Adaptive evolution of the human fatty acid synthase gene: Support for the cancer selection and fat utilization hypotheses?

Mary J. O'Connell [a], James O. McInerney [b,*]

[a] Bioinformatics and Molecular Evolution Laboratory, School of Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland
[b] Bioinformatics Laboratory, Department of Biology, National University of Ireland, Maynooth, Co. Kildare, Ireland

Abstract

Cancer may act as the etiological agent for natural selection in some genes. This selective pressure would act to reduce the success of neoplastic lineages over normal cell lineages in individuals of reproductive age. In addition, human's relatively larger brain and longer lifespan may have also acted as a selective force requiring new genotypes. One of the most important proteins in both processes is the fatty acid synthase (FAS) gene involved in fatty acid biosynthesis. A variety of other proteins, including PTEN, MAPK1, SREBP1, SREBP2 and PI are also involved in the regulation of fatty acid biosynthesis. We have specifically analysed variability in selective pressure across all these genes in human, mouse and other vertebrates. We have found that the FAS gene alone has signatures indicative of adaptive evolution. We did not find any signatures of adaptive evolution in any of the other proteins. In the FAS gene, we have detected an excess of non-synonymous over synonymous substitutions in approximately 6% of sites in the human lineage. Contrastingly, the substitution process at these sites in other available vertebrates and mammals indicates strong purifying selection. This is likely to reflect a functional shift in human FAS and correlates well with previously observed changes in FAS biochemical activities. We speculate that the role played by FAS either in cancer development or in human brain development has created this selective pressure, although we cannot rule out the various other functions of FAS.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Darwinian selection; Protein evolution; Metabolism; Brain size; Selection

1. Introduction

Cancer selection describes the situation where cancer acts as the agent of selection favoring novel mutations over the

wild type (Leroi et al., 2003). While most cancers are thought of as diseases of the elderly, with their onset being later than the end of normal reproductive age, some cancers affect the young and therefore can influence allele frequencies in the next generation. In addition, cancer selection theory predicts that new mutations would be favored as animals change their life-histories, evolve new morphologies and longer lifetimes (Leroi et al., 2003). This theory is reasonably falsifiable in the sense that there are recognizable signatures of selection and these should be found in some or many of the genes implicated in cancer. This can shed light on the likely historical impact of cancer in human evolution or at least the historical selective pressures on genes involved in tumorigenesis.

Fatty acid synthesis is an integral part of all living systems (Roncari, 1981). Fatty acids are used for energy storage, synthesis of cell membranes (O'Brien, 1967) and

generation of cell signaling molecules (Duplus and Forest, 2002). Fatty Acid Synthase (FAS) is one of the largest vertebrate proteins and consists of a single multifunctional protein found as a homodimer (Chirala et al., 2001). It is a key metabolic enzyme catalyzing the synthesis of long chain saturated fatty acid molecules from acetyl-CoA, malonyl-CoA and NADPH (Wakil, 1989). An increase in FAS expression is one of the earliest and most common events in the development of many human cancers and is closely linked to malignant transformation and to tumor virulence (Kumar-Sinha et al., 2003). In contrast to normal cells that preferentially use dietary fatty acids (Zhou et al., 2003), human cancer cells are dependent upon fatty acid synthesis for growth and survival. Inhibiting fatty acid synthesis by inhibiting FAS in cycling cancer cells leads to cell death, while normal cells survive (Pizer et al., 2001; Zhou et al., 2003). Over-expression of FAS in cancer cells is both linked to the inactivation of PTEN (phosphatase and tensin homolog) and to increases in epidermal growth factor signaling (Van de Sande et al., 2002). These act in major part through the mitogen-activated protein kinase and phosphatidylino-sitol-3 kinase signaling cascades (MAP and PI-3K, respectively) (Yang et al., 2002). Cellular cholesterol and Fatty acid metabolism in mammals is controlled by

Sterol Regulatory Element Binding Proteins (SREBPs) (Nohturfft et al., 1999). By inhibiting MAP and PI-3 kinases, SREBP1 levels are downregulated and tran-scription of the FAS promoter is decreased, thereby reducing fatty acid synthesis in cancer cells (Yang et al., 2002, 2003). FAS regulation is likely to be of major significance in the treatment of cancers (Kumar-Sinha et al., 2003; Qiao et al., 2003). If positive selection for novel FAS genotypes has been a major contributor to the evolution of the human lineage, it should be possible to identify this situation. Positive selection for new geno-types is caused by selection for new phenotypes. One theory in particular — the 'fat utilization hypothesis' (Horrobin, 1998, 1999) is interesting as it proposes that changes in fatty acid synthesis may have been a major feature of human evolution. Here we will examine a number of genes with roles central to fatty acid metabolism and cancer.

