REVIEW ARTICLE
THE CRADLE OF THE ΔF508 MUTATION
Danish Saleheen1,2, Philippe M. Frossard1
1 Department of Biological and Biomedical Sciences, Aga Khan University, Stadium Road, Karachi, Pakistan.
2 Molecular Epidemiology Unit, Strangeways Research Labs, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Cystic fibrosis (CF) is the most common autosomal recessive disorder caused due to mutation/s in the CFTR gene. The most common mutation in CFTR worldwide is ΔF508 and cystic fibrosis genetic analysis consortium revealed that this mutation is responsible for approximately 66% of all CF chromosomes in the world. Studies looking at the DNA polymorphic haplotypes created by CF linked markers suggest that ΔF508 has a single origin as this mutation has been found associated exclusively with one marker haplotype. Despite a high prevalence of this mutation in CF patients in northern parts of Europe, findings suggest that this mutation was not spread by Europeans but by a group that is speculated to have originated in the Middle East or a more eastern region in Asia (most likely subcontinent). Over here we have given a brief introduction to cystic fibrosis and classification of CFTR mutations and have further elaborated on the crucial issue about the spread of the ΔF508 mutation. We have reviewed findings that give clues about the origin of this mutation from the Baluch ethnicity residing in Pakistan.

Keywords: CFTR, Delta F508, Mutation

Cystic fibrosis (CF) is the most common life-limiting autosomal recessive disorder; it affects nearly 1/2500 live births in Caucasians.1 CF is a multi-system disease and can involve secretory cells, sinuses, lungs, pancreas, liver and the reproductive tract.1 The disease has a “classical” appearance in 90% of cases and patients present with a triad of chronic obstructive pulmonary disease, exocrine pancreatic insufficiency and elevation of sodium and chloride concentrations in the sweat.1,2 After extensive research in the field, in western countries, 60% of all CF patients have a diagnosis of the disease before their first birthday and 90% of all CF patients are diagnosed by 10 years of age. This has greatly affected the life expectancy of CF patients as compared to the past.1

The first evidence of CF being a monogenic disorder came from the reviews of CF pedigrees in Italian church records in 1985.2 Later on identification and cloning of the CFTR gene on chromosome 7 served as the basis for understanding the genetic aetiology of the disease.4–5 The gene causing CF spans 190 Kb of DNA and encodes a 1480 amino acid polypeptide named cystic fibrosis transmembrane regulator (CFTR).4 To date 1529 mutations have been identified in the gene.7 These mutations have a variety of effects at the cellular level. However they can be grouped into 6 categories8, in which they:

i. block the synthesis of CFTR,
ii. lead to defective processing
iii. cause abnormal regulation of CFTR Cl channel
iv. disrupt the normal conductive pathway of the channel
v. lead to complete or partial production of CFTR protein
vi. lead to defective regulation of other channels.

The most common mutation in CFTR worldwide is ΔF508.1 It is a class II mutation caused by deletion of three bases encoding a phenylalanine residue at position 508 within the nucleotide binding domain (NBD).9 Heterologous expression system analyses have shown that this mutation results in mis-folding of the CFTR and hence mislocalization and defective processing of the mature protein.9 It was postulated that the mutation is severe enough to result in absence of any mature protein in the affected cells.10 Recent immuno-cytochemical studies of intestinal, respiratory and hepatobiliary epithelia, however, show that in patients homozygous for ΔF508, a proportion of CFTR protein was shown to be targeted to the apical membranes.11 The fact was lent support by in vitro studies done on ΔF508 homozygous CF patients suggesting that at least some CFTR molecules reach the plasma membrane in these patients.12 Genotype-phenotype correlations clearly indicate that this mutation is severe as far as pancreatic status is concerned but conclusions on its effect on lung disease vary between studies.8

Cystic fibrosis genetic analysis consortium revealed that ΔF508 is responsible for approximately 66% of all CF chromosomes in the world.13 However its frequency varies from region to region.13 The highest frequency has been found in patients from Northern Europe, where it accounts for 75%-88% of all CF alleles.14 Indeed, the frequency reaches to a maximum of 100% in the isolated Faroe Islands of Denmark.15 Importantly, evaluation of studies that have been done on European populations reveal a clear-cut north-west to south-east gradient of ΔF508 prevalence in CF patients showing a maximum prevalence of 88% in Denmark (excluding Faroe
islands) to a minimum of 24.5% in the Turkish population.\textsuperscript{13–18}

Although studies investigating the origin of ΔF508 have failed to identify any founder population but have established strong evidence in ruling out various potential regions (for eg. Faroe islands) where the mutation could have possibly originated.\textsuperscript{13,16–19} These DNA based studies normally utilize single nucleotide polymorphisms (SNP’s) and several short tandem repeat polymorphisms which are also known as ‘microsatellites’ within a gene of interest to find and trace the origin of a mutation. By utilizing such a methodology, evolutionary origin of some CF causing mutations have been identified; for instance CFTR mutation G542X had a single origin in the ancient Phoenicians.\textsuperscript{20} Similarly CFTR mutation G551D has been associated with Celtic tribes.\textsuperscript{21}

SNP’s or microsatellite markers by virtue of linkage disequilibrium can be used to create haplotypic maps. A haplotype containing a CFTR mutation or a SNP linked with a CFTR mutation displays a geographical gradient starting from a higher frequency in the founder population residing in a specific region to a lower frequency in a region where migration from the prior region has occurred lately. Based on this principle ancestry of a CFTR mutation G551D was revealed, as one particular haplotype containing this mutation remained unaltered from more than 170 generations and displayed variable frequencies in different regions.\textsuperscript{20,22}

