

NEWS AND COMMENTARY

Cystic Fibrosis

Cystic fibrosis and lactase persistence: a possible correlation

Guido Modiano, Bianca M Ciminelli and Pier F Pignatti

European Journal of Human Genetics (2007) 15, 255–259.
doi:10.1038/sj.ejhg.5201749; published online 20 December 2006

Cystic fibrosis (CF) is the most common lethal genetic disease in Europeans and is transmitted as an autosomal recessive disorder caused by loss of function of the Cystic Fibrosis Transmembrane conductance Regulator (CFTR) gene. Since long it has been supposed that its quite high frequency (1/2500 implying an allele frequency equal to 0.02) is due to a Europe-restricted selective advantage of the heterozygous CF/+ individuals. Nevertheless, this hypothesis has never been directly demonstrated nor the hypothetical selective factor identified. Here, we provide compelling evidence that the CF/+ advantage hypothesis is right. As to the identification of the selective factor, as the CFTR protein is a Cl⁻ channel (and hence involved in water excretion), diarrhea appears a reasonable candidate for such a role. Cholera has, in fact, been suggested, but its area of diffusion had not been limited to Europe. Considering that initially all humans were lactose intolerant in the post-weaning life and Europeans became cattle breeders and adopted a dairy-milk diet, we propose that the hypothetical Europe-restricted CF/+ advantage consisted of a resistance to lactose-caused diarrhea in populations that adopted such a habit while they were still lactose intolerant. This initial emergency segregational load adaptation would have been followed by a not costly one consisting of the present European quasi-fixed post-weaning lactase persistence (due to the dominant P allele of the

lactase, LCT, gene), resulting in lactose tolerance. This hypothesis is supported by the positive correlation between the proportion, among the CF alleles, of the F508del allele and the P allele – recently identified with the T allele 14 Kb upstream to the LCT gene – frequency and provides a possible explanation for the well-known north–south cline of the F508del allele frequency.

Both adaptations would be examples of genetic responses to a culture-rooted adverse factor.

CF is the most common (1/2500 newborns) severe autosomal recessive disease in Europe. The global frequency of the CF-causing alleles of the CFTR gene, estimated as the square root of the prevalence of the disease, turned out to be about 0.02 in Europe, 0.01 in Africa and 0.002 in Far-Eastern Asia.¹

The high CF allele frequency in Europe could only be accounted for by a 'Europe-restricted' high mutation rate or by a 'Europe-restricted' heterozygous CF/+ advantage. The most plausible explanation seems to be the latter, mainly because a 'Europe-restricted' high mutation rate was judged extremely unlikely, although it was never directly disproved. In addition, the biological basis of the hypothetical 'Europe-restricted' CF/+ heterozygous advantage is far from being identified because of the lack of a convincing candidate adverse factor: bacteria-driven diarrheas – proposed on the basis of the functional properties of the CFTR

protein – do not seem to have had 'Europe-restricted' histories, a fundamental prerequisite for a good candidate.

In this paper, first we definitively reject the 'Europe-restricted' high mutation rate ($\mu_{+ \rightarrow CF}$) hypothesis; then we propose a 'Europe-restricted' biological factor that could account for the CF/+ heterozygous advantage.

Discarding the 'Europe-restricted high $\mu_{+ \rightarrow CF}$ hypothesis'

The mutation rates $CFTR^{+} \rightarrow CFTR^{CF}$ to be postulated in order to account, in terms of mutation only, for the present CF gene frequencies are 4×10^{-4} for Europe, 10^{-4} for Africa and 4×10^{-6} for Far-Eastern Asia.

A by-product of the 'Europe-restricted' high $\mu_{+ \rightarrow CF}$ hypothesis would be the *continuous and rapid substitution* of the pre-existing 'old' CF alleles with an equal number of 'new' CF alleles; this process would have taken place according to the function

$(F_{old}/F_{total})_t = (F_{old}/F_{total})_{t=0} e^{-0.02t}$, where F_{old} is the proportion of the 'old' CF alleles among all the CF alleles (the frequency of which is F_{total}); t is the time, expressed in terms of generations, elapsed from an arbitrary moment ($t=0$) when all the CF alleles were 'old' (thus F_{old}/F_{total} was 1); and 0.02 is the rate of substitution of the 'old' CF alleles by an equal number of 'new' CF alleles, per generation.

The 'Europe-restricted' heterozygous advantage would instead result in a substantially complete *maintenance* of the 'old' CF alleles, namely of their original spectrum and haplotype relationships.