The Neutral and Nearly Neutral theories of molecular evolution contend that the majority of sequence change is dictated by random genetic drift (Kimura, 1983). If this is true, then positive selective pressures that have been capable of overcoming random genetic drift are likely to be very rare, but important. Adaptive evolution of colubine lysozyme genes permits foregut fermentation in these

Table 1
Genes, accession numbers and lengths of alignments used in this study

Gene name	Human acc	Mouse acc	Other species_acc	Length
Fatty Acid Co-Enzyme A ligase (*F.A.C.*) (E.C.6.2.1.3)	L09229	BC056644	RAT_D90109	2196
	AB033899	BC031544	RAT_AB012933	
	BC026161	BC022959	RAT_D10041	
	AF030555	AJ243502	RAT_D85189	
	D89053	BC031529	RAT_D30666	
Acyl-Co-enzyme A Dehydrogenase (*A.C.A.D.*) (E.C.1.3.99.13)	M74096	U21489	PIG_D89478	1290
Liver Carnitine Palmitoyltransferase 1 (*L.C.P.T.*) (E.C.2.3.1.21)	BT009791	BC054791	SHEEP_Y18387	2319
Liver Fatty Acid Binding Protein (*LFABP*)	M10617	BC009812	PIG_X77640	381
Mitogen-Activated Protein Kinase-2 (*MAPK2*) (E.C.2.7.1.37)	M84489	M64300	COW_Z14089	1089
			FROG_M60977	
Phosphatidylinositol-3 kinase (*PI-3K*)	D30036	U96725	RABBIT_U12558	813
			RAT_M25758	
Sterol Regulatory Binding Protein 1 (*SREBP1*)	U00968	BC056922	RAT_AF286470	3471
			HAMSTER_U09103	
			CHICKEN_AY029224	
Sterol Regulatory Binding Protein 2 (*SREBP2*)	U02031	AF374267	HAMSTER_U12330	3432
			CHICKEN_AJ414379	
Fatty Acid Synthase (*F.A.S.*) (E.C. 2.3.1.85)	U26644	AF127033	CHICKEN_J03860	7581
RAS-related C3 botulinum toxin substrate 3 (*RAC3*)	BT019443	AB040819	RAT_AY491395	576
			COW_AF175262	
			ZEBRAFISH_BC044538	
Dicarbonyl/L-xylulose reductase (*DCXR*)	BT006881	D89656	HAMSTER_AB045204	732
			RAT_AB061719	
			GUINEAPIG_AB061720	
			FROG_BC062504	
Cop9 signalosome complex subunit 1 (*GPS1*)	U20285	BC003350	ORANGUTAN_CR861104	1581
			RAT_BC061746	
			CHICKEN_AJ719637	
Beta-1,3-*N*-acetylglucosaminyltransferase radical fringe (*RFNG*)	AF108139	U94350	RAT_AB016486	1143
			CHICKEN_GGU91850	
			ZEBRAFISH_AF510993	

animals and is responsible for a functional shift in the protein, specifically an increased pH tolerance and resistance to digestion by pepsin (Messier and Stewart, 1997). Positive selection has previously been reported in proteins with more central roles in metabolism, including extensive studies on the cytochrome oxidase (COX) complex in primates (Goldberg et al., 2003). The adaptive evolutionary events on the COX complex have been viewed as part of a series of coadaptive changes that optimized the anthropoid biochemical machinery for aerobic energy metabolism and alternatively the evolution of an expanded neocortex in anthropoid primates. There is strong evidence that the tumor suppressor gene *BRCA1* has undergone positive selection in the primate lineage, to a lesser extent along the chimpanzee lineage, but most prominently along the human lineage (Yang and Nielsen, 2002). During human evolution the *BRCA1* tumor suppressor gene has been under selective pressure for change perhaps conveying longevity to its host also showing the selective pressure cancer plays in evolution. This growing body of evidence indicates that positive selection was sufficiently strong on many occasions to overcome random genetic drift. These positive selection events are usually correlated with functional shifts in those proteins in which we find these changes.

In a comparative analysis of human and chicken FAS protein activities, it has been reported that there are some significant differences in the biochemical activities of these proteins (Jayakumar et al., 1995). The overall activity of FAS is much lower in humans than in chicken, with the FAS partial activities varying somewhat in their activities. For instance, the β-ketoacyl synthase activity is threefold lower in the human lineage and the enoyl reductase activity is twentyfold higher (Jayakumar et al., 1995).

In this study we sought to analyse variation in selective pressures across sites and lineages in a group of genes that are known to be involved in effecting and regulating fatty acid biosynthesis. Specifically, we wished to know if positive selection could have been a driving force in amino acid substitution in this cohort of proteins.

2. Materials and methods

2.1. Data assembly

We have created alignments of nine families of genes with documented roles in fat metabolism and in a variety of cancers. We have also analysed four neighboring genes of the *FAS* gene in the human and mouse genomes to determine if the GC content at this locus is influencing the evolutionary rate. These genes and their accession numbers and alignment lengths are listed in Table 1. In all cases, there was at least one human and one mouse sequence. These sequences were available in both the

GenBank and Swissprot databases, thereby minimizing the possibility of using incorrect sequences. All alignments were carried out at the amino acid level using ClustalX 1.81 (Thompson et al., 1994), and gap positions were placed into the nucleotide sequences according to where they were found in the protein alignment. The Fatty Acid Co-enzyme A ligase (FAC) data represents a multigene family. The FAC phylogenetic tree was reconstructed using the neighbor joining method and protein distances calculated from the PHYLIP package (Felsenstein, 1993), see Fig. 1.

