The Cystic Fibrosis Heterozygote Advantage: A Synthesis of Ideas

Kristen M. Beckett *McMaster University*

Abstract: It has been proposed that the high frequency of Cystic Fibrosis (CF) is due to a selective advantage in the heterozygous state. This research supports this proposition and attempts a holistic perspective in reviewing the various models proposed. The pathophysiology, molecular and population genetics and the epidemiology are all demonstrated to be key components when evaluating the carrier advantage of CF.

Various hypotheses are reviewed and critically evaluated. The most recent explanation proposed—that CF affords protection against cholera—is a main focus of this research. While this theory has been supported by molecular *in vivo* experiments, the historical epidemiology of both CF and cholera argue against the acceptance of this model. The conclusion is drawn that while the cholera hypothesis is attractive, further research will be necessary before it can either be accepted or rejected as the explanation for high frequencies of CF.

Résumé: Il a été envisagé que la haute fréquence de mucoviscidose est due à un avantage sélectif au niveau de l'état hétérozygote. Cette recherche appuie cette assertion et tente de donner une perspective holistique en faisant l'état des divers modèles proposés. Il a été démontré que la pathophysiologie, la génétique moléculaire et démographique ainsi que l'épidémiologie sont tous des éléments essentiels lorsqu'on évalue l'avantage qu'a le porteur de la mucoviscidose.

Diverses hypothèses ont eté passées en revue et évaluées de façon critique. L'explication la plus récente qui ait été proposée—la mucoviscidose offrant une protection contre le choléra—constitue le point central de cette recherche. Alors que cette théorie a été corroborée par des expériences moleculaires *in vivo*, l'épidémiologie historique de la mucoviscidose et celle du choléra argumentent contre l'acceptation de ce modèle. Il a été conclu que, bien que l'hypothèse du choléra soit séduisante, des recherches supplémentaires s'imposent avant de pouvoir la rejeter ou l'accepter comme explication de la haute fréquence de la mucoviscidose.

Introduction

Cystic Fibrosis is the most common autosomal recessive trait among Caucasians. The term "Caucasian" will be used in reference to those individuals of European descent. In Caucasian populations approximately 4 to 5 percent of individuals are heterozygous for CF, i.e., they carry a mutated copy of the cystic fibrosis gene. This results in approximately one in every 2 000 to 2 500 live births being homozygous for CF, i.e., expressing the disorder in full (Gabriel et al., 1994: 107). The frequency of a genetic disorder due to random mutation never exceeds 1 percent. The cystic fibrosis (CF) mutation in the homozygous state is lethal before reproduction. Because homozygous CF individuals have rarely survived to reproduce, the mutations they carry have not been passed on to the next generation. The deleterious nature of the disease should therefore maintain the mutation at a low frequency in the population. However, CF is being maintained at very high frequencies in Caucasian populations. This has led to speculation about a heterozygote (carrier) advantage for the cystic fibrosis mutation. If an additional pressure were selecting for the mutation then the disease would be maintained as a balanced polymorphism, which we may define as a state of equilibrium in which gene frequencies are maintained by a balance between mutation and selection (Saunders, 1994).

The prototype for a genetic disease with a heterozygote advantage is, of course, sickle cell anemia. Twenty percent of the population in malarial areas carry the sickle cell allele with up to 40 percent carrying it in some African populations. The explanation for the high frequency is that people who are heterozygous for sickle cell anemia are able to survive and reproduce in areas where malaria is endemic. This is a result of the sickle cell trait in the heterozygous state affording protection against the malaria parasite which matures in the red blood cells (Allison, 1954: 291).This is also the explana-

tion given for the high frequencies of G-6-PD and thalassemia in these areas.

The high frequency of CF coupled with the discovery of the location of the gene which causes the disease in 1989 (Rommens et al., 1989), has generated a great deal of interest in this disease among researchers, as well as in the popular press (Glausiusz, 1995: 30-31). The research presented in this paper attempts a holistic synthesis and review of the many disciplines which study this disease. When trying to discover what might be a possible heterozygote advantage for cystic fibrosis, a broad understanding of the disorder is necessary. Researchers must consider how the disease manifests itself clinically, how it works on a molecular level, the historical epidemiology of it and how it is distributed in populations around the world today. Understanding these aspects of the disease provides a strong basis from which inferences may be drawn.

Researchers seeking the heterozygote advantage of CF have come up with different hypotheses. These hypotheses include: the theory that CF provides protection against tuberculosis (Meindl, 1987); the idea that carriers of CF have increased fertility (Knudson and Wayne, 1967); and the theory that CF affords protection from cholera and *E.coli* (Chao et al., 1993; Hansson, 1988). If one reviews these hypotheses with all aspects of the disease in mind, it becomes apparent that there is more than one factor influencing selection for this gene. It is argued here that while the molecular data, which support the hypothesis that the CF gene may be selected for partly because it offers protection against cholera and/or *E. coli* bacteria, seem compelling, they are not enough on their own to support this theory. Before this or any other hypotheses can be accepted as an explanation for the high frequency of CF, the historical epidemiology of infectious disease in Europe must be calculated. In addition, the epidemiology of both CF and cholera must be understood and accounted for in both modern and historic times.

Pathophysiology of Cystic Fibrosis

General

The basic defect in CF has been associated with decreased chloride ion conductance across the apical membrane of the epithelial cells (Riordan et al., 1989: 106). This defect in the CFTR gene results in an increased viscosity of mucus gland secretion which causes many internal obstructions. Instead of forming a thin, freely flowing secretion, the mucus gland produces a thickened mucoprotein that accumulates in and dilates the glands. The small passages in organs, such as the pancreas and bronchioles, become obstructed as the secretions coagulate to form concretions in the glands and ducts. (Whalley, 1991: 1470)

Life expectancy for those who suffer from CF today is much longer then it has been. Due to advancing medical treatment life expectancy for those with CF has increased from one year in 1940, to 5.5 years in 1960, to 24 years in 1990 (Phillips, 1991: 937-939).

