Cystic Fibrosis in the African Diaspora

Cheryl Stewart & Michael S. Pepper

Department of Immunology and Institute for Cellular and Molecular Medicine, and SAMRC Extramural Unit for Stem Cell Research and Therapy, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa

Author for correspondence

Prof. Michael S. Pepper, Dept. of Immunology, Faculty of Health Sciences, University of Pretoria, P.O. Box 2034, 0001 Pretoria, South Africa.

Tel: +27 12 319 2179

Fax: +27 12 319 2946

E-mail: michael.pepper@up.ac.za

This work was supported by the University of Pretoria via the Vice Chancellor's Postdoctoral Fellowship, the Institute for Cellular & Molecular Medicine, and the Genomics Research Institute, and by the South African Medical Research Council.

Keywords: Americas, genetic testing, DNA sequencing, neonatal screening, targeted pharmacotherapy

Running Head: CF in the Diaspora

Word count: 3,496

Abstract

Identifying mutations that cause cystic fibrosis (CF) is important for making an early, unambiguous diagnosis, which in turn is linked to better health and a greater life expectancy. In patients of African descent, a molecular diagnosis is often confounded by the fact that the majority of investigations undertaken to identify causative mutations have been conducted on European populations, and CF-causing mutations tend to be population specific. We undertook a survey of published data with the aim of identifying causative CF mutations in patients of African descent in the Americas. We found that 1,584 chromosomes had been tested in only six countries of which 876 alleles (55.3%) still remained unidentified. There were 59 mutations identified – 41 of which have been shown to cause CF, 17 have no associated functional studies and one (R117H) is of varying clinical consequence. The most common mutations identified in the Diaspora were ΔF508 (29.4%; identified in America, Colombia, Brazil and Venezuela), 3120+1G>A (8.4%; identified in Brazil, America and Colombia), G85E (3.8% identified in Brazil), 1811+1.6kbA>G (3.7% identified in Colombia), and 1342-1G>C (3.1% identified in America). The majority of the mutations identified (81.4%) have been described in just one country. Our findings indicate that there is a need to fully characterise the spectrum of CF mutations in the Diaspora in order to improve diagnostic accuracy for these patients and facilitate treatment.

Introduction

"Woe to that child which when kissed on the forehead tastes salty. He is bewitched and soon must die." [1] This dire warning from the 1800s was about children born with cystic fibrosis (CF). In the absence of medical intervention, CF patients were not expected to see their first birthday – in fact life expectancy for CF patients in 1938 was only 6 months[2]. In addition to elevated sweat chloride levels, classic CF symptomology includes chronic lung disease and pancreatic insufficiency[3]. As our understanding of the underlying cause of the disease has grown, new diagnostic and treatment strategies have been devised which have raised life expectancy for CF patients around the world. This improvement has not been uniform: CF patient life expectancy can range from 20.5 years in South Africa[4] to 49.7 years in Canada[5].

Mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene interfere with the ability of this ion channel protein to regulate the chloride balance across apical membranes. This results in a depleted airway surface liquid layer, impeding the mucociliary clearance mechanism and creating an environment in which bacteria (usually *Pseudomonas aeruginosa*) can establish an infection[2]. Over time, *P. aeruginosa* ceases to exist solely in its planktonic form and forms a sessile community within a protective biofilm which is largely impervious both to antibiotics and the immune response[6]. It should be noted that respiratory complications are the leading cause of death among CF patients[7].

More than 2,000 mutations have been identified in CF patients around the world[8]. Causative mutations tend to be population specific, varying by both country of origin and race[9]. CF mutations have been organised into six classes based on the effect they have on the CFTR, which can range from no functional CFTR being produced (class I) to reduced amounts of functional CFTR at the cell surface (class VI)[2]. Until recently, CF treatment has focused on

symptom alleviation. However, the Food and Drug Administration's (FDA) approval of Ivacaftor in 2012 marked the first application of pharmacotherapy that directly addresses the molecular cause of the disease. Ivacaftor was developed to treat G551D, a class III gating mutation. Its use has since been expanded for the treatment of additional class III mutations[2]. Further investigations are underway to identify other class specific drugs for CF.

Cystic Fibrosis and Race

Cystic fibrosis was historically seen as a disease that was limited to a given race but research has since shown that CF is not ethnically linked[10]. Despite the fact that one of the earliest reports of CF in a Black patient was published in 1959[11], the notion that CF only occurs in White children has lingered. This assumption led to the majority of early investigations into the cause of CF being conducted in Caucasian populations which may be why the countries with a mutation detection rate in excess of 95% are all European[12]. Another consequence of this lack of data is the fact that there are more likely to be false negative results for a CF patient that is not of Caucasian descent[10,13]. This may result in a delayed diagnosis (or misdiagnosis) which can negatively impact patient health, longevity and treatment costs[7]. One study suggested that patients diagnosed after they were six weeks old had double the risk of developing severe pulmonary disease as they aged[14]. Delays in diagnosis have also been linked to worse pulmonary function, nutritional status, cognitive function, and premature death[7]. It has been estimated, based on CF registry data, that each patient in whom diagnosis is delayed would spend about one million euros more on treatment over their lifetime than a patient identified by a newborn screening programme[15].

There have since been many reports of CF in patients of varying ethnicities including reviews of the molecular epidemiology of CF in Africa[13] and Latin America[16]. Since members of the African Diaspora could be expected to have genomes derived from admixture between Europeans and Africans, and since all populations are known to be affected by this disease, it would be reasonable to assume that CF would be a source of morbidity in the Diaspora. We defined the African Diaspora as individuals of African descent in the Americas, especially the descendants of West Africans who were conscripted to work as part of the Trans-Atlantic Slave Trade. We focused on the Americas (North, South, and Central America, and the Caribbean) since this is the region where the majority of Africans were taken. We used Google Scholar to find papers that characterised mutations in African Diaspora patients. Given that identifying CF-causing mutations has become increasingly important for diagnosis and targeted pharmacotherapy, we reviewed the literature to determine the extent to which mutations have been identified in the Americas. We also wanted to determine if there were any region-specific issues facing CF patients.