2.2. Description of models employed in the Likelihood Ratio Test

The Likelihood Ratio Test (LRT) using nested models of sequence evolution was used to evaluate a variety of models of codon sequence evolution (Yang, 1997). The LRT proceeds by comparing nested models of sequence evolution. These models usually allow for variable non-synonymous to synonymous (Dn/Ds or ω) ratios among sites, along different branches of a phylogenetic tree or in combination. These models imply that there are a variety of classes of sites in a given set of aligned sequences and the LRT provides a method of identifying the model that best describes the evolution of the set of sequences. One
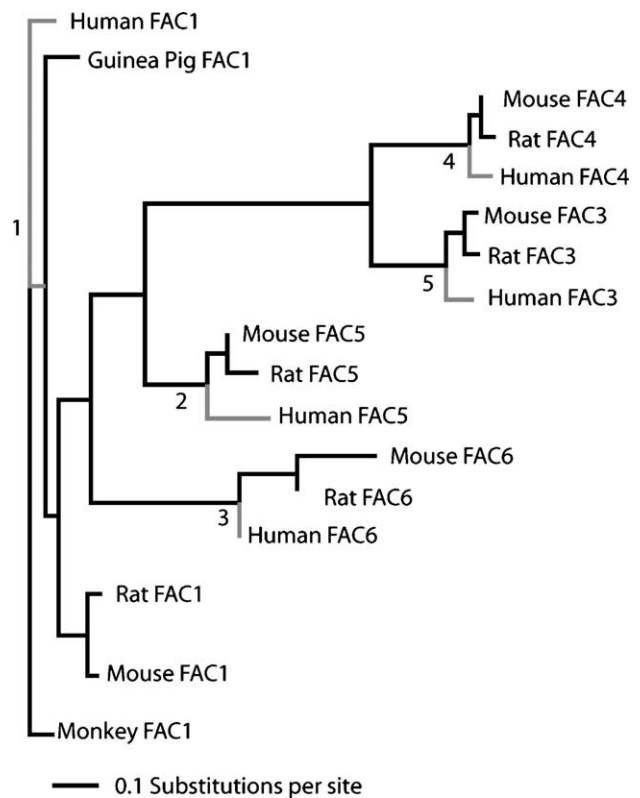

Fig. 1. Phylogenetic reconstruction of the Fatty Acid Co-enzyme A ligase (FAC) family. Branches labeled as foreground in the analyses are numbered 1 to 5.

model is usually constrained so that $\omega \leq 1$. The more general model allows at least one class of sites to exist where the ω value is dependent on the data. In those cases where the ω value exceeds unity and the resulting increase in the likelihood score is significant, we can conclude that positive selection has occurred. In order to ascertain significance of difference in likelihood score we can compare the likelihood statistic $2 \Delta l$ with χ^2_ν, where ν is the number of degrees of freedom and corresponds with the number of free parameters.

The simplest model (fewest free parameters) is called M0. In this model it is assumed that there is a single ω value at all sites and across all lineages. This corresponds to the Goldman and Yang model (Goldman and Yang, 1994). Model M1 assumes that there are two classes of sites — those with an ω value of zero and those with an ω value of 1. Model M2 allows for three classes of sites — one with an ω value of zero, one with an ω value of one and one with an ω value that is not fixed to any value. Given the relationship between M1 and M2, they can be tested for the significance of the difference of the fit of these two models using an LRT with $df=2$. Model M3 allows all ω values to vary freely. There are two variants of this model employed in this analysis. The first is where there are two classes of sites that are free to vary ($k=2$) and the second is where there are three classes of sites ($k=3$). M3 ($k=2$) can be tested for its fit against M0 with $df=2$. M3 ($k=3$) cannot be tested against any of the other models presented here using an LRT, however, its empirical comparison with M3 ($k=2$) can be interesting if the likelihood score is better.

We have also employed a number of models that use discrete approximations to continuous distributions in order to model variability in ω at different sites. The first of these, M7, assumes that variation in ω follows a beta distribution. A total of ten classes of sites are assumed to exist and their ω values are constrained to be between zero and one. The second model, M8, allows the existence of another class of site where the ω value is allowed to be greater than unity. M8 and M7 can be compared with one another using the LRT with $df=2$.

Finally, we used two models that allow the ω value to vary across sites and across different lineages. The first of these, Model A, is a lineage-specific extension of M1 and the second, Model B, is a lineage-specific extension of M3 ($k=2$). Both of these branch-site models can be compared with their site-specific counterparts using the LRT with $df=2$. We can attempt to identify those individual sites where positive selection has occurred using the empirical Bayes approach implemented in the software (Yang and Nielsen, 2002).

For all models used in the analysis where ω is estimated from the data, a variety of starting ω values were used for the calculation of likelihood estimates. This ensures that the global minimum is reached. The significance level for positively selected sites is taken as greater than 5% of any given alignment throughout the manuscript.