Respiratory Tract

In the normal respiratory tract, mucus production by goblet cells lubricates airways and entraps foreign particles (ibid.: 937). Owing to the increase of bronchial mucus and its thickness among individuals with CF, there is a greater resistance to ciliary action, a slower flow of mucus and an inability to cough up the mucus, all of which contribute to the formation of an obstruction (Whalley, 1991: 1470). The mucus obstruction reduces oxygen and carbon dioxide exchange causing varying degrees of hypoxia (a reduction of oxygen to tissue below physiological levels), hypercapnia (excess carbon dioxide in the blood) and acidosis (accumulation of acid, or depletion of alkaline residue in the blood and body tissue). Most importantly, the retained mucus serves as an excellent medium for bacterial growth. After repeated infections the bacteria develop resistance to multiple drugs and become impossible to eradicate. Eventually, bacterial infection leads to the destruction of the lung tissue (ibid.). Progressive lung involvement leads to compression of pulmonary blood vessels, and progressive lung dysfunction frequently leads to pulmonary hypertension, cor pulmonale, respiratory failure and death (ibid.).

Gastrointestinal Tract

In the pancreas of CF sufferers, the thick secretions block the ducts which leads to cystic dilation of the acici (the small lobes of the gland), which undergo degeneration and progressive diffuse fibrosis. This process prevents essential pancreatic enzymes from reaching the duodenum (the first portion of the small intestine). This causes significant difficulty in the digestion and absorption of nutrients, especially fats, proteins and to a lesser degree carbohydrates (Collins, 1992: 774). The incidence of diabetes mellitus is greater in CF children which is probably related to these changes in pancreatic architecture (Whalley, 1992: 1471).

Reproductive System

Both male and female reproductive systems are affected when an individual is homozygous for CF. Females have reduced fertility due to a thickening of cervical mucus which presents a barrier for sperm penetration. Males are sterile with few exceptions. Virtually all males are azoospermic (lack of spermatozoa in the semen). This is due to atrophy or obstruction of the epididymis, vas deferens and seminal vesicles (Tizzano et al., 1994: 906). The heterozygote may also have reduced fertility. Researchers in fertility clinics seem to have found evidence that the male heterozygote and compound heterozygote have similar symptoms to those afflicted with CF (Liu et al., 1994: 1859-1860).

Integumentary Systems

Individuals with CF display abnormally high sodium chloride concentrations in their sweat. They often taste salty to kiss (Whalley, 1991: 1472). In fact, the distinctive salty taste of children with CF was initially used in diagnosing this disease.

The defective chloride channels in sweat glands prevent reabsorption of sodium and chloride. This leaves the individual at risk for abnormal salt loss, dehydration and hypochloremic and hyponatremic alkalosis (decreased chloride and sodium in the blood) during hyperthermic conditions (ibid.). This would be extremely dangerous to CF individuals living in tropical areas. A study of CF children living in northern Spain showed that during the summer and early autumn CF children of all ages had acute attacks of hypochloraemia and metabolic alkalosis (a disturbance in which the acid-base status of the body shifts toward the alkaline side [Saunders, 1994]). Only the infants developed chronic cases of the same, and failed to thrive (Sojo et al., 1994: 825-826).

Genetic Methods of Study

The year 1989 was a breakthrough year in CF research. Researchers were able to locate the gene responsible for CF and investigate its function. Molecular cloning experiments and chromosome walking and jumping experiments permitted the isolation of a large contiguous segment of DNA spanning at least four transcribed sequences from a region thought to contain the CF locus (Riordan et al., 1989: 1066). Through genetic analysis and DNA sequencing, one of the four candidates was shown to correspond to the position of the CF gene locus (Tsui et al., 1991b: 11).

After finding the location at which the mutation occurred, Riordan et al. (1989) next pursued the detection of the mutation. Comparison between DNA sequences derived from CF and unaffected individuals was next conducted. The most striking difference found was a 3 base pair (bp) deletion which resulted in the loss of a phenylalanine residue, position 508 in the prediscided CF polypeptide. This mutation was observed in 68 percent of the chromosomes studied by this group. Direct sequencing of genomic DNA amplified by polymerase chain reactions revealed many additional mutant alleles (ibid.). These different mutations were found to mostly manifest themselves internally in an identical fashion to the major mutation, ΔF508; however, some missense mutations produced different clinical manifestations (see molecular genetics section).

Most of these laboratory methods continue to be used by those studying the molecular genetics of CF and by those who are trying to isolate additional mutations.

Mouse Models

Following the identification of the CF gene, researchers have continued experimenting in the hope that they may achieve a superior understanding the function of the gene and how it results in the various clinical manifestations. The use of mouse models has become a popular method to study the CF gene. Tsui et al. (1991b) describe two methods of study using mouse models. The first is a gene knock out experiment where a biochemically selectable marker is introduced into the coding region of the mouse gene via homologous recombination. The second method is creating a mouse model with the identical deletion to the major human mutation, ΔF508 (ibid.: 15).

Some claim success with these mouse models (Dori et al., 1992: 212); others believe that the typical 30-40 day survival rate of the mice is not sufficient time for symptoms to extend into the reproductive, pancreatic, hepatobiliary and respiratory systems (Collins and Wilson, 1992: 708).

Population Genetics and Epidemiology as a Method of Study

Certain researchers see value in examining the population distributions of CF (Beaudet et al., 1991: 53; Devoto, 1991: 63; Serre, 1991: 55). The geographic distribution of the ΔF508 and non-ΔF508 mutations and their associations with haplotypes are all being examined (see Serre et al., 1990 and Serre, 1991 for discussion).

This method of study is helping to trace the origins of the disease, discovering what populations are most affected, and trying to discover why it may be being selected for from a historical and geographical perspective.

Molecular Genetics of the Cystic Fibrosis Gene

Cystic fibrosis is caused by a single defective gene on chromosome 7 that codes for a 1480 amino acid protein named cystic fibrosis transmembrane conductance regulator or CFTR (Wine et al., 1991: 253). CFTR is a regulated, low-conductance chloride channel (Riordan, 1993: 615), the defect is the absence of the CFTR chloride channel from the apical membrane (ibid.: 622).

The major mutation which causes the defect is a 3bp (base pair) deletion in exon 10 which results in the loss of a phenylalanine residue, position 508, at the centre of the first putative ATP-binding site (Tsui et al., 1991b: 10). This deletion is known as ΔF508. Research has shown that up to 75 percent of CF chromosomes carry this mutation (Klinger et al., 1991: 39), and thus it has come to be considered the major mutation. Over 400 additional mutations have been identified since the discovery of the ΔF508 deletion in 1989 (Morral et al., 1994: 890). These additional mutations include various point mutations which can result in either severe or mild clinical manifestations (Kristidis et al., 1992: 1182). All of the additional mutations appear to be rare and generally associated with specific populations.