Fundamental – and testable – implications are that the high $\mu_{+ \rightarrow CF}$ hypothesis expects that the present spectrum of the CF alleles, besides being drastically different from the original one, consists of a very large number of different CF alleles, each extremely rare, and distributed, on the whole, at random with respect to the alleles of the polymorphic sites of the CFTR gene. On the contrary, the heterozygous advantage hypothesis expects that the present CF genes spectrum reflects 'faithfully' the original one in terms of both *composition* and *haplotypic relationships*. In other words, the few ancient CF alleles should be still common and largely located in their original haplotype.

To choose between these two sets of expectations, we have worked out the data reported in Ciminelli *et al.*,² where 358 non-F508del CF alleles had been studied in order to ascertain their association with the alleles of the M470V highly polymorphic site.

Two aspects of the data have been considered:

- (a) the percentage of the recent (mainly the singletons) and of the ancient CF alleles (all the others) and
- (b) the extent of their preferential association with the M allele.

Table 1 indicates that at least 90% of the CF alleles are ancient and that they are associated with the M allele 14 times (21.2/1.5) more frequently than the recent CF mutations. Both sets of results are incompatible with the 'Europe-restricted' high $\mu_{+ \rightarrow CF}$ hypothesis.

Therefore, these data conclusively disprove the 'European-restricted' high mutation rate hypothesis, thus proving, by exclusion, that the high CF gene frequency is due to a CF/+ heterozygous advantage.

Nature of the CFTR⁺/CFTR^{CF} heterozygote advantage

The CFTR protein is the Cl⁻ ion channel responsible for the displacement of that ion, and consequently of water, from the mucosa epithelial cells into the respiratory and intestinal lumen. It is therefore reasonable to believe that heterozygosity for a CFTR functionally silent allele (= a CF allele) makes the

intestinal epithelial cells less prone to secrete Cl⁻ ions and water, as shown in the mouse model,⁵ although not confirmed in humans.⁶ Chloride transport *in vivo* in CF heterozygotes is often below 50%, and at times normal,⁷ corresponding to the severity of the known mutations.⁸

Thus, considering that one of the events required to develop any type of diarrhea is the accumulation of water in the intestinal lumen and that the most severe diarrheas have been caused by cholera, it has been suggested that:

1. the CF heterozygotes are more resistant to the diarrhea-causing factors and this resistance would have conferred to them a selective advantage sufficient to maintain the CF alleles at frequencies very high for a lethal allele;⁹⁻¹²
2. the main diarrhea-causing factor has been tentatively identified with cholera epidemics.¹³

Here, we propose to accept the first of the two hypotheses only, namely that the cause (hereafter named adaptogen) of the CF/+ balanced polymorphism has been a diarrhea-causing factor. As for the second hypothesis, as there is no stringent indication that Europeans have been preferentially exposed to cholera (and/or allied diseases), it appears reasonable to assign the role of major candidate adaptogen, instead of to cholera, to any Europe-restricted intense and long-lasting diarrhea-causing factor. We hypothesized that the role of major adaptogen has been played by the dairy-milk diet adopted by Europeans when they were still lactose

intolerant. In fact, the vast majority of human populations, as all the other mammals, are lactose intolerant in the post-weaning life; lactose tolerance is a feature restricted to the European populations, especially to those of North-Western Europe, where it is near fixation.^{14,15} According to our hypothesis, when the cattle breeder populations of Europe began to make use of milk in the post-weaning life, they were lactose intolerant and the balanced polymorphism for the CF alleles has represented an 'emergency' (with a high cost/benefit ratio due to the high segregational load) adaptation to this dietary change. Clearly, the truly effective (etiological), and ultimately successful, adaptation to the milk diet has been a dramatic increase, even reaching fixation in North-Western Europe, of lactase persistence, that is, lactose tolerance, as suggested previously.¹⁶⁻¹⁹

Table 2 lists the main properties required to make a factor a good candidate for having played the role of adaptogen and compares cholera and lactose intolerance for these properties. Cholera and allied enterobacteria-caused diarrheas should be discarded mainly because of the lack of a 'Europe-restricted' distribution. Dairy milk plus lactose intolerance, instead, appears to be a better candidate.

Obviously, the selective 'usefulness' of such 'emergency' adaptation should have become less and less with time down to nullification as the frequency of lactose tolerance was increasing.