3. Results and discussion

3.1. FAS gene analysis

The alignment of the *FAS* gene resulted in 7581 aligned nucleotide positions (2527 codon positions). Our initial analysis involved using the original maximum likelihood codon-based model (Goldman and Yang, 1994). For this model, ω is assumed to be constant for all sites and all positions and was estimated to be 0.116, indicating that, on average, the ratio of non-synonymous to synonymous substitution is less than unity and reflecting that change in this protein is overall selected against. Relaxing the constraint that all lineages should have the same ω value, the human lineage was allowed to have a different ω value to the other two lineages. In this case the estimated values for ω in the human and mouse lineages were 0.099 and 0.118 for the 'background' lineages, respectively, and 0.206 and 0.09 for the human and mouse 'foreground branches', respectively, again suggesting strong purifying selection on average. The consequent decrease in the log-likelihood score was significant ($p<0.05$).

We then examined models of sequence evolution where the ω value is allowed to vary from site to site. These are indicated using the notation M1 to M8 in Table 2. M1, which has two classes of sites, one class where the ω value is fixed at zero, implying that changes at these sites are unable to occur, and another class of sites where the ω value is fixed at unity. This model is the poorest fit to the data of any model we have examined ($\ln L = -28,289.8085$). Model M2, allows an additional class of sites to exist, with ω estimated from the data. This model was a much better fit to the data with the likelihood score increasing by more than 587 units. This indicates that models that allow variation across the alignment provide a better description of the data than models that do not.

The next two models use an unconstrained discrete distribution in order to model variability at different sites in the alignment. The first of these models, M3 ($k=2$) allows two classes of sites to exist in the alignment and does not constrain any ω value. This model results in a likelihood score ($\ln L = -27,707.431$) that is not lower than the score obtained from model M2 ($\ln L = -22,702.463$). However, when we allow three classes of site (M3 $k=3$), we recover a likelihood score that is better than the simpler model (M3 $k=2$) by greater than seven likelihood units. Both models are significantly better than model M0 ($2 \Delta l=748$, $p<0.0005$, $df=2$ for $k=2$ site classes, and $2 \Delta l=746$ with $p<0.0005$ and $df=4$ for $k=3$ sites classes). Of interest is model M3 $k=3$. This is the site-specific model that gives the highest likelihood. It also indicates that the third class of sites has an ω value that is far less than unity ($\omega=0.609$). This proportion of sites is approximately 12%. Model M7 does not allow for positively selected sites, but allows variation in ω across sites in the interval (0,1). The likelihood score for this model is not particularly high ($\ln$

Table 2
Parameter estimates for fatty acid synthase gene

Model	p	$\ln L$	Estimates of parameters	Positively selected sites
M0: one ratio	1	$-28{,}081.5924$	$\omega=0.116$	None
Site-specific				
M1: neutral	1	$-28{,}289.809$	$p_0=0.53$	Not allowed
M2: selection	3	$-27{,}702.463$	$p_0=0.43$, $p_1=0.05$, $(p_2=0.51)$, $\omega_2=0.19$	None
M3: discrete ($k=2$)	3	$-27{,}707.431$	$p_0=0.62$, $(p_1=0.38)$, $\omega_0=0.02$, $\omega_1=0.34$	None
M3: discrete ($k=3$)	5	$-27{,}699.505$	$p_0=0.41$, $p_1=0.47$, $p_2=0.12$, $\omega_0=0$, $\omega_1=0.16$, $\omega_2=0.61$	None
M7: beta	2	$-27{,}702.395$	$p=0.38$, $q=2.116$	Not allowed
M8: beta and omega	4	$-27{,}701.241$	$p_0=0.76$, $p=0.74$, $q=3.03$, $(p_1=0.24)$, $\omega=0$	None
Branch-specific				
Human				
Two-ratios	2	$-28{,}061.095$	$\omega_0=0.099$, $\omega_1=0.206$	None
Mouse				
Two-ratios	2	$-28{,}080.279$	$\omega_0=0.118$, $\omega_1=0.091$	None
Branch-site				
Human				
Model A	3	$-28{,}184.953$	$p_0=0.533$, $p_1=0.41$, $(p_2=0.03$, $p_3=0.025)$, $\omega_2=10.52$	$119>0.5$
Mouse				
Model A	3	$-28{,}170.318$	$p_0=0.053$, $p_1=0.047$, $(p_2=0.48$, $p_3=0.42)$, $\omega_2=0.02$	None
Human				
Model B	5	$-27{,}629.452$	$p_0=0.60$, $p_1=0.34$, $(p_2=0.04$, $p_3=0.02)$, $\omega_0=0.02$, $\omega_1=0.32$, $\omega_2=5.84$	$101>0.5$
Mouse				
Model B	5	$-27{,}701.555$	$p_0=0.37$, $p_1=0.22$, $(p_2=0.26$, $p_3=0.15)$, $\omega_0=0.02$, $\omega_1=0.36$, $\omega_2=0$	None

$L=-27{,}702.395$). However, model M8 is not significantly better (2 $\Delta l=2.3$). It is important to remember here that these models do not allow for lineage-specific evolutionary rates and as the preliminary analysis with the 2 ratio model indicated there may be a lineage-specific rate variation in the *FAS* gene.