Functions of the CF Gene

In its most basic sense the function of the CF gene is to regulate chloride channels. A mutation to the gene results in defective chloride channels. CFTR is a chloride channel which is regulated by cyclic AMP—dependent phosphorylation and by intracellular ATP (Collins, 1992: 776). Mutations in CFTR cause CF, partly through the loss of cyclic AMP (cAMP) regulated chloride permeability from the plasma membrane or affected epithelia (Denning et al., 1992: 761). Generally speaking, CFTR controls chloride secretions in the epithelia of the pulmonary tract, gastrointestinal tract and reproductive organs. Studies in biosynthesis and localization of CFTR ΔF508 indicate that the mutant protein is not processed correctly and is not delivered to the plasma membrane (ibid.). CFTR must reach the cell surface to function. Normal CFTR assembles in the endoplasmic reticulum (ER) then travels via the golgi complex to the cell surface, faulty CFTR gets trapped in the ER (Armstrong, 1992: 709). On this basis some argue that the problem lies not so much in what the mutant protein cannot do as with where it ends up in the cell (ibid.).

Population Genetics and History of CF

Cystic fibrosis is believed to have spread throughout Europe with the migration of Neolithic farmers. This migration began from the Middle East and progressed towards the north and northwest of Europe (Devoto, 1991: 70). This hypothesis has recently been supported by archaeological data and biochemical typing (ibid.).

Since the presumed introduction of this disease to Europe it has spread from the continent to European colonies, including the Americas (see Table 1). The highest known estimates for the frequencies of CF today are in England and France. From these two locations CF frequencies decline in all geographic directions in Europe. However, CF frequencies still remain very high throughout Europe. Among Caucasians living in the United States, Canada, Australia and New Zealand CF frequencies are also high. It is estimated that between 1 in 2 500 to 3 000 births in these areas are homozygous for CF (De Braekleer and Daigneault, 1992: 167-168).

Table 1
Estimated CF Gene Carrier Frequencies in Various Populations Using the Hardy-Weinberg Formula

Geographic Location	Estimated Carrier Frequency
Brittany	1:9.7
Ohio (U.S. Amish)	1:12.5
South West Africa (Dutch)	1:13
United States	1:23
Australia	1:23
England	1:25
Czechoslovakia	1:26
France	1:22
Germany	1:29
Pakistani (in England)	1:50
African-Americans	1:82
Oriental (in Hawaii)	1:150

Source: Rodman and Zamudio, 1991: 255.

CF Mutations

The ΔF508 mutation accounts for 70-80 percent of the CF mutations in western Europe (British Isles, France, Belgium, West Germany and Netherlands). The ΔF508 mutation varies between populations and regions. It reaches a maximum of 87 percent in Danish populations (ibid.: 170), while it accounts for only 30 percent of CF mutations in Ashkenazi Jews (Beaudet et al., 1991: 53; Cutting et al., 1992). Its frequency is highest in northern Europe with lower frequencies in southern Europe (Serre, 1991: 55). The mutation is thought to have arisen

on a B haplotype around 5 000 years ago (Serre et al., 1990). By the Middle Ages the mutation is thought to have been well established in European populations (Rodman and Zamudio, 1991: 254).

Over 400 additional mutations have been observed since the discovery of the major mutation in 1989 (Morral et al., 1994: 890). The frequencies of these mutations vary between populations, and some appear to be unique to specific populations (see Table 2).

Table 2
Mutations Shared by Populations and Mutations Unique to Specific Populations

United States Caucasians	African Americans	Northern Irish	Israeli
ΔF508	ΔF508	ΔF508	ΔF508
G542X	G542X	G542X	G542X
6551D	—	6551D	—
W1282X[a]	—	—	—
N1303K	—		
N1303K	N1303K		
A455E	—	—	—
V520F	—	—	—
Y563N	—	—	—
E1371X	—	—	—
—	—	R560T	—
—	—	G3849⇒A	—
—	1342⇒1G⇒C	—	—
—	S539N	—	—
—	RS53X	—	—
—	AS59T	—	—
—	S1255X	—	—
—	W1316X	—	—

a Only observed in individuals of Ashkenazi origins
Source: Modified from Cutting et al., 1992: 1185-1189.

The W1282X mutation seems to be only present in those of Ashkenazi origins, and in this group it is responsible for over 50 percent of CF mutations (Cutting et al., 1992: 1199). This could be due to founder effect. Further investigation by Sereth et al. (1993) found that 96 percent of CF chromosomes in Ashkenazi Jews are due to only six mutations; W1282X, ΔF508, C542X, N1303K, 1717⇒G⇒A and 3849 + 10KbC⇒T (Sereth et al., 1993: 294).

In examining Table 2, it is evident that the distribution of mutations clearly differed between the African American populations and the three groups of Caucasian origins. Two thirds of the mutations found in the African American populations appear to be unique to that group (Cutting et al., 1992: 1190). This then means that gene flow alone does not account for the occurrence of CF in African American populations (ibid.).

French Canadians living in Quebec have also been found to have high frequencies of non-ΔF508 mutations, especially 621 + IG–T, A455E and 711 + IG–T (ibid.: 1192). Founder effect or genetic drift are assumed to be responsible for this. Researchers are not surprised by the ethnic variations they have found. A number of other autosomal recessive disorders such as Beta-thalassemia, Ornithine aminotransference deficiency and Tay-Sachs have all shown the same pattern (ibid.).

The type of mutation present in an individual seems to play a role in the clinical manifestations of the disease (Dean et al., 1991: 47). The ΔF508 mutation (a deletion mutation) is associated with a severe form of the disease including pancreatic insufficiency (PI) and an early age of presentation. It is interesting that most mutations which cause severe forms of CF occur at frequencies greater than 1 percent, while other mutations including all of those which result in mild clinical cases, occur at frequencies less than 1 percent (Kristidis et al., 1992: 1182; Morral et al., 1994: 890). This could mean that the latter are random mutations and irrelevant to the investigation of a possible heterozygote advantage.