Investigations in the Diaspora

In the Americas, there were 18 papers describing the identification of mutations in CF patients from six countries: the United States of America[17-25], Haiti[26], French Guiana[27], Brazil[28-33], Colombia[34] and Venezuela[35]. Together, these authors assessed 1,584 CF chromosomes in members of the African Diaspora but used a variety of methods to identify mutations. None of the papers included relied solely on an unbiased sequencing approach (supplementary table E1) and thus 876 alleles remained unidentified (55.3% of all chromosomes tested). Although the Caribbean has the highest percentage of African descendants in the general population in the Americas, there are no published records of any local attempts to identify CF causing mutations in members of the Diaspora living in this geographical region.

There were 59 mutations identified in the Diaspora, 41 of which have been shown to cause CF, 17 have no associated functional studies and one (R117H) is of varying clinical consequence (table 1). As expected, mutations were identified in the Diaspora that had previously been associated with predominantly White (such as ΔF508) or Black (3120+1G>A, A559T) patients. When compared with mutations identified in African CF patients[13] there was an overlap of 18 mutations of which only I1203V has no associated functional study. Of the West African countries largely involved in the Trans-Atlantic Slave Trade, only Senegal has any published reports of CF causing mutations[13] and none of these were identified in the Diaspora (table 2). It is possible that the mutations in the overlap could also be present in the relevant West African populations which could explain their presence in the Americas. There were 41 mutations found in the Diaspora that have not yet been identified in African patients.

Mutations that are associated with a severe phenotype usually result in pancreatic insufficiency. The CFTR2 database (www.cftr2.org) has information on more than 66,000 CF patients, largely from European and American CF clinics. Of the mutations identified in the Diaspora, 42 were present in this database. We searched CFTR2 for each of these 42 mutations in order to determine how many of them were associated with pancreatic insufficiency (a marker of mutation severity) in the majority of the patients having that particular mutation in the database. Most of the mutations (92.5%) identified in the Diaspora patients and included in CFTR2 are associated with pancreatic insufficiency, which indicates a severe phenotype.

The most frequent test result was an unknown mutation (U; 56.5%). The most common mutations identified in the Diaspora were ΔF508 (29.4%; identified in the United States[17-22], Colombia[34], Brazil[28-32] and Venezuela[35]), 3120+1G>A (8.4%; identified in Brazil[29, 31], the United States[18, 19, 21] and Colombia[34]), G85E (3.8% identified in Brazil[29, 31]), 1811+1.6kbA>G (3.7%

identified in Colombia[34]), and 1342-1G>C (3.1% identified in the United States[22]). Most of these mutations (79.7%) have been found in just one country, which supports the idea that mutations tend to be population specific. The majority of the alleles assessed were from African Americans (764 chromosomes) or Afro-Brazilians (710 chromosomes).

There are some limitations imposed by the data. The majority of the data has been derived from Brazil and the USA but the Caribbean, where significant proportions of the population are of African descent, is grossly under represented. Additionally, no study utilised an unbiased sequencing strategy that covered the entire *CFTR* gene and many relied on kits that could only detect certain mutations. There are several consequences to this. First, novel mutations would not be identified. Second, African genomes are known to possess unparalleled genetic diversity[36]. It is therefore reasonable to assume that CF causing mutations would be present in the Diaspora that would not be included on any available genetic test. Third, since commercial kits are largely based on European data[37], it may explain the observed low mutation detection rate. Fourth, since there has not been a standard approach to identifying mutations over the 19-year span of these publications, it is possible that some of the unknown mutations may have been uncovered if all patients had been tested for all 59 mutations identified. Finally, given the population specificity of these mutations[9], having so much of the Caribbean and Latin America excluded from these investigations represents a gap in the literature that could negatively impact on the ability of clinicians to accurately diagnose CF patients in these regions and in the broader Diaspora.

Issues Surrounding Diagnosing CF in the Diaspora

The Trans-Atlantic Slave Trade was responsible for the movement of an estimated 12 million people, largely from West Africa, to the Americas to work as slaves on plantations. This forced exodus from the Mother Continent altered the course of history in many ways including bringing African genomes to the West in large numbers. It should be possible to trace the ancestry of CF patients in the Diaspora using the mutations they carry. However, the molecular epidemiology of CF in Africa remains largely unexplored. The majority of the work on the continent has been done in northern and southern African states[13], which precludes using the data to speculate about the ancestry of CF patients in the Diaspora. For instance, 3120+1G>A has been identified in Rwanda, South Africa and Zimbabwe, and in Brazil, Colombia, and the United States of America (table 2). There are no obvious historical migratory patterns linking these eastern and western countries. While available data cannot be used to trace ancestry, it does highlight that CF patients of African descent are at a distinct diagnostic disadvantage which negatively affects their prognosis.

Molecular diagnosis of CF patients – almost 30 years after the identification of the responsible gene – continues to be inadequate because of the sheer diversity of causative mutations (more than 2,000 variants have been identified in CF patients around the world[8]) and their population specific nature[9]. This is further compounded when working in a clinical setting that serves a genetically diverse population, such as might be found among Africans and members of the African Diaspora. Admixture also contributes an additional layer of complexity. Although African genomes are known to be highly diverse[36] they also tend to be understudied which is the case with CF, especially since it was assumed that this disease could only affect those of European or Caucasian descent[37].

If we examine the data from Brazil, we also see that when a set of known mutations is used to identify CF patients, White patients have a significantly higher mutation detection rate (80.7%) than Blacks (21.1%)[31]. This may be because the majority of the mutations on the panel had previously been identified in Europeans. The diversity resident in African genomes is likely to play a role here. These findings underscore the need to investigate the mutations relevant in the Diaspora in order to reduce the likelihood that CF patients will receive a delayed diagnosis or be misdiagnosed.