The next category of models that were used were those branch-site models described by Yang and Nielsen (2002). These are designated 'Model A' and 'Model B' in Table 1. The human branch and mouse branch were labeled in turn as "foreground" for both Model A and Model B with the rest of the phylogeny treated as "background". Model A is an extension of the site-specific neutral model (M1) and can be compared with it using a χ^2 test with 2 degrees of freedom. The difference in log-likelihood value is significant in both human and mouse cases (2 $\Delta l=209$, $p=0$ for human, 2 $\Delta l=239$, $p=0$ for mouse). This indicates that site- and lineage-specific models provide a better description of the data, indicating that the molecule has evolved differently in different lineages.

Model B is the model that results in the best likelihood score of any model examined. It is significantly better than the discrete model M3 ($k=2$), with which it can be compared using a χ^2 test (2 $\Delta l=155.957$, $p=0$, $df=2$ for human). This model also indicates that some sites are evolving under the influence of positive selection, specifically in the human lineage. A total of 6% of sites have an ω value of 5.84 in the human lineage. This is in contrast to model M3 ($k=2$) itself and M3 ($k=3$) neither of which indicate site-specific positive selection. Model B for the

mouse lineage is also significantly better than M3 ($k=2$) (2 $\Delta l=11.75$, $p<0.005$, $df=2$), this model indicates strong purifying selective pressure in the mouse lineage for this protein. The lowest likelihood score of all models tested is for model B with the human branch labeled as foreground. The inference is that the molecule has been under an increased positive selective pressure in the human lineage.

Interestingly, of the 15 sites that have a posterior probability greater than 0.95 of being in the class of sites under positive selection, many are very close to known active sites of the protein, see Fig. 2(b). Five substitutions are of particular interest. One of the substitutions identified as having a high likelihood of being under positive selection in the human lineage ($p<0.05$) occurs at position 877 and is located adjacent to the β-hydroxyacyl dehydratase active site of this protein at position 876. This is a change from a polar cysteine residue (in chicken, mouse, rat and cow) to a non-polar Threonine (in human). Although we can only speculate on the functional significance of this change, its radical nature and close proximity to one of the most important residues in the entire protein make this an interesting event.

In addition, two substitutions occur at sites that are close to the Malonyl transferase dehydratase active site at position 580 that is responsible for the conversion of malonyl-CoA and ACP to CoA and malonyl-ACP. This is the initial reaction in de novo fatty acid synthesis, see Fig. 2(a). These substitutions, which are located at positions 587 and 588 involve a change from a polar, neutral Tyrosine (in chicken, mouse, rat and cow) to a non-polar, neutral Valine (in
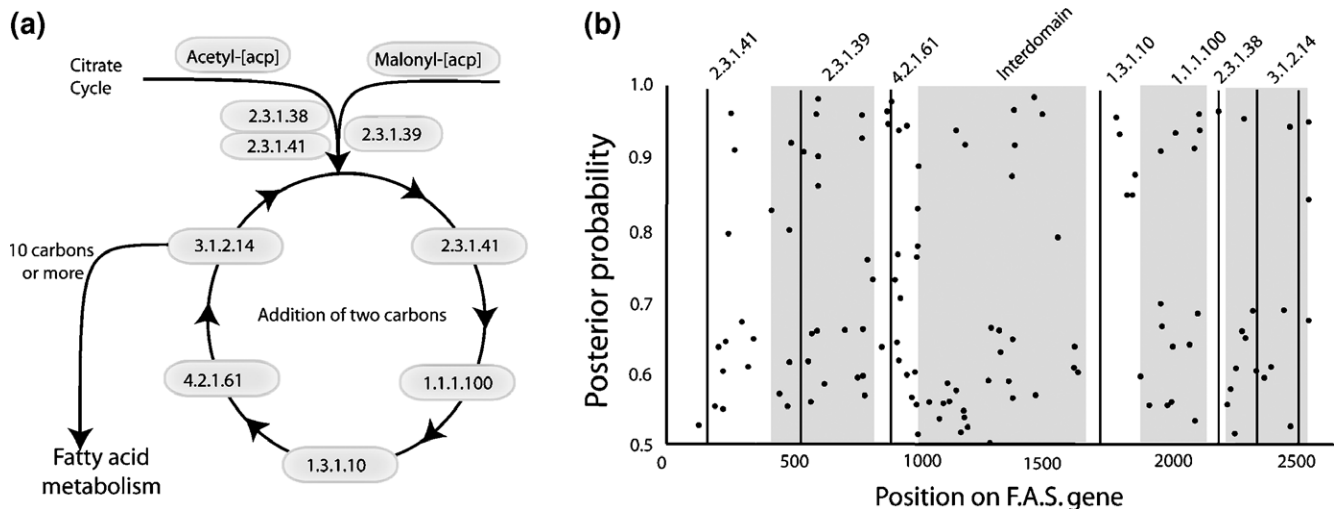
Fig. 2. (a) A representation of the Fatty Acid Synthase pathway taken from the KEGG web page http://www.genome.ad.jp/kegg/kegg2.html. Outlines the process of long chain fatty acid synthesis from acetyl-CoA, malonyl-CoA and NADPH. (b) Plot of the exact location of the positively selected sites and where they occur. The seven classifications across the top of the plot represent the seven major catalytic domains of the protein, the unlabelled 649 amino acid region in the center of the molecule represent the interdomain (ID). From the left of the chart the following are the domains of FAS, BKA (Beta-ketoacyl synthase), AMT (Acyl and Malonyl Transferase), Beta-hydroxyacyl dehydratase (DH), ER (Enoyl Reductase), BKAR (Beta-Ketoacyl Reductase), AC (Acyl carrier) and TE (Thioesterase), each of which has at least one of these sites within its boundaries. The vertical bars represent the active sites and binding sites of FAS. The 5 active sites are as follows: Beta-Ketoacyl synthase at position 161, Malonyltransferase at position 580, Beta-Hydroxyacyl dehydratase at position 876, Thioesterase at position 2302 and Thioesterase at position 2475. The two binding sites are Pyridoxal phosphate at position 1699 and Phosphopantetheine at position 2151.