Studies looking at the DNA polymorphic haplotypes created by CF linked markers suggest that ΔF508 has a single origin.\textsuperscript{16} This mutation has been found associated exclusively with one marker haplotype. Morral \textit{et al.} have suggested that the original time of this mutation ranges from 3000 years to 53000 years ago.\textsuperscript{16} Despite a high prevalence in northern parts of Europe, findings suggest that the mutation is more ancient and did not originate in Europe.\textsuperscript{16–18} Moreover, analyses of wild type chromosome containing CFTR gene of various European populations suggest that the general genetic background of these populations is different than the diseased CFTR chromosome and thus the origin of most frequent CF mutations should be non-European.\textsuperscript{17–18}

Based on these findings, Mateu \textit{et al.} investigated the origins of five common CFTR mutations including ΔF508 in 949 unrelated autochthonous healthy individuals of various geographical origins and concluded that the haplotype containing ΔF508 was present in very low frequencies in the European populations.\textsuperscript{19} However, the highest frequency was detected in one Middle Eastern population.\textsuperscript{19} Although Mateu \textit{et al.} did not report the data on the different Middle Eastern ethnicities, a comprehensive pattern of CF causing mutations in two Middle Eastern regions has been reported by Frossard \textit{et al.}\textsuperscript{23,24} The data from Middle East show high frequencies of ΔF508 in particular ethnicities. Despite the fact that data reported by Frossard \textit{et al.} does not show the haplotype frequencies, it is now widely believed that ΔF508 is more ancient and that it was not spread by Europeans but by a group that preceded them.\textsuperscript{16–19,21,23} This group is speculated to have originated in the “Middle East” or a more eastern region in Asia (most likely subcontinent).\textsuperscript{23,25} Thus it is important to study this mutation along with specific haplotypes in the CF patients of Middle East, Pakistani and Indian origin to discover more about its origins.

Studies on CF in indigenous subcontinental population are severely lacking. Two studies have been done in this regard looking at the prevalence of ΔF508 prevalence in CF patients. Both studies were done on patients who were admitted to tertiary care hospitals. In both the studies the patients were diagnosed of CF by pilocarpine iontophoresis test initially. Study done by Bhutta \textit{ZA et al.} on Pakistani population was limited to very small sample size.\textsuperscript{26} They evaluated 15 patients belonging to different ethnicities in Pakistan and found a 60% prevalence of this mutation in the Pakistani CF population. The other study done by Kabra \textit{SK et al.} looked at a patient population coming from different regions of India and Pakistan. ΔF508 was identified in 19% chromosomes out of 290 tested.\textsuperscript{27} It is surprising to note that ΔF508 was highest (56%) in patients of Pakistani origin as compared to a 12% prevalence of this mutation in Indian CF patients.\textsuperscript{27} Findings of both studies could be challenged on the fact that ΔF508 is a severe mutation as far as pancreatic status is concerned and patients carrying this mutation are hence more likely to be hospitalized compared to those with milder mutations, which would confer a spuriously high prevalence of ΔF508 in a hospital-based sample population. A 42% difference in prevalence between Pakistani and Indian CF populations, however, cannot be ignored.

It must be kept in mind that the ‘Pakistani population’ does not have a homogenous distribution and is composed of people belonging to different ethnicities. It is highly essential to look at the individual ethnic data while studying the genetic mutation spectrum of a disorder like CF. Studies looking at the ΔF508 prevalence in CF patients of Pakistani origin living overseas have been variable and have looked at heterogeneous Pakistani populations rather than individual ethnicities. The most striking evidence of a high prevalence of ΔF508 in a Pakistani ethnic group comes from the findings of Frossard \textit{et al.}, who studied CF genetic mutations in the United Arab Emirates.\textsuperscript{23} They investigated 17 unrelated families and found a 86% prevalence (14/16 alleles) of ΔF508 in Baluchis
residing in UAE. Out of this Baluch population, all Pakistani Baluch CF patients carried ΔF508 mutation (100% prevalence). These findings were subsequently confirmed in the Omani population, in which 100% of CF patients of Baluch origin were also ΔF508 homozygotes. Keeping in view the findings discussed earlier about a non-European origin of ΔF508 and a high prevalence of ΔF508 found in Pakistani Baluchis, it has been proposed that this mutation might have originally arisen in Baluchis. This hypothesis is analogous to the findings of Quaife et al. who studied the spectrum of beta-thalassemia in the UAE population and concluded that some specific mutations were introduced into UAE population by Baluchi immigrants.

The Baluch ethnicity currently resides in Iran, Pakistan (Baluchistan) and Afghanistan, and its people are known to have immigrated through successive waves to the Punjab, India and Gulf countries (Oman and the UAE). However, the actual geographical location of Baluchis refers only to the western province of Pakistan, Baluchistan. To prove a founding effect of ΔF508 mutation in the Pakistani Baluch population, it is highly essential to study Baluchis of Pakistani province Baluchistan using appropriate haplotype analyses. This will answer the crucial question in the field of CF to know whether indeed ‘Baluchistan’ was the cradle of the ΔF508 mutation.

Figure 1: Starred line represents approximate boundaries of the Baluch ethnicity. The arrow indicates the migrations that occurred historically from Baluchistan to Oman and the UAE.

REFERENCES


Address for Correspondence:
Dr. Danish Saleheen, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge CB2 8RN. Tel: 01223 741302, Fax: 01223 741339.
Email: ds436@medschl.cam.ac.uk