Further assessment of the data also revealed that 48 of the 59 mutations (81%) had been identified in only one country (table 1). This is not unexpected since causative mutations have repeatedly been demonstrated to exhibit population specificity. This does mean that genetic tests need to be developed for a given population instead of taking a one-test-fits-all approach. There was only one country (USA) where the most common test result was not an unknown mutation (supplementary table E1). As is the global trend, a few mutations in this data set account for the majority of the alleles tested, while the majority of mutations identified occurred at frequencies below 3% (table 1).

Misdiagnosis of CF patients of African descent is possible because it is a relatively rare disease and other phenocopic illnesses are more likely to be perceived as the cause of the patient's symptoms[31,38]. If the attending physician is unaware that CF is not an ethnically linked disease, this also increases the likelihood that members of the Diaspora may be misdiagnosed[13,38]. The phenotypes of CF patients exist along a continuum[37] in which a single test may not be able to lead to a conclusive diagnosis. For instance, a CF patient may have a borderline sweat test result, normal or intermediate faecal elastase result, and none or only one mutation identified by the genetic test in use in his/her country. It may therefore be difficult for a physician to diagnose

CF, especially for patients that do not present with the classic symptoms. Sequencing of these patients' *CFTR* may be helpful in resolving their diagnosis[39].

Currently the highest mutation detection rates are among European populations (about 100% in Finland, 80% in Europe, 60% in South Asia, 40% in the Americas and about 45% in Africa)[37]. Using the panel recommended by the American College of Medical Genetics (ACMG)[40] only about 30% of CF carriers in the Americas (an admixed group from North, South, and Central America) would be detected[37]. This percentage increases to 40%[37] if all the CFTR2 mutations[41] are used to screen these individuals. Additional research is definitely needed to identify the set of mutations that would raise the detection rate above 95% in the Diaspora. If the ACMG panel[40] had been used to screen this Diaspora cohort, 65.3% of the mutations would have remained unidentified, which raises the question: should we continue to use genetic screening for diagnosis, especially in non-Caucasian populations?

Using next generation sequencing (NGS) to diagnose CF outperforms conventional genetic testing in terms of mutation detection rates[42] and positive predictive value[43]. One approach combined sequencing 182 of the 189 kb that constitutes *CFTR* with a bioinformatics pipeline that was able to identify SNPs, indels and gross rearrangements to achieve a diagnostic rate of 98.9%. The authors estimated that a traditional multi-tier diagnostic algorithm could cost €400 and take 2-3 months whereas their strategy cost about €200 and was completed in 14 days[44]. A more recent paper covered significantly fewer bases using NGS (16.5 kb as opposed to 182 kb) but the associated bioinformatics pipeline had a positive predictive value of 100% and identified SNPs, indels and large deletions. This protocol was estimated to cost $15 per sample if 95 samples were sequenced per run and the analysis would be complete within 3 days[45]. It may therefore be time to consider a similar approach to diagnosing CF patients, especially those that

are not of European descent who are underserved by current genetic tests. It should also be borne in mind that with the advent of Ivacaftor, (and with other class specific drugs being developed[2,46]) being able to resolve the molecular status of CF patients has become an important precursor for targeted pharmacotherapy.

Reducing the length of the diagnostic odyssey in these patient groups will have a positive impact on morbidity and mortality. CF patients that were identified via newborn screening had better height and age-for-weight Z-scores during the first six months post-partum when compared to those diagnosed based on their symptoms (and thus diagnosed later), a gap which persisted throughout the course of the study. This effect was most pronounced with height-for-age suggesting that delayed diagnosis may be associated with permanent stunting[47]. Delays in diagnosis also result in prolonged vitamin deficiency which can affect cognition. CF patients with α-tocopherol deficiency at diagnosis were stratified into those identified by screening or those identified symptomatically. When these patients were given a Test of Cognitive Skills, the delayed diagnosis group had Cognitive Skills Index scores 12.5 points less than their screened counterparts[7]. Approximately 50% of symptomatically diagnosed patients had potentially irreversible lung damage by age two as opposed to 29% of patients who had been diagnosed via newborn screening. Screened patients had better FEV_1% scores than those diagnosed later but this difference shrank with time until the better predictor of lung function was *Pseudomonas aeruginosa* infection[7,47].

The Way Forward

Based on the available data, CF patients in the African Diaspora are at a distinct disadvantage when compared to their European counterparts[11,37]. Global data suggests that the number of CF causing mutations in patients of European descent may be approaching a plateau. Conversely, there seems to be many more mutations yet to be identified in the *CFTR* of patients

of African descent. CF patients in the Diaspora face two major obstacles to an early and accurate diagnosis. First, they have a relatively rare disease which occurs at a lower frequency in their population group than in those of European descent[10]. This may increase their chances of being misdiagnosed especially in areas where there are more rampant phenocopic illnesses such as malnutrition, viral or parasitic infection or tuberculosis[11,38]. Second, there is not enough information available about the mutations present in these sub-populations for the design of suitable genetic tests[12,37]. This also increases the probability that these patients may be misdiagnosed particularly if they don't present with the classic triad of CF symptoms or if they have milder forms of the disease.

At the very least, it would be useful to sequence the *CFTR* in individuals in the Diaspora suspected of having CF (the World Health Organisation recommends at least 50-100 patients[12]). Since causative mutations tend to be population specific[9] and given the diversity known to reside in African genomes[36], it would be reasonable to assume that several mutations would be identified by this survey and that some of them may well be novel. This information could be used to design genetic tests that are tailored to each sub-population in the Diaspora which would raise the mutation detection rate. These tests could serve an important role in reducing the false negative rates among Diaspora patients and could assist in reducing the length of time required for these patients to receive an accurate diagnosis. As has been repeatedly demonstrated, early diagnosis is correlated with better prognosis and lower lifetime treatment costs[7,14,15]. Being aware of which mutations each patient carries also opens up new therapeutic options as more class specific drugs are developed[2,46]. As recent literature suggests[44, 45], sequencing has several advantages over panel tests and may be a cheaper and faster way to arrive at a diagnosis.

