neolithic-Late Bronze Age
France	2583 BCE	Early Bronze Age	4558–1283 BCE	Late Neolithic-Late Bronze Age
Denmark	2583 BCE	Early Bronze Age	4383–1333 BCE	Late Neolithic-Late Bronze Age
Austria	1558 BCE	Middle Bronze Age	3058 BCE–558 CE	Early Bronze Age-Early Middle Ages
Czech Repub	1183 BCE	Late Bronze Age	2933 BCE–108 CE	Early Bronze Age-Roman Imperial
Albania	717 CE	Middle Ages	33 BCE–1167 CE	Roman Imperial-Middle Ages
Greece	842 CE	Middle Ages	267–1217 CE	Roman Imperial-Middle Ages

^aArcheological periods are as traditionally assigned for Europe [21, 22]

^bMean chronologic years and 95% CI values are listed based on the data of Table 3 and using 2017 as the present year

population in the USA (Wisconsin) = 2608 BCE, France = 2583 BCE, Denmark = 2583 BCE [obviously, none of these are statistically different]; (2) Austria = 1558 BCE and the Czech Republic = 1183 BCE [intermediate]; and (3) Albania = 717 AD and Greece = 842 AD [significantly different from the western populations with no overlap of the 95% CI values]. The mean values for the first category represents the early Bronze Age, while the central European populations dated to the Middle and Late Bronze Ages. Lastly, the p.(Phe508del) variant was likely introduced into the southeastern population during and/or after the Roman Imperial Era of 31 BCE to 476 CE.

Data repository

These results are available in a database available to the public that can be accessed through the European Nucleotide Archive (<https://www.ebi.ac.uk/ena>) using study accession number: PRJEB27683 and study unique name: ena-STUDY-INSERM UMR1078-09-07-2018-20:39:26:525-2156.

Discussion

As part of a larger investigation entitled “The Ancient Origin of Cystic Fibrosis,” we designed this project to gain a better understanding of when the p.(Phe508del) variant may have first arisen in Europe, i.e., an estimate of its age, by examining more geographically distributed and distinct trio populations than in previous studies [6, 7] focused on this issue. Although the populations sampled may not be truly pan-European, they represent regions separated by ~2000 miles, i.e., from Ireland to Athens, Greece. In addition, we selected European CF patients and their parents from regions in which geographic and thus genetic distances were previously documented by others [24, 25] using genome wide high density microarray data. In doing so, we

hoped to shed light on where the p.(Phe508del) variant may have arisen and its pattern of dissemination. Addressing these issues, we reasoned, might in turn provide insights about the presumed p.(Phe508del) heterozygote selective advantage [13] or at least help guide future studies as occurred with geographical and historical/temporal evidence [16, 17] confirming the “malaria hypothesis” regarding hemoglobin S heterozygosity. To elucidate the origin and explain the frequency of the p.(Phe508del) CFTR variant, we believe that it would be necessary to answer three “W questions,” namely *when*, *where* and then *why* did p.(Phe508del) allele become so prevalent in northern European populations and their descendants.

Our results revealed tMRCA average values ranging from 4725 to 1175 years ago and support the estimates of Serre et al. (3000–6000 years ago) [11], rather than Morral et al. (52,000 years ago) [6], but the latter figure was challenged by Kaplan et al. [26] because of disagreement with assumptions used in their calculations. In addition, the tMRCA values from western European regions reported herein refine the results of Fichou et al. [7] from a study of Breton CF patients in which the *Estiage* analysis suggested that the most common recent ancestor lived 115 generations ago. That tMRCA value, however, may have underestimated the age of p.(Phe508del) in Brittany due to consideration of all the haplotypes, even those that were reconstructed with ambiguities, as well as a potential bias associated with consanguinity due to including both haplotypes in homozygous families. In the more stringent *Estiage* analyses reported herein, those potential biases were avoided for all populations, leading to estimates of the oldest tMRCA values corresponding to the Early Bronze Age in western Europe, which is generally agreed to begin around 3000 BCE. This finding extends our results from a direct investigation of aDNA in teeth from Iron Age burials near Vienna around 350 BCE and allow us to conclude that p.(Phe508del) was present in that region long before then. More specifically, in the Austrian families studied, the

Estiage data revealed a mean tMCRA value of 3575 years ago, which converts to 1558 BCE (Middle Bronze Age) [22].