Investigating the Carrier Advantage of CF

Fertility Hypothesis

The fertility hypothesis proposes that the reason for the high frequency of CF is not because of a heterozygote advantage, but because male carriers have increased fertility and produce more progeny (Knudson and Wayne, 1967). This hypothesis is based on the researchers' observation that CF children seem to have more siblings than non-CF children in the same population. However, recent research using the Utah Genealogical data base presents evidence that CF heterozygotes have normal fertility. The principle flaw in the previous analysis by Knudson and Wayne appears to be "ascertainment bias," which means that larger families are more likely to provide the index cases for recessive diseases irrespective of fertility (Jorde and Lathrop, 1988: 808).

Recent studies on infertility in men may shed light on the reproductive phenotype of CF heterozygotes. It has been shown that men with congenital bilateral absence of the vas deferens (CBAVD) have a higher than normal frequency of the ΔF508 mutation. This seems to confirm a link between CF and CBAVD. Further studies concluded that CBAVD is a genital form of CF in otherwise normal healthy men with no other manifestations of CF. In this study these men were compound heterozygotes or homozygotes for CF (Liu et al., 1994: 1858). In another study of only two men the researchers found they only carried the ΔF508 mutation, but still had

CBAVD. However, it is possible they were compound heterozygotes for a less common mutation which was undetected (ibid.: 1859-1860).

The theory of increased fertility for the CF heterozygote holds little credibility today. It has been demonstrated that in the intestinal tract CF heterozygotes express partial manifestations of the disease (Gabriel et al., 1994: 109). If this can be demonstrated to be true for the reproductive tract, then heterozygotes should have decreased fertility since CF homozygotes are most often infertile. New studies being conducted by fertility clinics should help expand our understanding of this. If it is found that heterozygotes do not have any symptoms in the reproductive tract then other factors remain to be considered. Religious orientation and attitudes towards contraception must be considered in order to understand beliefs regarding desired family size. Additionally, an interview process with the families may reveal that the parents of children with CF have more children in hopes of having a child that does not have CF. It can be concluded that the theory that CF in the heterozygote state increases fertility needs much more data to support its claims.

Tuberculosis Hypothesis

In 1987, Richard S. Meindl published his hypothesis stating that the selective advantage for the CF heterozygote may be resistance to tuberculosis (TB). Meindl's hypothesis was concerned with TB of the lungs, or pulmonary TB, caused by *Mycobacterium tuberculosis*. The symptoms of pulmonary TB are weight loss, lassitude and fatigue, night sweats and wasting with purulent sputum, hemoptysis and chest pain (Saunders, 1994).

Meindl's research focused on the epidemiology of CF, the historical distribution of TB, pathophysiology and relations to excess mucopolysacchride (MPS) secretion (Meindl, 1987: 40). The historical distribution of TB and the epidemiology of CF certainly seem to coincide, but there is a dearth of published, experimental data to sustain the molecular side of his argument. His hypothesis can be simplified as follows: when an individual is exposed to TB the tissues inflame and MPS is synthesized as a response. In a normal individual the MPS becomes thick and creates a blockage. The TB would presumably thrive in the mucus buildup. In the CF individual, and heterozygote, Meindl believes MPS production is at least doubled, which at first thought would seem to be a disadvantage. However, when the accumulation of fluids swells to 30-50 percent the MPS breaks up and free-fluid spaces appear. This, Meindl believes, would allow the CF heterozygote resistance to pulmonary TB.

The hypothesis lacks substance for some of its claims, and further research on other assertions is warranted. The idea that CF individuals have increased MPS seems to be based on a 1968 study conducted on only four CF children by Matalon and Dorfman. Moreover, there is no reference suggesting how Meindl determined the phenotype of the heterozygotes. Contradictory research suggests that CF patients have an impaired antibody response that may predispose them to persistent endobronchial infection (Moss et al., 1987: 708). The most compelling evidence Meindl offers is the historical data of the distribution of TB and the epidemiological studies on CF, but this too is a weak ecological correlation. He notes that TB was the leading cause of death in adults at the turn of this century, and "was a major factor in child deaths" (Meindl, 1987: 40). Clearly, TB was not the only disease which could have influenced selection and there were others, perhaps those affecting primarily children, that could also be contenders for the selective advantage. With these caveats, it would be premature to disregard the TB hypothesis entirely (see Discussion).

Cholera Hypothesis

It has recently been proposed that wild-type CFTR (normal CFTR) controls chloride secretions in the small intestine (Chao et al., 1994). The CFTR controlling these secretions is activated by guanylin and cyclic GMP which in turn are activated by a common subtype of a small molecular weight, heat stable enterotoxin produced by *Escherichia coli* (*E. coli*) bacteria, which causes diarrhea and vomiting. This enterotoxin is referred to as STa. When a person is infected with *E. coli* they stimulate the above described repose in that individual and cause severe diarrhea and vomiting which leads to dehydration and in many cases death. It has been demonstrated *in vivo* that there is failure to secrete chloride in response to these agents in the CF intestine, following exposure to heat-stable *E. coli* (Goldstein et al., 1991: 260).

These agents, STa, guanylin and cyclic GMP (cGMP), induce secretion only in cells expressing wild-type CFTR, since mutated CFTR results in defective chloride channels. Chloride secretion is decreased when less wild-type CFTR is expressed by the cells and more ΔF508 is expressed (Chao et al., 1994: 1068). Therefore, it may be that individuals exposed to enterotoxins who only have wild-type CFTR will secrete chloride in response to intestinal bacteria, where individuals with some or all mutated CFTR will not secrete chloride as a response to exposure to certain bacteria. Therefore with fluid secretion decreased in those individuals heterozygous and homozygous for a major CF mutation, it is pos-

sible that these individuals would not suffer as much dehydration as a result of the enterotoxin. This would have been a notable advantage for those living in Europe for at least the last 200 years (when cholera was endemic there), especially infants who are exceptionally susceptible to death from diarrhea.