There may be objections raised regarding changing current diagnostic algorithms, but this must be considered in light of the cost of doing nothing. In Latin America for example the mean age of diagnosis for CF is 3.7 years while the mean age at which patients are first seen in clinics is 4.5 years[48]. This kind of delayed diagnosis and lack of access to appropriate treatment is known to have a negative impact on prognosis and quality of life[7,14]. The median age at death is 6.68 years and only 10% of patients survived to adulthood[48], which is in stark contrast to Canada where CF life expectancy is almost 50 years[5]. Approximately 25% of these Latin American patients had normal lung function (FEV_1% above 80%) and 80% reported microbial colonisation of their airways[48]. These dire health statistics concur with the World Health Organisation's (WHO) statement that "the societal costs of inaction in genetics, measured in terms of avoidable human suffering and the burden on public health are very high." The WHO therefore suggested that countries "recognise that there are approaches to their [genetic disease] management and prevention that can significantly reduce their burden in a cost-effective manner" [49]. Improving diagnostic accuracy and helping patients access appropriate care is one way to improve the health of CF patients in the Diaspora.

In the clinical setting, it is therefore important not to exclude CF as a diagnosis based on race. When dealing with genetically diverse populations, sequencing will be needed to identify relevant mutations that are absent from commercially available genetic tests[37,39]. This process would generate data that could resolve the patient's CF status, be useful in designing a genetic test with a higher mutation detection rate and determine if a patient may benefit from one of the class-specific drugs that have recently become available[2,46]. Alternatively, given recent data that suggests that pairing NGS with an appropriate bioinformatics pipeline could significantly reduce both the time and cost involved in CF diagnosis[44,45], it may be time to begin switching to sequencing as the primary diagnostic method. This should have particular utility in diverse

populations such as are found in Africa and her Diaspora, which have suffered from the inherent European bias in the current genetic tests. If this approach is taken within the context of a larger public health policy, it could lower the age at diagnosis for CF patients which should both decrease morbidity and raise life expectancy[13]. Critically, it would gradually eliminate the immediate diagnostic disadvantage faced by non-Caucasian CF patients.

References

1. Hodson M, Bush A, Geddes D, editors. Cystic Fibrosis, 3rd ed. Boca Raton, Florida: CRC Press; 2007.
2. Ikpa PT, Bijvelds MJC, de Jonger HR. Cystic fibrosis: toward personalised therapies. *Int J Biochem Cell Biol* 2014;52:192-200.
3. Westwood T, Henderson B, Ramsay M. Diagnosing cystic fibrosis in South Africa. *S Afr Med J* 2006;96:304-306.
4. Westwood ATR. The prognosis of cystic fibrosis in South Africa: a 33 year study. *J Cyst Fibros* 2008;7(S2):458.
5. Cystic Fibrosis Canada - Annual Report of The Canadian Cystic Fibrosis Registry. 2012 [Internet]. [accessed February 3, 2015]. Available from: http://www.cysticfibrosis.ca/wp-content/uploads/2014/03/Canadian-CF-Registry-English-FINAL-FOR-WEB1.pdf.
6. Costerton JW, Stewart PS, Greenberg EP. Bacterial biofilms: a common cause of persistent infections. *Science* 1999;284:1318-1322.
7. Grosse SD, Boyle CA, Botkin JR, Comeau AM, Kharrazi M, Rosenfeld M, Wilfond BS. Newborn screening for cystic fibrosis: evaluation of benefits and risks and recommendations for state newborn screening programs. *MMWR Recomm Rep* 2004;53:1-36.
8. Cystic Fibrosis Mutation Database. [Internet]. [accessed June 15, 2016]. Available from: http://www.genet.sickkids.on.ca/cftr/HelpPage.html#faq.
9. 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.
10. Schrijver I, Pique L, Graham S, Pearl M, Cherry A, Kharrazi M. The spectrum of *CFTR* variants in non-white cystic fibrosis patients – implications for molecular diagnostic testing. *J Mol Diagn* 2016;18:39-50.

11. Stewart C, Pepper MS. 2015. Cystic fibrosis on the African continent. *Genet Med* 2015;doi:10.1038/gim.2015.157.
12. Wang SS, O'Leary LA, FitzSimmons SC, Khoury MJ. The impact of early cystic fibrosis diagnosis on pulmonary function in children. *J Pediatr* 2002;141:804-810.
13. Grové SS. Fibrocystic disease of the pancreas in the Bantu. *S Afr J Lab Clin Med* 1959;5:113-118.
14. *The molecular genetic epidemiology of cystic fibrosis*. Genoa: World Health Organisation; 2002.
15. van der Ploeg CPB, van den Akker-can M, Vernooij-van Langen AMM, Elvers LH, Gille JJP, Verkerk PH, Dankert-Roelse, JE. Cost-effectiveness of newborn screening for cystic fibrosis determined with real-life data. *J Cyst Fibros* 2015;14:194-202
16. Pérez MM, Luna MC, Pivetta OH, Keyeux G. *CFTR* gene analysis in Latin American CF patients: heterogeneous origin and distribution of mutations across the continent. *J Cyst Fibros* 2007;6:194-208.
17. Alper OM, Wong L-JC, Young S, Pearl M, Graham S, Sherwin J, Nussbaum E, Nielson D, Platzker A, Davies Z, Lieberthal A, Chin T, Shay G, Hardy K, Kharrazi M. Identification of novel and rare mutations in California Hispanic and African American cystic fibrosis patients. *Hum Mutat* 2004;24:353-364.
18. Sugarman EA, Rohlfs EM, Silverman LM, Allitto BA. *CFTR* mutation distribution among U.S. Hispanic and African American individuals: evaluation in cystic fibrosis patient and carrier screening populations. *Genet Med* 2004;6:392-399.
19. Heim RA, Sugarman EA, Allitto BA. Improved detection of cystic fibrosis mutations in the heterogeneous U.S. population using an expanded, pan-ethnic panel. *Genet Med* 2001;3:168-176.
20. Freidman KJ, Leigh MW, Czarnecki P, Feldman GL. Cystic fibrosis transmembrane-conductance regulator mutations among African Americans. *Am J Hum Genet* 1998;62:195-196.