Perhaps most remarkably, the estimated ages of p.(Phe508del) in the three western European regions (France, Ireland, and Denmark) were similar with closely overlapping 95% CI values. This observation is also in line with previously documented spatial autocorrelograms expressing genetic and geographical distance for these populations [24]. Such data provide more insight about the ancient origin of CF in our judgment—both *when* and *where*—and lead us to propose that CFTR p.(Phe508del) is derived from ancestors who lived in western Europe during the Bronze Age, as early as 2700 BCE, and that its relatively rapid dissemination occurred because of human migrations around the northwestern Atlantic trading routes [21] and then towards central and eastern Europe [22]. Diffusion from northwestern to central Europe in approximately 1000 years is consistent with the prominent Bronze Age migrations evident in the archeological record [21, 22] and from genomic studies of aDNA [27]. On the other hand, we are assuming a discrete origin of the principal CF-causing variant, but it is possible that p.(Phe508del) arose more than once or earlier, and then reached western Europe subsequently through Neolithic migrations.

Considering potential explanations from archeological evidence regarding prehistoric settlements and migrations, and based on opinions from consulted European archeologists, we believe that the most likely phenomenon of Bronze Age human activities that could account for our CFTR p.(Phe508del) tMCRA observations is the Bell Beaker culture [22, 28–30]. Prehistorians have concluded that Bell Beaker folk appeared at the transition from the Late Neolithic period to the Early Bronze Age during the third millennium BCE somewhere in the western Europe [22], although the exact region is uncertain [29]. They were distinguished by their ceramic beakers, pioneering metallurgy north of the Alps, and great mobility [30, 31]. Over ~1000 years, a network of small families and/or elite tribes spread their culture from west to east throughout western Europe and into regions that correspond closely to the present-day European Union, where the highest incidence of CF is found [32]. More specifically, their distinctive Bell Beaker pottery appeared and spread across western and central Europe beginning around 3000–2750 BCE and then disappeared between 2200 and 1800 BCE [22, 29]. Their migrations are linked to the advent of western and central European metallurgy, as they manufactured and traded metal goods, especially weapons, while traveling over long distances [30]. Most relevant to our study is the evidence that they migrated in a direction and over a time period that fits well with the pattern of tMCRA data we found for the p.(Phe508del) variant. Olalde et al. [29] have

shown that both migration and cultural transmission played a major role in diffusion of the “Beaker Complex” and led to a “profound demographic transformation” of Britain after 2400 BCE. Moreover, the cultural elements that unite the widely distributed Beaker folk are so obvious that some have considered them a distinct ethnicity of Bronze Age people [33].

From our results, we propose the novel concept that large scale, long term west-to-east migrations of the Bell Beaker Europeans [22, 28–30] during the Bronze Age, could explain the dissemination of p.(Phe508del) in Europe and its documented northwest-to-southeast gradient [4]. In fact, our tMCRA data show a temporal gradient also. Determining *when* the p.(Phe508del) variant was first introduced in Europe and discovering *where* it arose should provide new insights about the high prevalence of p.(Phe508del) heterozygotes. For instance, Bronze Age Europeans migrated extensively and apparently were not exposed to endemic infectious diseases or epidemics; thus, microbial-related selection as in sickle hemoglobin seems unlikely [34]. As more information on Bronze Age people and their practices during migrations [21, 22] become available through archeological and aDNA genomics research [29], more clues about selection factors should emerge.

Acknowledgements We thank three archeologists who contributed to this study with encouragement, project design, and/or interpretation of results coupled to manuscript revisions: Sir Barry Cunliffe (Oxford University), William Aylward (University of Wisconsin-Madison), and T. Douglas Price (University of Wisconsin-Madison and the University of Aarhus). In addition, Patrick Sosnay (Johns Hopkins University) and Mark Loudon (University of Wisconsin-Madison) contributed significantly to the manuscript. We are also grateful to Anita Laxova for her important efforts during the organization phase and for securing approvals from the Human Subjects IRB at the University of Wisconsin-Madison. The haplotype and *Estiage* analyses were performed with support from Inserm UMR 1078 under the leadership of Drs. Férec and Génin, while Dr. Farrell was funded by NIH grant DK 34108 and Dr. Macek by IP00064203/6003, CZ.02.1.01/0.0/0.0/16_013/0001634, LM2015091.