The question then arises: 'Why the onset of the lactase-persistence adaptation

Table 1 Estimate of the relative proportions of RECENT and ANCIENT CF alleles and of the extent of the preferential association with the M470 allele

Type of CF alleles	Frequency		Nos. of CF associated with				P	Degree of the preferential association with the M allele ^b
	Abs	Rel	M		V			
			Obs	Exp ^a	Obs	Exp ^a		
Recent (mainly singletons)	75	0.105	36	28.9	39	46.1	~0.1	1.5
Ancient	641 ^c	0.895	596	246.8	45	394.2	0	21.2 ^d

^aOn the basis of the M and V allele frequencies in the random European population (0.385 and 0.615: see Modiano *et al.*³).

^bno. of CF-M haplotypes/no. of CF-V haplotypes) × (V freq/M freq).

^cComputed as follows from Table 1 of Ciminelli *et al.*²: the sum of the Italian and Czech Grand totals (272+86 = 358); then, by assigning to F508del alleles (all 'ancient') an absolute frequency of 358 (assuming that they make up half of the CF alleles:), the total number of CF alleles becomes $358 \times 2 = 716$ and the no. of 'ancient' CF alleles is $716 - 75 = 641$.

^dThe intensity of such overall preferential association of the 'ancient' CF alleles with the M allele is comparable to that observed for the random CFTR alleles (Pompei *et al.*⁴).

Table 2 Comparison between two diarrhea-causing factors, cholera and dairy milk plus lactose intolerance, for their compliance to the prerequisites of an ideal adaptogen candidate in the CF/+ heterozygous advantage

Adaptogen Candidate	Severity of the single episodes	Required properties			
		Overall selective pressure severe episodes in one individual's lifespan	Exposed generations	Antiquity	Europe restriction
Identikit of the 'ideal' candidate	At least +	+	Many	++	+++
Cholera and other severe and epidemic enterotoxigenic diseases	++++	< +	Few	++	—
Dairy milk+LACTOSE INTOLERANCE	+(+)	+++	All ^(a)	+++	++++

^aIn cattle breeding populations only.

was so much delayed with respect to the CF 'emergency' adaptation?

A reasonable answer is that CF alleles, being 'loss of function' alleles, were certainly present when the selective pressure due to the combined effect of lactose intolerance and dairy milk started, whereas lactose-intolerance, a sophisticated phenotypic modification (a single gene ontogenetically regulated in all mammals to express itself only during weaning had to be 'convinced' to continue to work through post-weaning life), was not so easily available or even absent in the exposed populations.

In summary, the most economical scenario to explain two 'Europe-restricted' features, the extremely high frequency of lactose tolerance and the high frequency of the CF alleles, consists in considering them as facets of the same coin: two profoundly different adaptations to the same stringent, culture-rooted adaptogen, the abundant intake of lactose during the post-weaning life. Both adaptations would avoid severe diarrhea attacks following lactose ingestion. Lactase persistence acts by preventing the arrival of lactose into the colon and the consequent hyperfermentation owing to its utilization by the intestinal bacteria; CF heterozygosity would instead act by reducing the diarrhoeal effect of the hyperfermentation. The time course of events might have been the following:

- Initially all humans were lactose intolerant and the cumulative CF gene frequency was maintained at a value of about 0.002 by a $\mu_{+ \rightarrow CF} \approx 4 \times 10^{-6}$ (as in Japan).
- 5000–10 000 years ago cattle-breeding and dairy milk post-weaning diet started,^{19,20} resulting in a stringent selective pressure.

- The onset of an 'emergency' adaptive response consisting of an immediate expansion of the already available CF alleles (the antiquity of the F508del allele has been estimated to be 52,000 years²¹). They would have attained and maintained a frequency unknown, but possibly higher than the present one, resulting from a balance between the advantage of being more resistant to the adverse consequences of being unable to digest lactose (hence to benefit from an important food as the milk) and a heavy segregational load.
- The onset of the not costly lactase-persistence adaptation, which attained rapidly a very high frequency (up to fixation in some European areas).^{14,15}
- A progressive decrease of the CF gene frequency down to the present (still high) values, which can be considered a relic of the past 'emergency' adaptation.

The fact that the global CF gene frequency in Africa, where some ancient pastoralist cultures do occur, is believed to be about 0.01 – a value intermediate between those of Europe and Japan – would be in favor of the present hypothesis if a positive correlation between pastoralism and CF gene frequency distribution in Africa were found. However, owing to the lack of information about the spectrum of CF alleles in African populations and the low prevalence of the disease, a search in this direction could only be based on extensive epidemiological surveys, not to mention diagnostic problems.