21. Macek Jr. M, Mackova A, Hamosh A, Hilman BC, Selden RF, Lucotte G, Freidman KJ, Knowles MR, Rosenstein BJ, Cutting GR. Identification of common cystic fibrosis mutations in African-Americans with cystic fibrosis increases the detection rate to 75%. *Am J Hum Genet* 1997;60:1122-1127.

22. Cutting GR, Curristin SM, Nash E, Rosenstein BJ, Lerer I, Abeliovich D, Hill A, Graham C. Analysis of four diverse population groups indicates that a subset of cystic fibrosis mutations occur in common among Caucasians. *Am J Hum Genet* 1992;50:1185-1194.

23. White MB, Krueger LJ, Holsclaw Jr. DS, Gerrard BC, Stewart C, Quittell L, Dolganov G, Baranov V, Ivaschenko T, Kapronov NI, Sebastio Gianfranco, Catiglione O, Dean M. Detection of three rare frameshift mutations in the cystic fibrosis gene in an African-American (CF444delA), an Italian (CF2522insC), and a Soviet (CF3821delT). *Genom* 1991;10:266-269.

24. Cutting GR, Kasch LM, Rosenstein BJ, Zielenski J, Tsui L-C, Antonarakis SE, Kazazian Jr. HH. A cluster of cystic fibrosis mutations in the first mucleotide-binding fold of the cystic fibrosis conductance regulator protein. *Nat* 1990a;346:366-369.

25. Cutting GR, Kasch LM, Rosenstein BJ, Tsui L-C, Kazazian Jr. HH, Antonarakis SE. Two patients with cystic fibrosis, nonsense mutations in each cystic fibrosis gene and mild pulmonary disease. *New Engl J Med* 1990b;323:1685-1689.

26. Schwartz KM, Pike-Buchanan LL, Muralidharan K, Redman JB, Wilson JA, Jarvis M, Cura MG, Pratt VM. Identification of cystic fibrosis variants by polymerase chain reaction/oligonucleotide ligation assay. *J Mol Diagn* 2009;11:211-215.

27. Feuillet-Fieux MN, Ferrec M, Gigarel N, et al. Novel *CFTR* mutations in black cystic fibrosis patients. *Clin Genet* 2004;65:284-287.

28. Costa FMM, Santana MA, Lemos ACM, Galvão-Castro B, Acosta AX. Low frequency of the ΔF508 mutation of the CFTR gene in a highly admixed population in Bahia, Brazil. *Hum Biol* 2007;79:293-297.

29. Cabello GMK, Cabello PH, Otsuki K, Gombarovitis ME, Llerena Jr. JC, Fernandes O. Molecular analysis of 23 exons of the *CFTR* gene in Brazilian patients leads to the finding of rare cystic fibrosis mutations. *Hum Biol* 2005;77:125-135.

30. de Araújo FG, Novaes FC, dos Santos NPC, Martins VC, de Souza SM, dos Santos SEB, Ribeiro-dos-Santos AKC. Prevalence of ΔF508, G551D, G542X, and R553X mutations among cystic fibrosis patients in the North of Brazil. *Brazilian J Med Biol Res* 2005;38:11-15.

31. Raskin S, Pereira L, Reis F, Rosario NA, Ludwig N, Valentim L, Phillips III JA, Allitto B, Heim RA, Sugarman EA, Probst CM, Faucz F, Culpi L. High allelic heterogeneity between Afro-Brazilians and Euro-Brazilians impacts cystic fibrosis genetic testing. *Genet Test* 2003;7:213-218.

32. Cabello GMK, Moreira AF, Horovitz D, Correia P, Santa Rosa, A, Llerena Jr. J, Greg J, Grody WW, Degrave WM, Fernandes O, Cabello PH. Cystic fibrosis: low frequency of DF508 mutation in two population samples from Rio de Janeiro, Brazil. *Hum Biol* 1999;71:189-196.

33. da Silva Martins R, Fonseca ACP, Acosta FES, Folescu TW, Higa LYS, Sad IR, de Miranda Chaves CRM, Cabello PH, Cabello GMK. Severe phenotype in an apparent homozygosity caused by a large deletion in the *CFTR* gene: a case report. *BMC Res Notes* 2014;7:583-586.

34. Keyeux G, Rodas C, Bienvenu T, Garavito P, Vidaud D, Sanchez D, Kaplan J-C, Aristizábal G. *CFTR* mutations in patients from Colombia: implications for local and regional molecular diagnosis programs. *Hum Mutat* 2003;22:259-265.

35. Restrepo CM, Pineda L, Rojas-Martínez A, Gutiérrez CA, Morales A, Gómez Y, Villalobos MC, Borjas L, Delgado W, Myers A, Barrera-Saldaña HA. CFTR mutations in three Latin American countries. *Am J Med Genet* 2000;91:277-279.

36. Tishkoff SA, Kidd KK. Implications of biogeography of human populations for "race" and medicine. *Nat Genet* 2004;36:S21-S27.

37. Lim RM, Silver AJ, Silver MJ, Borroto C, Spurrier B, Petrossian TC, Larson JL, Silver LM. Targeted mutation screening panels expose systematic population bias in detection of cystic fibrosis risk. *Genet Med* 2016;18:174-179.

38. Mutesa L, Bours V. Diagnostic challenges of cystic fibrosis in patients of African origin. *J Trop Pediatr* 2009;55:281-286.