Author contributions PF, as the overall project leader for investigation of “The Ancient Origin of Cystic Fibrosis,” proposed an assessment of the “age” of the p.Phe508del variant across Europe, obtained IRB approval, requested and secured the specimens as well as coordinating shipments with the Laboratoire de Génétique in Brest, and completed the initial manuscript draft and final version, while participating in data interpretations and consulting with archeologists. CF has been responsible for the design, genetics laboratory leadership, selection of European regions for evaluation, haplotype determinations, data interpretations, manuscript drafting, and revising. MM reviewed/approved the expanded design, provided appropriate DNA specimens from Czech CF patients and their parents, participated in data interpretation, and then during the final manuscript preparation process contributed an important comparison of the geographic, genetic, and archaeologic interpretations with insights that added significantly to our conclusions. TF and SR, with the assistance of KR, after reviewing and approving the design, obtained ethics committee approval in

Vienna, selected carefully the native Austrian patient/parent trios, obtained the blood specimens, harvested the leukocyte fractions, and then reviewed and approved the manuscript. DB, with the assistance of MR, reviewed/approved the design, provided appropriate DNA specimens from Irish CF patients and their parents, participated in data interpretation, and then reviewed and revised the manuscript. TR reviewed/approved the design, provided appropriate specimens from native Albanian CF patients and their parents who attend the CF clinic in Florence, and lastly reviewed and revised the manuscript. MT reviewed/approved the design, provided DNA specimens from Greek and Albanian CF patients and their parents, then participated in data interpretation, and lastly reviewed and revised the manuscript. KG performed the lab work and the genetic analysis, and approved the manuscript. MD reviewed/approved the design, provided appropriate DNA specimens from Danish CF patients and their parents, participated in data interpretation, and reviewed and revised the manuscript. HL provided appropriate DNA specimens from Wisconsin-based, CF patients and their parents with known ancestry, participated in data interpretation, and reviewed and revised the manuscript. MS contributed to the data management and prepared the data for the statistical analyses and approved the manuscript. YF designed the haplotype experiments, supervised the essential lab work to generate data for *Estiage* analyses, and approved the manuscript. CLM designed the genetics experiments, supervised the lab work, reviewed/revised and approved the manuscript, and has provided important advice for the overall project. EG developed the essential *Estiage* method, led the design efforts, was responsible for the strategy employed for haplotype data analyses via *Estiage*, performed the analyses and interpretations of results, created most of the tables, wrote the initial Methods and Results sections of the manuscript, and assumed responsibility for the manuscript revisions and the final version.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Kerem B, Rommens JM, Buchanan JA, et al. Identification of the cystic fibrosis gene: genetic analysis. *Science*. 1989;245:1073–80.
- Bobadilla JL, Macek Jr M, Fine JP, Farrell PM. Cystic fibrosis: a worldwide analysis of CFTR mutations-correlation with incidence data and application to screening. *Hum Mutat*. 2002;19:575–606.
- De Braekeleer M, Daigneault J. Spatial distribution of the DF508 mutation in cystic fibrosis: a review. *Hum Biol*. 1992;64:167–74.
- Lucotte G, Hazout S, De Braekeleer M. Complete map of cystic fibrosis mutation DF508 frequencies in western Europe and correlation between mutation frequencies and incidence of disease. *Hum Biol*. 1995;67:797–803.
- Farrell P, Le Marechal C, Ferec C, Siker M, Teschler-Nicola M. Discovery of the principal cystic fibrosis mutation (F508del) in ancient DNA from Iron Age Europeans. *Nat Preced*. 2007;1276:1 <http://hdl.nature.com/10101/npre>.
- Morral N, Bertranpetit J, Estivill X, et al. The origin of the major cystic fibrosis mutation (delta F508) in European populations. *Nat Genet*. 1994;7:169–75.