human) and a non-polar, tiny Alanine (chicken, mouse, rat and cow) to an large, basic Arginine (in human) respectively. At position 2155, which is four residues away from the phosphopantetheine binding site, there is another radical substitution from a Glutamate to Proline in the human lineage. The last replacement substitution that occurs in this protein is a change from a neutral, non-polar Valine to a basic, large Arginine at position 2501, on the boundary of the last domain (Thioesterase) of FAS. However it has been shown that positions identified in this way should be treated with some caution (Zhang, 2004).

The GC content between mammals and birds varies substantially and is known to influence the mutation rate (Eyre-Walker and Hurst, 2001). The GC3 content of the human, mouse, rat, cow and chicken for the *FAS* gene are 85%, 69%, 67%, 82% and 47%, respectively. However, we have conducted the analysis of the FAS dataset with the exclusion of the highly divergent chicken sequence. Considering the FAS data for the mammalian species only the results indicate that positive selection has influenced the evolution of this protein, with an estimated ω value of 8.98 in the human lineage in 5% of the sites while the mouse lineage shows evidence of strong purifying selection ($\omega = 0.02$). This approach reduces the concern over the use of a very divergent outgroup species. However, it is important to keep in mind that the result may be due to an overall increase in mutation rate in the human lineage.

The genomic location of the *FAS* gene may be influencing the mutation rate rather than a shift in selective pressure. To further investigate this possibility we have analysed four neighboring genes in the human and mouse genomes. These

genes are listed at the bottom of Table 1, they are RAS-related C3 botulinum toxin substrate 3 (RAC3), Beta-1,2-*N*-acetylglucosaminyltransferase radical fringe (RFNG), Dicarbonyl/L-xylulose reductase (DCXR), and Cop9 signalosome complex subunit 1 (GPS1). The GC3 content for RAC3 is as follows: Human 85%, Mouse 70%, Rat 56%, Cow 66% and Zebrafish 52%. The GC3 content for RFNG is as follows: Human 88%, Mouse 70%, Rat 69%, Chicken 63% and Zebrafish 60%. The GC3 content for GPS1 is as follows: Human 84%, Orangutan 83%, Mouse 65%, Rat 67%, and Chicken 52%. The GC3 content for DCXR is as follows; Human 80%, Mouse 70%, Golden Hamster 74%, Rat 69%, Guinea Pig 70% and Frog 46%. The variation in GC3 content in these neighboring genes is very similar to that found in the *FAS* gene dataset. Analyses of these neighboring genes will indicate whether the GC3 content in this region is contributing to the detection of false positive results. However, we find that the analysis of these genes does not identify significant levels of positive selection (as defined as 5% or more of sites in the alignment), thereby strengthening the argument for positive selection in the *FAS* gene (see supplementary information for likelihood scores and parameter estimates).

3.2. Analyses of interacting proteins

Although the adaptive evolutionary events we see in the FAS protein may be due to a number of reasons, it is difficult to ascribe causality to this effect. It is obvious that primate lineages were influenced by changes in this gene, with clear evidence for selection for those lineages that had

specific amino acid changes. In order to examine the possibility that the role of FAS in cancer development was the cause of positive selection in this gene, we sought to examine cancer-implicated genes that are known to interact with FAS. We analysed MAP, PTEN, PI-3K and SREBP1 using the same methods described previously. We wished to test the hypothesis that if cancer had been a major influence on primate evolution, there may have been some evidence for adaptive evolution in many genes involved in the same processes.

The same models of codon evolution were employed and the same LRTs were carried out on a variety of gene sequences whose proteins are known to interact with *FAS*, what follows is a description of the maximum likelihood model that best fit the data for each protein-coding gene. The finding was that, for the most part, these proteins have remained under strong purifying selection for the entirety of the time period under examination. The results of these analyses are summarized in Table 3 and the likelihood model for each gene in the study is described below, the full details of parameter estimates for each of these genes is given in the supplementary information.