In addition to the cGMP activated STa induced chloride secretion, it has been reported that chloride secretion in the small intestine can also be activated by cyclic AMP (cAMP) in response to other toxins such as *E. coli LT* (a large molecular weight, heat labile enterotoxin produced by *E. coli*), and cholera toxin (Baxter et al., 1988; Bijman and DeJonge, 1988; Field and Semrad, 1993: 633; Hansson, 1988; Rodman and Zamudio, 1991). Studies in this area concluded that since the pathophysiology of CF is that of defective cAMP-dependent chloride secretion across the apical membrane of secretory epithelial cells, then, secretory diarrheas mediated by toxins such as *E. coli LT* and cholera toxin, which increase cellular cAMP, meet the criteria for a selective advantage for the CF mutation (Chao et al., 1994: 1071). This research was further confirmed through the use of CF mice. In 1994, Gabriel et al. used CF mice to demonstrate that fluid and chloride secretion in response to cholera toxin varies directly with the number of CFTR alleles in each mouse. Thus, in the heterozygote with only 50 percent active CFTR alleles, the chloride secretions in response to cholera toxin were reduced by half (Gabriel et al., 1994: 109).

Cholera is caused by a potent enterotoxin elaborated by *V. cholera* in the small intestine where it acts on epithelial cells to cause secretions of large quantities of isotonic fluid from the mucosal surface. It is marked by severe, painless watery diarrhea resulting in massive gastrointestinal fluid loss and saline depletion, vomiting and cramps. It is spread by feces and contaminated water and food (Saunders, 1994). The CF heterozygote would have less intestinal chloride secretions due to a smaller number of susceptible chloride channels, which would result in less diarrhea and presumably lower death rates from cholera (Rodman and Zamudio, 1991: 255).

Today cholera can be treated with oral rehydration therapy and/or antibiotics and is usually curable. However, cholera is known to have reached epidemic proportions in Europe and North America in the recent past and could result in death in hours. The largest obstacle to the acceptance of the cholera hypothesis is that cholera is believed to have first appeared in Europe only as recently as 1817 (Greenwood, 1977: 165). However, while this appears to be the case for "Indian Cholera" or Asiatic Cholera, it is possible that other strains of cholera or cholera-like illnesses infected Europe much earlier. Cholera is a crowd disease associated with poor sewage disposal in villages, towns and cities. Based on the nature and habitat of the bacteria, one could argue that it has been infecting humans as long as people have lived in city centres. Cases of "deadly" cholera were described in the Hippocratic collection as occurring in Alkmar in 1548, Nimes in 1645, London in 1669 and 1676 and in Vienna in 1786 (ibid.). However, it would be a very difficult task to firmly establish cholera in Europe prior to the 1800s (see Discussion).

Cholera was endemic in Europe and the Americas as recently as the last century (Bilson, 1980: 1). In England and Wales in 1831, there was a 26.3 per 1 000 morbidity rate due to cholera, and there was a 33.8 per 1 000 rate in Scotland (Durey, 1979: 39). In Canada, between June 9 and July 21, 1832, 1 615 people died in Lower Canada from cholera (ibid.: 179). In the same year 5 432 deaths caused by cholera were reported in England and 800 in Belgium (Molnar, 1992: 326). In France, cholera was even more serious. Some 18 000 deaths in Paris in 1832 were attributed to cholera (Delaporte, 1986: 5). In 1834, 258 deaths in Lower Canada and 555 deaths in Upper Canada were caused by cholera (Bilson, 1980: 181). After 1834, there were two more major outbreaks in Canada in 1849 and 1854, two minor outbreaks in 1851 and 1852, with local incidents common from 1866-1871. In 1866, 50 000 died from cholera in the United States of America (Molnar, 1992: 326) and in 1873 there was another major epidemic (Bilson, 1980: 114).

It is important to note that these outbreaks occurred only within the last 121-162 years. As only four to five generations have passed since cholera was endemic, frequencies of CF would not have had time to decline. The incidence of CF could very well be on the decline now, since cholera can be treated. However, not enough time has passed to significantly lower the frequencies. Also, because we have no records of the frequencies of CF until recently, it cannot be known if CF frequencies were significantly higher between 121 and 162 years ago.

Discussion

The frequency at which the ΔF508 mutation is occurring among Caucasians is much too high to be ascribed to any genetic process other than selection. Any mutations occurring at a frequency above 1 percent in a population must be seriously considered for a possible selective advantage. With CF the situation is complex. The clinical manifestations are extremely variable. There seems to be no consistent pattern even among those homozygous

for the major and more serious mutation, ΔF508. Some patients have pancreatic problems, some have respiratory problems, and some have a combination of both in varying degrees. The additional mutations further complicate the search for the heterozygote advantage, because added to the variation of clinical manifestations within the major mutation is a variation of clinical symptoms between mutations. This complicates the search for a heterozygote advantage because the advantage may lie in any of the clinically affected areas. Alternatively, it may be selected for because there is a selective advantage that falls in the same section of chromosome as the CF gene.

In reviewing the hypotheses put forward to explain the high frequency of CF in Caucasians, it is apparent that more investigations and research are necessary. The fertility hypothesis and TB hypothesis have both been criticized here on the basis of insufficient evidence. Likewise, while the cholera hypothesis is supported by more laboratory experiments than the other theories, it too warrants further investigations—-especially in the realm of epidemiology.

The cholera hypothesis is supported mainly by laboratory studies. *In vivo* studies in mice have demonstrated that individuals heterozygous for CF would presumably suffer less and have reduced mortality from exposure to cholera and *E. coli,* due to a decrease in the amount of active chloride channels in the individual (Gabriel et al., 1994). However, this evidence alone is insufficient.

Before this hypothesis can be accepted two major concerns must be addressed. First and foremost is the proposal that cholera was endemic in Europe prior to the 1800s. This proposal must be withdrawn until there is sufficient evidence to support it. In order to suggest that cholera was present in northern Europe prior to the 1800s in proportions great enough to induce a selective advantage, an intensive study of reconstituted families from parish records, and other sources of information on fertility and cause of death, would have to be undertaken. Even if this could be accomplished, and if data on families with and without deaths from cholera could be compiled through generations, it is unlikely that medical diagnosis prior to this century was accurate. The best one could hope to establish would be the presence of cholera-like illness.

If it cannot be demonstrated that cholera was present in Europe before the 1800s, the hypothesis that cholera is the selective advantage for CF weakens. In order for a disease to increase the frequency of a genetic trait, when no other pressures are present, several generations of selection are required. However, since the molecular data supports the notion that CF affords protection against cholera, other possibilities should be considered. One possibility is the idea of pre-adaptation. Pre-adaptation can be defined as a characteristic useful for conditions in which the organism does not yet live (Poirier, 1993: 335). If the idea of pre-adaptation is accepted then two possibilities concerning the cholera hypothesis emerge. First, CF was already established in Europe prior to the first cholera epidemic, maintained by a different and unknown source. The mutation acted as a pre-adaptation during the cholera epidemics, consequently further increasing the frequency of CF rapidly.