- Fichou Y, Génin E, Le Maréchal C, Audrézet MP, Scotet V, Férec C. Estimating the age of CFTR mutations predominantly found in Brittany (Western France). *J Cyst Fibros*. 2008;7:168–73.
- Strom CM, Strom CM, Crossley B, et al. Cystic fibrosis testing 8 years on: lessons learned from carrier screening and sequencing analysis. *Genet Med*. 2011;13:166–72.
- Castellani C, Massie J, Sontag M, Southern KW. Newborn screening for cystic fibrosis. *Lancet Respir Med*. 2016;4:653–61.
- Gregg RG, Wilfond BS, Farrell PM, Laxova A, Hassemer D, Mischler EH. The application of DNA analysis in a population screening program for neonatal diagnosis of cystic fibrosis: comparison of screening protocols. *Am J Hum Genet*. 1993;52:616–26.
- Serre JL, Simon-Bouy B, Mornet E, et al. Studies of RFLP closely linked to the cystic fibrosis locus throughout Europe lead to new considerations in population genetics. *Hum Genet*. 1990;84:449–54.
- Morral N, Nunes V, Casals T, et al. Microsatellite haplotypes for cystic fibrosis: mutation frameworks and evolutionary tracers. *Hum Mol Genet*. 1993;2:1015–22.
- Romeo G, Devoto M, Galletta LVJ. Why is the cystic fibrosis gene so frequent? *Hum Genet*. 1989;84:1–5.
- Rodman DM, Zamudio S. The cystic fibrosis heterozygote--advantage in surviving cholera? *Med Hypotheses*. 1991;36:253–8.
- Cuthbert AW, Halstead J, Ratcliff R, Colledge WH, Evans MJ. The genetic advantage hypothesis in cystic fibrosis heterozygotes: a murine study. *J Physiol*. 1995;15:449–54.
- Piel FB, Patil AP, Howes RE, et al. Global distribution of the sickle cell gene and geographical confirmation of the malaria hypothesis. *Nat Commun*. 2010;1104:1–7.
- Bitoungui VJ, Pule GD, Hanchard N, Ngogang J, Wonkam A. Beta-globin gene haplotypes among cameroonians and review of the global distribution: is there a case for a single sickle mutation origin in Africa? *OMICS*. 2015;19:171–9.
- Pauling L, Itano HA, Singer SJ, Wells IC. Sickle cell anemia, a molecular disease. *Science*. 1949;110:543–8.
- Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *Am J Hum Genet*. 2007;81:1084–97.
- Génin E, Tullio-Pelet A, Lyonnet S, Abel L. Estimating the age of rare disease mutations: the example of Triple A syndrome. *J Med Genet*. 2004;41:445–9.
- Cunliffe B. *Europe between the Oceans*. New Haven: Yale University Press; 2008.
- Price TD. *Europe before Rome*. New York: Oxford University Press; 2013.
- Cordovado SK, Hendrix M, Greene CN, et al. *CFTR* mutation analysis and haplotype associations in DNA used for proficiency testing. *Mol Genet Metab*. 2012;105:249–54.
- Lao O, Lu TT, Nothnagel M, et al. Correlation between genetic and geographic structure in Europe. *Curr Biol*. 2008;18:1241–8.
- Heath SC, Gut IG, Brennan P, et al. Investigation of the fine structure of European populations with applications to disease association studies. *Eur J Hum Genet*. 2008;16:1413–29.
- Kaplan NL, Lewis PO, Weir BS. Age of the $\Delta F508$ cystic fibrosis mutation. *Nat Genet*. 1994;8:216–7.
- Allentoft ME, Sikora M, Sjögren KG, et al. Population genomics of Bronze Age Eurasia. *Nature*. 2015;522:167–72.
- Müller J, van Willigen S. New radiocarbon evidence for European Bell Beakers and the consequences for the diffusion of the Bell Beaker phenomenon. In: Nicolis F, editors. *Bell Beakers today. Pottery, people, culture, symbols in prehistoric Europe (Proceedings of the International Colloquium, Riva del Garda, Trento, Italy); 2001. pp. 59–80.*
- Olalde S, Brace ME, Allentoft I, et al. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature*. 2018;555:190–6.
- Price TD, Knipper C, Grupe G, Smrcka V. Strontium isotopes and prehistoric human migration: the Bell Beaker period in central Europe. *Eur J Archaeol*. 2004;7:9–40.

31. Needham S. Transforming Beaker culture in north-west Europe: processes of fusion and fission. *Proc Prehist Soc.* 2005;71:171–217.
32. Farrell PM. The prevalence of cystic fibrosis in the European Union. *J Cyst Fibros.* 2008;7:450–3.
33. Harrison RJ. *The Beaker Folk: Copper Age archaeology in Western Europe.* London, UK: Thames and Hudson; 1980.
34. Poolman EM, Galvani AP. Evaluating candidate agents of selective pressure for cystic fibrosis. *J R Soc Interface.* 2007;4: 91–8.