The molecular basis of the lactose-persistence phenotype (consisting in a

T/C polymorphism 14 kb upstream to the LCT gene, with the dominant T allele conferring lactase persistence) has been identified recently²² so that directly testing the hypothesis by comparing CFTR⁺/CFTR⁺ with CFTR⁺/CF subjects all with the C/C lactose-intolerant genotype would be a feasible, but not an easy, task.

Geographic aspects

Both LCT**T* and F508del alleles are monophyletic.^{23,24} This is equivalent to saying that they were born in a single geographic area where, being selectively advantageous, they attained relatively high frequencies and were then exported to other areas. The spectrum of the potentially adaptive alleles of the LCT gene is likely to be very limited, so the fact that the T allele was the one selected was *a priori* a very likely event; on the contrary, as the potential spectrum of pathologic (CF) alleles is extremely large, it appears likely that the F508del allele was the successful one by pure chance.

The question then arises: 'Why the various European populations are all adapted to a dairy milk diet mainly with the same CF allele (F508del) instead of being each adapted with its own CF allele(s)?'

The simplest and most economical explanation is that a dairy-milk diet became established in a single area and remained restricted to that area for a period of time sufficient to allow the T and the F508del alleles to attain high values. Then, in a second phase, the population of that area exported to the rest of Europe its dairy-milk diet culture

together with the two adaptive genes, that is, the adaptogen and the two genetic adaptations to it. These two alleles would have then been amplified in the recipient populations because of their adaptive value owing to the co-imported dairy milk diet.

As to the identification of the 'donor area', there is no doubt that by far the best candidate is Northern Europe, where these alleles have their highest frequencies. This is perfectly in line with the proposal of Beja-Pereira *et al.*²⁵ (based on data on six milk protein genes of cows), according to which dairy cow breeding was first developed in Northern Europe.

This interpretation is consistent with the highly significant ($P < 0.001$) positive correlation between the LCT*T frequency and the proportion, among the CF alleles, of the F508del allele, as shown in Figure 1, and would account for the well-known F508del allele frequency north-south cline.

It may be worth noticing that the rest of Europe has not reached the 'donor area' neither for the T allele nor for the F508del allele frequency, but it did for the combined CF alleles frequency (about 0.02 in the whole of Europe) through the complementary (to 0.02) accumulation of CF alleles other than F508del. This can be explained by considering that, whereas it would have been difficult to mount an adaptation at the LCT level (in addition to that provided through importation), there were a number of opportunities for adaptation through CF alleles other than F508del ■

Acknowledgements

We thank LL Cavalli-Sforza, L Luzzatto and L Terrenato for their comments on the manuscript.

Guido Modiano and Bianca M Cimenelli are at Department of Biology, University of Roma-Tor Vergata, Italy;
Pier F Pignatti is at Department of Mother and Child and of Biology-Genetics, University of Verona, Italy

Correspondence: Professor G Modiano, Dipartimento di Biologia, University of Roma-Tor Vergata, Via della Ricerca

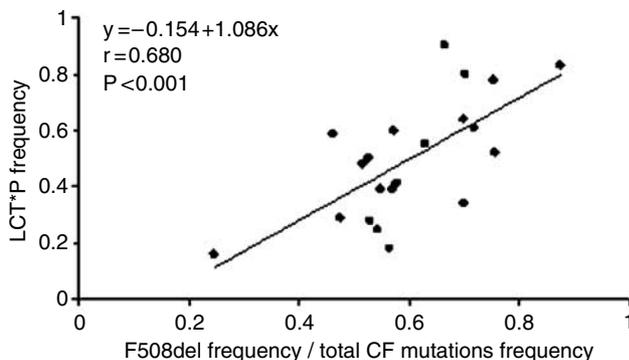


Figure 1 Positive correlation between F508del and LCT*P. Data for different European populations from Swallow and Hollox¹⁴ and Bobadilla *et al.*²⁶ LCT*P indicates the dominant lactase-persistence allele identified at the phenotypic level (ability to digest lactose); at a first approximation it can be considered to coincide with the LCT*T allele identified at the DNA level.¹⁵

Scientifica, s.n.c., 00133 Roma, Italy.