Alternatively, the CF gene may lie on a highly mutable region of the chromosome which would allow for speedy selection once cholera was introduced. In the second case, the high mutability would be the pre-adaptation. The second scenario seems to be supported by the fact that while over 400 mutations have been found to cause CF, certain populations seem to have one mutation which occurs at greater frequencies. In northern Europeans it is the ΔF508; in French-Canadian Quebecois it is 621 + IG–T, A455E and 711 + IG–T; and in those of Ashkenazi origins it is W1282X (Cutting et al., 1992: 1192). However, if CF does lie on a highly mutable region of the chromosome, high frequencies of this mutation should be found in all populations around the world.

The second problem with the cholera hypothesis which needs to be addressed is the population genetics of CF with regards to the epidemiology of cholera. CF is believed to have its origins in the Middle East, and to have been brought with the Neolithic farmers through southern Europe, north and northwest into northern Europe where it would become more prevalent (Devoto, 1991: 70). Cholera has been or is presently endemic in all of these areas. In the last century cholera was endemic throughout Europe, and today it is endemic in the Middle East and parts of southern Europe. The cholera hypothesis does not explain why CF is only polymorphic in Caucasian populations and not in Asians and Africans who have both suffered a history of greater exposure to cholera than other populations. Likewise, it fails to explain why CF is more prevalent in northern Europe as opposed to southern Europe where cholera remains a present-day threat.

Climate may be the key to understanding the epidemiology of CF. The southern areas of the world exhibit much lower frequencies of CF than areas in the northern hemisphere. It may be that the southern climates exhibit lower frequencies of CF because individuals with the disease would be at some disadvantage in a hot climate. It is likely that CF would be selected against in the southern

climates because the defective chloride channels in the sweat glands prevent reabsorption of sodium and chloride. In hyperthermic conditions this would leave the individual with CF at a greater risk than a person without CF for decreased sodium and chloride in the blood and dehydration. More simply, if an individual with CF were to be placed in an environment that increased sweating they would lose more salt then the average individual, and, in the absence of modern medical intervention, such salt was not easily replaced. It must be noted that CF heterozygotes as well as homozygotes suffer an increased risk of dehydration when exposed to excessive heat. Behm et al. (1987) found that, when CF hetrozygotes were artificially stimulated to sweat, their sweat secretions were reduced by 65 percent compared with individuals who did not carry a copy of any CF mutation (cited in Rodman and Zamudio, 1991: 257). Furthermore, Smith et al. (1995: 579-580) report the case of an individual heterozygous for both ΔF508 and R117H who succumbed to heat exhaustion when increased sweating was induced. This type of chloride imbalance that affects the individual's cooling system would greatly outweigh any advantage the CF heterozygote may have. In areas such as Australia and New Zealand the Caucasians represent a recent migration from northern climates, so frequencies could be expected to be on the decline. While this explanation may explain the low frequencies of CF in cholera endemic areas in the present day, further research will be necessary to test the idea.

In northern climates CF heterozygotes would not have this disadvantage. It is also possible that they may have an additional selective advantage. As Meindl hypothesized, tuberculosis, which is prevalent in northern climates, may have acted as an additional selective pressure. As CF alters the molecular architecture of the respiratory tract, perhaps it also causes some changes in the local endobronchial environment which are unfavourable for the growth of tuberculosis (Rodman and Zamudio, 1991: 254).

Because cholera and other diarrheal diseases can now be medically treated, it is possible that CF frequencies are on the decline, although this cannot be known for certain. Historical diagnosis is one of the obstacles to gaining a true understanding of the prevalence of specific diseases such as cholera, *E. coli* and CF. Cystic fibrosis was not diagnosed until about 60 years ago (Lloyd-Still, 1983: 1). Prior to that time, CF deaths would have occurred before the age of one and would not be recognized as a genetic disorder. Similarly, for the diarrheal diseases in early records, cause of death would often be no more descriptive than "diarrhea." Infections with *E. coli* bacteria or cholera toxin produces symptoms which closely parallel each other as well as most other diarrheal disease. As a result they were likely diagnosed simply as diarrhea by earlier doctors. This serious impairs our ability to truly understand the historical epidemiology of CF, cholera and *E. coli* in Europe prior to the second half of the 20th century.

Conclusion

The study of cystic fibrosis is becoming increasingly more complex. This research has attempted to compile data from the different fields of study in order to understand what force may be maintaining this deleterious gene at such a high frequency. This holistic approach has incorporated evidence from biochemistry, population and molecular genetics, epidemiology, bacteriology, archaeology and medical anthropology. From this broad-based research it can be concluded that it is likely that no single force is maintaining CF at a carrier frequency of 5 percent in the Caucasian population. The conclusion of molecular data, that CF provides protection against cholera and *E. coli* in the heterozygous state, is not enough on its own to explain the population distribution of the disease. Further study of the population genetics and the pathophysiology of CF are necessary to explain why the disease would be selected against in the tropics where cholera is the greatest threat today.

The various explanations provided for the heterozygote advantage of CF are all speculative and none have been confirmed with the certainty that the sickle cell-malaria scenario has been. In fact, many investigations into selective advantage for genetic disorders have yet to yield results as conclusive as the sickle-cell model. The trouble rests in the nature of the historical data on infectious disease—we do not know enough about epidemiological patterns in Europe prior to this century.

The high frequency of CF supports the contention that there may be a carrier advantage for this disease, but so far no single hypothesis has been accepted. This being said, the cholera/*E. coli* hypothesis, with its caveats, does provide one of the better-researched and documented theories. However, further testing of the theory is necessary. In particular, the historical and modern epidemiology of cholera should be consistent with the epidemiology of CF; if it is not consistent, a rigorously tested explanation must be provided to explain the anomaly.

Acknowledgments

I would like to express my gratitude to Dr. Anne Zeller at the University of Waterloo. Dr. Zeller encouraged my investigation into the cystic fibrosis heterozygote advantage and provided continual support throughout my research. I would also like to thank Dr. Ann Herring (McMaster University) whose experience and guidance saw me through the final revisions and drafts of this article. Finally, I offer my thanks to the reviewers of my work whose suggestions not only improved the quality of this article, but have also challenged me to think in new ways about other endeavours.