39. Sosnay PR, Castellani C, Penland CM, Rommens JM, Lewis M, Raraigh KS, Corey M, Cutting GR. Bias in *CFTR* screening panels. *Genet Med* 2015; doi:10.1038/gim.2015.105

40. Watson MS, Cutting GR, Desnick RJ, Driscoll DA, Klinger K, Mennuti M, Palomaki GE, Popovich BW, Pratt VM, Rohlfs EM, Strom CM, Richards CS, Witt DR, Grody WW. Cystic fibrosis population carrier screening: 2004 revision of American College of Medical Genetics mutation panel. *Genet Med* 2004;6:387-391.

41. Sosnay PR, Siklosi KR, Van Goor F, Kaniecki K, Yu H, Sharma N, Ramalho AS, Amaral MD, Dorfman R, Zielenski J, Masica DL, Karchin R, Millen L, Thomas PJ, Patrinos GP, Corey M, Lewis MH, Rommens JM, Castellani C, Penland CM, Cutting GR. Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet* 2013;45:1160-1167.

42. Hughes EE, Stevens CF, Saavedra-Matiz CA, Tavakoli NP, Krein LM, Parker A, Zhang Z, Maloney B, Vogel B, DeCelie-Germana J, Kier C, Anbar RD, Berdella MN, Comber PG, Dozor AJ, Goetz DM, Guida Jr L, Kattan M, Ting A, Voter KZ, The New York State Cystic Fibrosis Newborn Screening Consortium, van Roey P, Caggana M, Kay DM.

Clinical sensitivity of cystic fibrosis mutation panels in a diverse population. *Hum Mutat* 2016;37:201-208.

43. Baker MW, Atkins AE, Cordovado SK, Hendrix M, Earley MC, Farrell PM. Improving newborn screening for cystic fibrosis using next generation sequencing technology: a technical feasibility study. *Genet Med* 2016;18:231-238.
44. Trujillano D, Ramos MD, González J, Tornador C, Sotillo F, Escaramis G, Ossowski S, Armengol L, Casala T, Estivill X. Next generation diagnostics of cystic fibrosis and *CFTR*-related disorders by targeted multiplex high-coverage resequencing of *CFTR*. *J Med Genet*2013;50:455-462.
45. Lefterova MI, Shen P, Odegaard JI, Fung E, Chiang T, Peng G, Davis RW, Wang W, Kharrazi M, Schrijver I, Scharfe C. Next-generation molecular testing of newborn dried blood spots for cystic fibrosis. *J Mol Diagn* 2016;18:267-282.
46. Bell SC, De Boeck K, Amaral MD. New pharmacological approaches for cystic fibrosis: Promises, progress, pitfalls. *Pharmacol Ther* 2015;145:19-34.
47. Farrell PM, Lai HCJ, Li Z, Kosorok MR, Laxova A, Green CG, Collins J, Hoffman G, Laessig R, Rock MJ, Splaingard ML. Evidence on improved outcomes with early diagnosis of cystic fibrosis through neonatal screening: Enough is enough! *J Pediatr* 2005;147:S30-S36.
48. Macri CN, Pivetta OH, Gentile A, Cafferata EG, Luna MC. Cystic fibrosis in Latin America. In: Bauernfeind A, Marks MI, Strandvik B, editors. Cystic fibrosis pulmonary infections: lessons from around the world. Basel, Switzerland: Birkhäuser; 1996. p. 119-138.
49. *Services for the prevention and management of genetic disorders and birth defects in developing countries.* Geneva: World Health Organisation; 1999.

Table 1: Cystic fibrosis mutations identified in the African Diaspora.

Mutation	# of alleles tested	# of + alleles	Allele Frequency	Nationality
IVS22+1G>A	2	2	100.00	French Guianese[27]
U	1580	892	56.46	Brazilian, American, Haitian, Colombian, Venezuelan
c.3718-2530A>G	2	1	50.00	Haitian[26]
ΔF311	2	1	50.00	American[20]
-12_10del23	2	1	50.00	Brazilian[33]
Genzyme87	92	31	33.70	American[17]
ΔF508	1431	421	29.42	Brazilian[28-32], American[17-22], Colombian[34], Venezuelan[35]
3120+1G>A	886	74	8.35	Brazilian[29,31], American[18,19,21], Colombian[34]
G85E	266	10	3.76	Brazilian[29]
1811+1.6kbA>G	54	2	3.70	Colombian[34]
1342-1G>C	32	1	3.13	American[22]
3662delA	32	1	3.13	American[22]

Mutation	# of alleles tested	# of + alleles	Allele Frequency	Nationality
I1203V	32	1	3.13	American[22]
W1316X	66	2	3.03	American[22]
2307insA	566	13	2.30	American[19]
164+2T>A	92	2	2.17	American[17]
A559T	688	14	2.03	American[17-19, 21, 22, 24]
R1158X	202	4	1.98	American[19]
1323_1324insA	54	1	1.85	Colombian[34]
3500-2A>G	54	1	1.85	Colombian[34]
R553X	668	12	1.80	American[18, 21, 22, 24, 25]
G542X	838	14	1.67	Brazilian[29, 31, 32], American[18, 21, 22, 25], Venezuelan[35]
S549N	216	3	1.39	American[21, 22, 24]
405+3A>C	148	2	1.35	American[21]
S1255X	632	7	1.11	American[18, 19, 21, 22]
Q98X	92	1	1.09	American[17]
S466X	92	1	1.09	American[17]
3791delC	566	6	1.06	American[18,19, 21]
R334W	756	8	1.06	Brazilian[29, 31], American[18, 19, 21]
444delA	568	6	1.06	American[18, 19, 21, 23]
P205S	190	2	1.05	Brazilian[29]