For MAPK1 the best model was the site-specific selection model M2. Under this model, it is estimated that there are 95% of the sites under strong purifying selection with an ω value of 0, no sites neutrally evolving, and 5% of sites with an ω value estimate from the data of 0.30. For PI-3K, the maximum likelihood model is the one ratio model. Under this model, the average ω value for the entire alignment is 0.025, indicating strong purifying selection. The maximum likelihood model for ACAD is model B with the human branch labeled as foreground, and under this model it is estimated that all sites in the human lineage have an ω value of 0.17444. This indicates strong purifying selection. The LCPT protein is for the most part under strong purifying selection (likelihood model is the discrete model with two site classes) with an ω value of 0.0547, however there are a tiny proportion of the sites (1%) showing evidence of positive selection with an ω value estimated as 4.08. From Bayesian statistics it is estimated

that 6 sites have greater than 0.5 posterior probability of being under positive selection, 3 of these sites with strong support (0.95 posterior probability). The likelihood model for LFABP is the discrete model with 2 site classes (M3 $k=2$). Under this model, both proportions of sites and ω values are estimated from the data. It is estimated that 66% of the sites in the LFABP protein are under purifying selection with an ω value of 0.019 and 26% of sites are evolving almost neutrally with an ω value of 0.8. The likelihood model for SREBP1 is the one ratio model with ω estimated as 0.247 for the entire alignment. For the SREBP2 protein the likelihood model is the discrete model with three site classes. The majority of the SREBP2 protein is under strong purifying selection with 60.5% of the alignment under strong purifying selection ($\omega=0.0001$), and 39.2% of the alignment under slightly weaker purifying selection with an estimated ω value of 0.201. There is a very tiny proportion, only 0.03% of the protein estimated to have an ω value of 5.9. Finally the results of the analysis of the FAC protein family indicated that the site-specific model for the data was the discrete model with 3 site classes showing purifying selection of various strengths across the protein. There are 25% of sites with an ω value estimated as 0.009, 51% of sites with an ω value estimated as 0.09 and 23% of the protein has an ω value estimated as 0.31. The phylogenetic tree used in the analysis of FAC is shown in Fig. 1. Each of the labeled branches was analysed in turn. For branches 1, and 3 on the phylogeny the likelihood model is model B, showing strong purifying selection across 98.5% and 98% the protein for branches 1 and 3, respectively. However, for both branch 1 and 3 there are 1.5% and 2% of the protein, respectively, showing signatures of positive selection.

The results show that considering all the known proteins that interact with FAS in the cancer-causing pathway the dominant selective pressure is purifying and this has been the case for the entirety of the time period under examination. None of these other proteins have any substantial evidence of positive selection at an appreciable number of sites.

Table 3
Genes known to interact with FAS and its role in Cancer, or metabolism

Model	MAPK1	PI	ACAD	LCPT	LFABP	SREBP1	SREBP2	FAC	FAS
M0: one ratio	None	None	None	None	None	None	None	None	None
2 ratios	None	None	None	None	None	None	None	None	None
M2: selection	None	None	None	None	None	None	None	None	None
M3: discrete $k=2$	None	None	None	None	None	None	None	None	None
M3: discrete $k=3$	None	None	None	None	None	None	None	None	None
M8: beta and omega	None	None	None	None	None	None	None	None	None
Model A	None	1	None	None	None	None	None	None	None
Model B	None	1	None	None	None	None	None	None	101

The left-most column represents those models used and described in the Materials and methods section which allow for positive selection. For each gene, the results of adaptive evolution analysis using maximum likelihood analysis is listed in terms of whether sites were identified as having high probability of being in a class of sites likely to be under positive selection. *FAS* is the only gene which shows a significant number of sites under positive selection. Using Model B which best fits the data, 101 sites are identified with probability greater than 0.5 of being under positive selection, 15 of these have extremely high posterior probabilities of greater than 0.95.

For this protein interaction network, the FAS protein is unique in terms of its evolutionary history (see Table 3). The number of vertebrate genes displaying evidence of adaptive evolution remains small, e.g., ape cytochrome *c* oxidase (Goldberg et al., 2003), lysozyme of the foregut fermenters (Messier and Stewart, 1997), human SRY (Wang et al., 2002). This low number is either a reflection of the sparse sampling of vertebrate genomes or because of the relative infrequency of adaptive events within a phylum or simply because analyses of this kind have been rare. However, in this manuscript we report strong evidence for positive selection in a gene that plays a central role in metabolism and in the development of many cancers. Of particular significance is the fact that similar analyses of a variety of other genes that control the expression of FAS or interact with it have revealed only evidence for strong purifying selection in these genes. Therefore, we suggest that it is not changes in the control of FAS that has conveyed some advantage, but rather it is the function of this gene product. Analysis of genes at the same locus as the *FAS* gene do not show significant levels of positive selection, the mutation rate in this gene is consistent with that of an adaptive evolutionary event rather than the influence of GC3 content or genome location.