References Cited

Allison, A.C.
1954 Protection Afforded by Sickle Cell Trait Against Subtertian Malarial Infection, *British Medical Journal*, 1: 290-294.

Armstrong, J.
1992 Another Protein Out in the Cold, *Nature*, 358: 709-710.

Baxter P.S., J. Goldhill, J. Hardcastle, P.T. Hardcastle and C.J. Taylor
1988 Accounting For Cystic Fibrosis, *Nature*, 335: 211.

Beaudet, A. L. et al.
1991 Mutation Analysis for CF in a North American Population, *The Identification of the CF Gene*, Lap-Chee Tusi et al. (eds.), New York: Plenum Press: 53-54.

Behm. J.K., G. Hagiwara, N.J. Lewiston, P.M. Quinton and J.J. Wine.
1987 Hyposecretion of Beta-adrenergically Induced Sweating in Cystic Fibrosis Heterozygotes, *Pediatric Research*, 22: 271.

Bijman, J, and H. Dejonge.
1988 Cystic Fibrosis Advantage, *Nature*, 336: 430.

Bilson, G.
1980 *A Darkened House: Cholera in 19th Century Canada*, Toronto: University of Toronto Press.

Chao, A.C., F.J. De Sauvage, Y.J. Dong, J.A. Wagner, D.V. Goeddel and P. Gardner
1994 Activation of Intestinal CFTR Cl—Channel by Heat Stable Enterotoxin and Guanylin via cAMP-dependent Protein Kinase, *The Embo Journal, European Molecular Biology Organization*, 13(5): 1065-1072.

Collins, F.S.
1992 Cystic Fibrosis: Molecular Biology and Therapeutic Implications, *Science*, 256: 774-779.

Collins, F.S. and J.M. Wilson
1992 A Welcome Animal Model, *Nature*, 348: 708-709.

Cutting, G.R., S.M. Curristin, E. Nash, B.J. Rosenstein, I. Lerer, D. Abeliovich, A. Hill and C. Graham
1992 Analysis of Four Diverse Population Groups Indicates that a Subset of Cystic Fibrosis Mutation Occur in Common Among Caucasians, *American Journal of Human Genetics*, 50: 1185-1194.

The Cystic Fibrosis Genetic Analysis Consortium
1994 Population Variation of Common Cystic Fibrosis Mutations, *Human Mutation*, 4: 167-177.

Dean, M. et al.
1991 Identification of Cystic Fibrosis Mutations, *The Identification of the Cystic Fibrosis Gene*, Lap-Chee Tsui et al. (eds.), Plenum Press: New York: 45-51.

De Braekleleer, M., and J. Daignaeult
1992 Spatial Distribution of the DF508 Mutation in Cystic Fibrosis: A Review, *Human Biology*, 64(2): 167-174.

Delaporte, F.
1986 *Disease and Civilization: The Cholera in Paris 1832*, Cambridge, MA: MIT Press.

Denning, G.M., M.P. Anderson, J.F. Amara, J. Marshall, A.E. Smith and M.J. Welsh
1992 Processing of Mutant Cystic Fibrosis Transmembrane Conductance Regulator is Temperature-Sensitive, *Nature*, 348: 761-764.

Devoto, M.
1991 Origin and Diffusion of the Major CF Mutations in Europe, *The Identification of the Cystic Fibrosis Gene*, Lap-Chee Tsui et al. (eds.), New York: Plenum Press: 63-75.

Dori, J.R., P. Dickinson, E.W.F.W. Alton, S.N. Smith, D.M. Geddes, B.J. Stevenson, W.L. Kimber, S. Flamming, A.R. Clarke, M.L. Hooper, L. Anderson, R.S.P. Beddington and D.J. Porteous
1992 Cystic Fibrosis in the Mouse by Targeted Insertional Mutagenesis, *Nature*, 359: 211-215

Durey, M.
1979 *The Return of the Plague: British Society and the Cholera 1831-32*, Dublin: Gill and McMillan.

Egan, M.E, E.M. Schwiebert and W.B. Guggino
1995 Differential Expression of Orcc and CFTR Induced by Low Temperature in CF Airway Epithelial Cells, *American Journal of Physiology*, 268 (1, part 1): C243-C251.

Field, M., and C.E. Semrad
1993 Toxigenic Diarrheas; Congenital Diarrheas; and Cystic Fibrosis: Disorders of Intestinal Ion Transport, *Annual Review of Physiology*, 55: 631-655.

Gabriel, S.E., K.N. Brigman, B.H. Koller, R.C. Boucher and M.J. Stutts
1994 Cystic Fibrosis Heterozygote Resistance to Cholera Toxin in the Cystic Fibrosis Mouse Model, *Science*, 266: 107-109.

Glausiusz, J.
1995 Hidden Benefits, *Discover*, March: 30,31.

Goldstein, J.L, M.B. Bhuva, T.J. Layden and M.C. Rao
1991 E. Coli Heat Stable Enterotoxin Stimulated Chloride Secretion in Abnormal in Cystic Fibrosis (CF) Rectal Mucosa in Vivo, *Pediatric Pulmonology. Supplement*, 6: 260.

Greenwood, M.
1977 *Epidemics and Crowd Diseases*, New York: Arno Press.

Hansson, G.C.
1988 Cystic Fibrosis and Chloride Secreting Diarrhoea, *Nature*, 333: 711.