Mutation	# of alleles tested	# of + alleles	Allele Frequency	Nationality
S4X	192	2	1.04	Brazilian[29, 33]
1898+1G>A	202	2	0.99	American[19]
R117H	202	2	0.99	American[19]
R1162X	738	7	0.95	Brazilian[29, 31], American[18, 19], Colombian[34]
Y1092X	320	3	0.94	Brazilian[29, 31], Colombian[34]
711+5G>A	216	2	0.93	American[18]
G330X	566	5	0.88	American[18, 19, 21]
G551D	910	8	0.88	Brazilian[29, 30, 32], American[18, 19, 21]
1717-1G>A	364	3	0.82	American[18, 21]
1812-1G>A	494	4	0.81	Brazilian[31], American[18, 19]
N1303K	270	2	0.74	American[18], Colombian[34]
G480C	566	4	0.71	American[18, 19, 21]
1002-3T>G	148	1	0.68	American[21]
1119delA	148	1	0.68	American[21]
1504delG	148	1	0.68	American[21]
1618T	148	1	0.68	American[21]
2734delGinsAT	148	1	0.68	American[21]

Mutation	# of alleles tested	# of + alleles	Allele Frequency	Nationality
621G>A	148	1	0.68	American[21]
DI507	148	1	0.68	American[21]
R764X	148	1	0.68	American[21]
S364P	148	1	0.68	American[21]
W19C	148	1	0.68	American[21]
Y563D	148	1	0.68	American[21]
S549R	460	3	0.65	Brazilian[29], American[18], Colombian[34]
3905insT	202	1	0.50	American[19]
621+1G>T	202	1	0.50	American[19]
Q493X	202	1	0.50	American[19]
R1066C	202	1	0.50	American[19]
E60X	216	1	0.46	American[18]
R560T	216	1	0.46	American[18]

Mutations highlighted in red have been shown to cause CF; mutations in blue are of varying clinical consequence; mutations in black have no associated functional study; U = unknown mutation. Genzyme87 refers to a commercially available panel. In the publications that used this panel, the individual mutation frequencies were not reported.

Table 2: Mutations identified in Africa and the African Diaspora.

Mutation	Identified in African Country(ies)	Identified in Diaspora Country(ies)
ΔF508	Egypt, Algeria, Tunisia, Libya, Morocco, Namibia, South Africa (48.34%)	Brazil, USA, Colombia, Venezuela (29.42%)
G542X	Tunisia, South Africa (4.87%)	Brazil, USA, Venezuela (1.67%)
N1303K	Egypt, Algeria, Tunisia, Libya, South Africa (4.50%)	USA, Colombia (0.74%)
3120+1G>A	Rwanda, South Africa, Zimbabwe (11.46%)	Brazil, USA, Colombia (8.35%)
G85E	Tunisia (1.11%)	Brazil (3.76%)
R1066C	Tunisia (0.19%)	USA (0.50%)
G551D	South Africa (1.02%)	Brazil, USA (0.88%)
R553X	South Africa (1.00%)	USA (1.80%)
1717-1G>A	South Africa (0.25%)	USA (0.82%)
621+1G>T	South Africa (0.25%)	USA (0.50%)
Q493X	South Africa (0.25%)	USA (0.50%)
R1162X	South Africa (0.25%)	USA (0.95%)
R117H	South Africa (0.25%)	USA (0.99%)
S549N	South Africa (0.25%)	USA (1.39%)
I1203V	Tunisia (1.47%)	USA (3.13%)
R1158X	Tunisia (1.47%)	USA (1.98%)

Mutation	Identified in African Country(ies)	Identified in Diaspora Country(ies)
S549R	Morocco (11.76%)	Brazil, USA, Colombia (0.65%)
1812-1G>A	Algeria (5.00%)	Brazil, USA (0.81%)

Mutations highlighted in red have been shown to cause CF; mutations in blue are of varying clinical consequence; mutations in black have no associated functional study. Numbers in parentheses are the overall allele frequencies.

Data Supplement

Cystic Fibrosis in the African Diaspora

Cheryl Stewart & Michael S. Pepper

Supplementary Table E1: Cystic fibrosis mutations in the African Diaspora by Country.

Mutation	# of alleles tested	Method	# of + alleles	Allele Frequency	Country
ΔF311	2	DGGE then Sanger	1	50.00	USA
ΔF508	615	DGGE then Sanger, ASO for 93 mutations, reverse dot strip hybridisation, Genzyme 87 panel, PCR & sequencing of exons 1, 2, 9-12, 19-23	255	41.46	USA
U	762		283	37.14	USA
Genzyme87	92	Genzyme 87	31	33.70	USA
3120+1G>A	566	ASO for 93 mutations, PCR-RFLP, Genzyme 87	64	11.31	USA
3662delA	32	PCR & sequencing of exons 1, 2, 9-12 & 19-23	1	3.13	USA
1342-1G>C	32	PCR & sequencing of exons 1, 2, 9-12 & 19-23	1	3.13	USA
I1203V	32	PCR & sequencing of exons 1, 2, 9-12 & 19-23	1	3.13	USA
W1316X	66	PCR & sequencing of exons 1, 2, 9-12 & 19-23	2	3.03	USA
2307insA	566	ASO for 93 mutations, PCR-RFLP, Genzyme 87	13	2.30	USA

Mutation	# of alleles tested	Method	# of + alleles	Allele Frequency	Country
164+2T>A	92	TTGE	2	2.17	USA
		Genzyme 87, PCR & sequencing of exons 1, 2, 9-12 & 19-23, ASO for 93 mutations, sequencing			USA
A559T	634	of exons 9-12 & 20-22	13	2.05	
R1158X	202	ASO for 93 mutations	4	1.98	USA
		Reverse dot strip hybridisation, Genzyme, PCR & sequencing of exons 9-12 & 20-22, PCR & sequencing of exons 1, 2, 9-12, 19-23, ASO for			USA
R553X	668	93 mutations	12	1.80	
		PCR & sequencing of exons 1, 2, 9-12, 19-23, reverse dot strip			USA
S549N	216	hybridisation	3	1.39	
405+3A>C	148	ASO	2	1.35	USA
		Genzyme 87, sequencing of exon 11, ASO, PCR &			USA
G542X	430	sequencing of exons 1, 2,	5	1.16	