Tel.: +39 0672594341;

Fax: +39 062023500;

E-mail: modiano@uniroma2.it

References

- 1 WHO Report. 2004, www.who.int/genomics/publications/en/.
- 2 Ciminelli BM, Bonizzato A, Bombieri C *et al*: Highly preferential association of the non-F508del CF mutations with the M470 allele. *J Cystic Fibrosis* 2006; in press.
- 3 Modiano G, Bombieri C, Ciminelli BM *et al*: A large-scale study of the random variability of a coding sequence: a study on the CFTR gene. *Eur J Hum Genet* 2005; 13: 184–192.
- 4 Pompei F, Ciminelli BM, Bombieri C *et al*: Haplotype block structure study of the CFTR gene. Most variants are associated with the M470 allele in several European populations. *Eur J Hum Genet* 2006; 14: 85–93.
- 5 Gabriel SE, Brigman KN, Koller BH, Boucher RC, Stutts MJ: Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. *Science* 1994; 266: 107–109.
- 6 Hogenauer C, Santa Ana CA, Porter JL *et al*: Active intestinal chloride secretion in human carriers of cystic fibrosis mutations: an evaluation of the hypothesis that heterozygotes have subnormal active intestinal chloride secretion. *Am J Hum Genet* 2000; 67: 1422–1427.
- 7 Sermet-Gaudelus I, Dechaux M, Vallee B *et al*: Chloride transport in nasal ciliated cells of cystic fibrosis heterozygotes. *Am J Respir Crit Care Med* 2005; 171: 1026–1031.
- 8 Hallberg K, Reims A, Strandvik B: Electro-genic ion transport in duodenum, an aid in cystic fibrosis diagnosis. *Scand J Gastroenterol* 2000; 35: 1106–1109.
- 9 Baxter PS, Goldhill J, Hardcastle J, Hardcastle PT, Taylor CJ: Accounting for cystic fibrosis. *Nature* 1988; 335: 211.
- 10 Hansson GC: Cystic fibrosis and chloride-secreting diarrhoea. *Nature* 1988; 333: 711.
- 11 Field M, Semrad CE: Toxigenic diarrheas, congenital diarrheas, and cystic fibrosis: disorders of intestinal ion transport. *Annu Rev Physiol* 1993; 55: 631–655.
- 12 Quinton PM: Human genetics. What is good about cystic fibrosis? *Curr Biol* 1994; 4: 742–743.
- 13 Rodman DM, Zamudio S: The cystic fibrosis heterozygote – advantage in surviving cholera? *Med Hypotheses* 1991; 36: 253.
- 14 Swallow DM, Hollox EJ: The genetic polymorphism of intestinal lactase activity in adult humans; in CR Scriver, AL Beaudet, WS Sly, D Valle (eds): *Metabolic and Molecular basis of Inherited Disease*. New York: McGraw Hill, pp 1651–1663.
- 15 Swallow DM: Genetics of lactase persistence and lactose intolerance. *Ann Rev Genet* 2003; 37: 197–219.
- 16 Cavalli-Sforza LL: Analytic review: some current problems of human population genetics. *Am J Hum Genet* 1973; 25: 82–104.
- 17 Flatz G: Genetics of lactose digestion in humans; in Harris H, Hirschhorn K (eds): *Advances in Human Genetics*, Vol 16. New York: Plenum Press, 1987, pp 1–77.
- 18 Hollox EJ, Poulter M, Zvarik M *et al*: Lactase haplotype diversity in the Old World. *Am J Hum Genet* 2001; 68: 160–172.
- 19 Klein RG 2nd edn. Chicago: The University Chicago Press, 1999.
- 20 Klein RG: Archeology and the evolution of human behavior. *Evol Anthropol* 2000; 9: 17–36.
- 21 Morral N, Bertranpetit J, Estivill X *et al*: The origin of the major cystic fibrosis mutation (delta F508) in European populations. *Nat Genet* 1994; 7: 169–175.
- 22 Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I: Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 2002; 30: 233–237.
- 23 Bersaglieri T, Sabeti PC, Patterson N *et al*: Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 2004; 74: 1111–1120.

- 24 Sereth H, Shoshani T, Bashan N, Kerem BS: Extended haplotype analysis of cystic fibrosis mutations and its implications for the selective advantage hypothesis. *Hum Genet* 1993; **92**: 289–295.
- 25 Beja-Pereira A, Luikart G, England PR *et al*: Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nat Genet* 2003; **35**: 311–313, Erratum in: *Nat Genet* 2004; **36**: 106.
- 26 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.