Harris, H.
1980 *The Principles of Human Biochemical Genetics*, 3rd ed., New York: Elsevier/North-Holland Biomedical Press.
Jorde, L.B., and G.M. Lathrop
1988 A Test of the Heterozygote Advantage Hypothesis in Cystic Fibrosis Carriers, *American Journal of Human Genetics*, 42: 808-815.
Kerem, B.-S., J.M. Rommens, J.A. Buchanana, D. Markiewicz, T.K. Cox, A. Chakravarti, M. Buchwald and Lap-Chee Tsui
1989 Identification of the Cystic Fibrosis Gene: Genetic Analysis, *Science*, 245: 1073-1080.
Klinger, K.W. et al.
1991 Molecular and Genetic Analysis of Cystic Fibrosis, *The Identification of the CF Gene*, Lap-Chee Tsui et al. (eds.), New York: Plenum Press: 39-45.
Knudson, A.G., and L. Wayne, Jr.
1967 On Selective Advantage of Cystic Fibrosis Heterozygotes, *American Journal of Human Genetics*, 19: 388-392.
Kristidis, P., D. Bozon, M. Corey, D. Markiewicz, J. Rommens, L.-C. Tsui and P. Durie
1992 Genetic Determination of Exocrine Pancreatic Function in Cystic Fibrosis, *American Journal of Human Genetics*, 50: 1178-1184.
Liu, J., W. Lissens, S.J. Silber, P. Devroey, I. Liebaers and A. Van Steirteghem
1994 Birth after Pre Implantation Diagnosis of the Cystic Fibrosis ΔF508 Mutation by Polymerase Chain Reaction in Human Embryos Resulting from Intracytoplasmic Sperm Injection with Epididymal Sperm, *Journal of the American Medical Association*, 272(23): 1858-1860.
Lloyd-Still, J.D.
1983 *Textbook of Cystic Fibrosis*. Boston: John Wright PSG.
Matalon, R., and A. Dorfman
1968 Acid Mucopolysaccharides in Cultured Fibroblasts of Cystic Fibrosis of the Pancreas, *Biochemical and Biophysical Research Communications*, 33: 954-958.
Meindl, R.S.
1987 Hypothesis: A Selective Advantage for Cystic Fibrosis Heterozygotes, *American Journal of Physical Anthropology*, 74: 39-45.
Morral, N., R. Llevadot, T. Casals, P. Gasparini, M. Macek, Jr., T. Dörk and X. Estivill
1994 Independent Origins of Cystic Fibrosis Mutations R334W; R347P; R1162X; and 3849=10kbC⇒T Provide Evidence of Mutation Recurrence in the CFTR Gene, *American Journal of Human Genetics*, 55: 890- 898.
Mosby's Dictionary
1990 *Mosbys Dictionary: Medical; Nursing; and Allied Health*, 3rd ed., St. Louis: Mosby.
Moss, R.B, Y.P. Hsu, P.H. Van Eede, A.M. Van Lewiston, N.J. Lewiston and G. De Lange
1987 Altered Antibody Isotype in Cystic Fibrosis: Impaired Natural Antibody Response to Polysaccharide Antigens, *Pediatric Research*, 22: 708-713.
Phillips, W.J.
1991 *Medical-Surgical Nursing: Concepts and Clinical Practice*, 4th ed., St. Louis: Mosby Year Book.
Poirier, F.E.
1993 *Understanding Human Evolution*, 3d ed., New Jersey: Prentice Hall.
Riordan, J.R, J. Rommens, B.-S. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plavisic; J.-L. Chou, M.L. Drumm, M.C. Iannuzzi, F.S. Collins and L.-C. Tsui
1989 Identification of the Cystic Fibrosis Gene: Cloning and Characterization of Complementary DNA, *Science*, 245: 1066-1072.
Riordan, J.R.
1993 The Cystic Fibrosis Transmembrane Conductance Regulator, *Annual Review of Physiology*, 55: 609-630.
Rodman, D.H., and S. Zamudio
1991 The CF Heterozygote: Advantage in Surviving Cholera? *Medical Hypothesis*, 36: 253-258.
Rommens, J., M.C. Iannuzzi, B.-S. Kerem, M.L. Drumm, G. Melmer, M. Dean, R. Rozmahel, J.L. Cole, D. Kennedy, N. Hidaka, M. Zsiga, M. Buchwald, J.R. Riordan, L.-C. Tsui and F.S. Collins
1989 Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping, *Science*, 245: 1059-1072.
Saunders, W.B. (Dictionary Staff)
1994 *Dorlands Medical Dictionary*, 28th ed., Philadelphia: W.B. Saunders.
Sereth, H., T. Shoshani, N. Bashan and B.-S. Kerem
1993 Extended Haplotype Analysis of CF Mutations and Its Implications for the Selective Advantage Hypothesis, *Human Genetics*, 92: 289-295.
Serre, J.L.
1991 Towards a Geographical History of the Predominant and Secondary Mutations in Europe, *The Identification of the CF Gene*, Lap-Chee Tsui et al. (eds.), New York: Plenum Press: 55-63.
Serre, J.L., B. Simon-Bouy, E. Mornet, B. Kaume-Roig, A. Balassopoular, M. Schwartz, A. Taillandier, J. Boue and A. Boue
1990 Studies of RFLP Closely Linked to the Cystic Fibrosis Locus Throughout Europe Lead to New Considerations in Population Genetics, *Human Genetics*, 84: 449-454.
Smith, H.R, G.S. Dhatt, W.M.A. Melia and J.G. Dickindon
1995 Cystic Fibrosis Presenting as Hyponatraemic Heat Exhaustion, *British Medical Journal*, 310: 579-580.
Sojo, A., J. Rodriguez-Soriano, J.C. Vitoria, C. Vazquez, G. Ariceta and A. Villate
1994 Chloride Deficiency as a Presentation or Complication of Cystic Fibrosis, *European Journal of Pediatrics*, 153: 825-828.
Tizzano, E., M.M. Silver, D. Chityat, J.-C. Benichou and M. Bachwald
1994 Differential Cellular Expression of CFTR in Human Reproductive Tissues, *American Journal of Pathology*, 144(5): 906-914.
Tsui, L.-C. et al. (eds.).
1991a *Advances in Experimental Medicine and Biology*, Vol. 290: *The Identification of the CF Gene: Recent Progress and New Research Strategies*, New York: Plenum Press.
1991b Molecular Genetic of CF, *Identification of the CF Gene*, Lap-Chee Tsui et al. (eds.), New York: Plenum Press: 9-19.

Whalley, L.F., and D.L. Wong
1991 *Nursing Care of Infants and Children*, 4th ed., St. Louis: Mosby Year Book.
Wine, J.J. et al.
1991 Cystic Fibrosis; The CFTR; and Rectifying Chloride Channels, *Identification of the CF Gene*, Lap-Chee Tsui et al. (eds.), New York: Plenum Press: 253-273.
Wine, J.J.
1993 Indictment of Pore Behaviour, *Nature*, 366: 18-19.
Zielenski, J., D. Markiewicz, F. Rininsland, J. Rommens and L.C. Tsui
1991 A Cluster of Highly Polymorphic Dinucleotide Repeats in the Intron 17b of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Gene, *American Journal of Human Genetics*, 49: 1256-1262.