Mutation	# of alleles tested	Method	# of + alleles	Allele Frequency	Country
		9-12 & 19-23, reverse dot strip hybridisation			
		Genzyme 87, PCR & sequencing of exons 1, 2, 9-12, 19-23, ASO for 93 mutations, PCR & sequencing of exons 20			USA
S1255X	632	& 21	7	1.11	
Q98X	92	TTGE	1	1.09	USA
S466X	92	TTGE	1	1.09	USA
3791delC	566	Genzyme 87, ASO for 93 mutations, sequencing of exon 19	6	1.06	USA
444delA	568	Genzyme 87, SSCP of exons 4, 13 & 19, ASO for 93 mutations	6	1.06	USA
R117H	202	ASO for 93 mutations, ASO for R117H	2	0.99	USA
1898+1G>A	202	ASO for 93 mutations	2	0.99	USA
711+5G>A	216	Genzyme 87	2	0.93	USA

Mutation	# of alleles tested	Method	# of + alleles	Allele Frequency	Country
G330X	566	Genzyme 87, ASO for 93 mutations, sequencing of exon 7	5	0.88	USA
G551D	566	ASO for 93 mutations, reverse dot strip hybridisation, Genzyme 87	5	0.88	USA
1717-1G>A	364	Reverse dot strip hybridisation, Genzyme 87	3	0.82	USA
1812-1G>A	418	ASO for 93 mutations, Genzyme 87	3	0.72	USA
G480C	566	ASO for 93 mutations, Genzyme 87, PCR, chemical mismatch cleavage & sequencing	4	0.71	USA
DI507	148	Reverse dot strip hybridisation	1	0.68	USA
W19C	148	Sequencing of exon 2	1	0.68	USA
621G>A	148	Sequencing of exon 4	1	0.68	USA
1002-3T>G	148	ASO	1	0.68	USA
1119delA	148	Sequencing of exon 7	1	0.68	USA

Mutation	# of alleles tested	Method	# of + alleles	Allele Frequency	Country
S364P	148	RFLP, ASO	1	0.68	USA
1504delG	148	RFLP	1	0.68	USA
Y563D	148	RFLP, ASO	1	0.68	USA
1618T	148	RFLP, ASO	1	0.68	USA
R764X	148	Sequencing of exon 13	1	0.68	USA
2734delGinsAT	148	RFLP	1	0.68	USA
		Genzyme 87, ASO for 93 mutations, reverse dot strip hybridisation,			USA
R334W	566	Genzyme 87	3	0.53	
621+1G>T	202	ASO for 93 mutations	1	0.50	USA
Q493X	202	ASO for 93 mutations	1	0.50	USA
R1066C	202	ASO for 93 mutations	1	0.50	USA
3905insT	202	ASO for 93 mutations	1	0.50	USA
R1162X	418	Genzyme 87, ASO for 93 mutations	2	0.48	USA
E60X	216	Genzyme 87	1	0.46	USA
S549R	216	Genzyme 87	1	0.46	USA
N1303K	216	Genzyme 87	1	0.46	USA
R560T	216	Genzyme 87	1	0.46	USA
3718-2530A>G	2	CF genetic test panels, sequencing	1	50.00	Haiti

Mutation	# of alleles tested	Method	# of + alleles	Allele Frequency	Country
U	2		1	50.00	Haiti
IVS22+1G>A	2	PCR, DGGE	2	100.00	French Guiana
U	708		550	77.68	Brazil
-12_10del23	2	PCR & sequencing	1	50.00	Brazil
ΔF508	708	ΔF508 screening, SSCP	129	18.22	Brazil
G85E	266	70 mutation screen, SSCP, sequencing	10	3.76	Brazil
3120+1G>A	266	70 mutation screen, RFLP	9	3.38	Brazil
R334W	190	70 mutation screen, SSCP, sequencing	5	2.63	Brazil
G542X	354	Inno-Lipa CF2 kit, 70 mutation screen, SSCP, sequencing	7	1.98	Brazil
1812-1G>A	76	70 mutation screen	1	1.32	Brazil
R1162X	266	70 mutation screen, SSCP, sequencing	3	1.13	Brazil
P205S	190	SSCP, sequencing	2	1.05	Brazil
S4X	192	SSCP, sequencing	2	1.04	Brazil
G551D	344	PCR-RFLP	3	0.87	Brazil
Y1092X	266	70 mutation screen	2	0.75	Brazil

Mutation	# of alleles tested	Method	# of + alleles	Allele Frequency	Country
U	54		22	40.74	Colombia
ΔF508	54	Sequencing of exon 11	21	38.89	Colombia
1811+1.6kbA>G	54	Sequencing of intron 12	2	3.70	Colombia
R1162X	54	DGGE & sequencing	2	3.70	Colombia
Y1092X	54	DGGE & sequencing	1	1.85	Colombia
3120+1G>A	54	DGGE & sequencing	1	1.85	Colombia
1323_1324insA	54	DGGE & sequencing	1	1.85	Colombia
S549R	54	DGGE & sequencing	1	1.85	Colombia
A559T	54	DGGE & sequencing	1	1.85	Colombia
N1303K	54	DGGE & sequencing	1	1.85	Colombia
3500-2A>G	54	DGGE & sequencing	1	1.85	Colombia
U	54		36	66.67	Venezuela
ΔF508	54	16 mutation kit	16	29.63	Venezuela
G542X	54	16 mutation kit	2	3.70	Venezuela

Mutations highlighted in red have been shown to cause CF; mutations in blue are of varying clinical consequence; mutations in black have no associated functional study; U = unknown mutation. Genzyme87 refers to a commercially available panel. In the publications that used this panel, the individual mutation frequencies were not reported. Kit = commercial screening diagnostic kit; PCR-RFLP = polymerase chain reaction restriction fragment length polymorphism; DGGE = denaturing gradient gel electrophoresis; SSCP = single strand

conformation polymorphism; ASO = allele specific oligonucleotide hybridisation; TTGE = temporal temperature gradient gel

electrophoresis.