In summary, the selective pressures on the fatty acid synthase gene have been highly variable among sites and in different lineages. Some sites have been under strong purifying selection throughout the phylogeny, while other sites appear to be under positive selection only in the human lineage, with the number of amino acid altering substitutions in excess of the silent substitutions. The branch leading to the human lineage is under much stronger positive selective pressure than the remainder of the phylogeny, with no sites in the background lineages showing evidence of positive selection. According to the likelihood model for the *FAS* gene, Model B, 101 sites have a probability greater than 0.5 of having been positively selected in the human lineage. This finding is not incompatible with, yet does not prove the fat utilization hypothesis (Erren and Erren, 2004; Horrobin, 1998, 1999), which suggests that changes in phospholipid and fatty acid metabolism played a major role in human evolution, namely the development of the large human brain. It is also important to note that this method has reduced power with such a small sample size, in this case four or five sequences, and as such to find positive selection it must show a significantly strong signal (Anisimova et al., 2001). With more FAS sequences from other species future analysis will provide even more insight into the evolution of this central metabolic gene.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2005.06.020.

References

Anisimova, M., Bielawski, J.P., Yang, Z., 2001. Accuracy and power of the Likelihood Ratio Test in detecting adaptive molecular evolution. Mol. Biol. Evol. 18, 1585–1592.

Chirala, S.S., Jayakumar, A., Gu, Z.W., Wakil, S.J., 2001. Human fatty acid synthase: role of interdomain in the formation of catalytically active synthase dimer. Proc. Natl. Acad. Sci. U. S. A. 98, 3104–3108.

Duplus, E., Forest, C., 2002. Is there a single mechanism for fatty acid regulation of gene transcription? Biochem. Pharmacol. 64, 893–901.

Erren, T.C., Erren, M., 2004. Can fat explain the human brain's big bang evolution? Horrobin's leads for comparative and functional genomics. Prostaglandins Leukot. Essent. Fat. Acids 70, 345–347.

Eyre-Walker, A., Hurst, L.D., 2001. The evolution of isochores. Nat. Rev. Genet. 2, 549–555.

Felsenstein, J., 1993. PHYLIP: Phylogeny Inference Package. Distributed by Author, Seattle.

Goldberg, A., Wildman, D.E., Schmidt, T.R., Huttemann, M., Goodman, M., Weiss, M.L., Grossman, L.I., 2003. Adaptive evolution of cytochrome *c* oxidase subunit VIII in anthropoid primates. Proc. Natl. Acad. Sci. U. S. A. 100, 5873–5878.

Goldman, N., Yang, Z., 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11, 725.

Horrobin, D.F., 1998. Schizophrenia: the illness that made us human. Med. Hypotheses 50, 269–288.

Horrobin, D.F., 1999. Lipid metabolism, human evolution and schizophrenia. Prostaglandins Leukot. Essent. Fat. Acids 60, 431–437.

Jayakumar, A., et al., 1995. Human fatty acid synthase: properties and molecular cloning. Proc. Natl. Acad. Sci. U. S. A. 92, 8695–8699.

Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

Kumar-Sinha, C., Ignatoski, K.W., Lippman, M.E., Ethier, S.P., Chinnaiyan, A.M., 2003. Transcriptome analysis of HER2 reveals a molecular connection to fatty acid synthesis. Cancer Res. 63, 132–139.

Leroi, A.M., Koufopanou, V., Burt, A., 2003. Cancer selection. Nat. Rev. Cancer 3, 226–231.

Messier, W., Stewart, C.B., 1997. Episodic adaptive evolution of primate lysozymes. Nature 385, 151–154.

Nohturfft, A., DeBose-Boyd, R.A., Scheek, S., Goldstein, J.L., Brown, M.S., 1999. Sterols regulate cycling of SREBP cleavage-activating protein (SCAP) between endoplasmic reticulum and Golgi. Proc. Natl. Acad. Sci. U. S. A. 96, 11235–11240.

O'Brien, J.S., 1967. Cell membranes—composition: structure: function. J. Theor. Biol. 15, 307–324.

Pizer, E.S., Pflug, B.R., Bova, G.S., Han, W.F., Udan, M.S., Nelson, J.B., 2001. Increased fatty acid synthase as a therapeutic target in androgen-independent prostate cancer progression. Prostate 47, 102–110.

Qiao, S., Pennanen, P., Nazarova, N., Lou, Y.R., Tuohimaa, P., 2003. Inhibition of fatty acid synthase expression by 1alpha,25-dihydroxyvitamin D3 in prostate cancer cells. J. Steroid Biochem. Mol. Biol. 85, 1–8.

Roncari, D.A., 1981. Fatty acid synthase from human liver. Methods Enzymol. 71, 73–79.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Van de Sande, T., De Schrijver, E., Heyns, W., Verhoeven, G., Swinnen, J.V., 2002. Role of the phosphatidylinositol 3ʹ-kinase/PTEN/Akt kinase pathway in the overexpression of fatty acid synthase in LNCaP prostate cancer cells. Cancer Res. 62, 642–646.

Wakil, S.J., 1989. Fatty acid synthase, a proficient multifunctional enzyme. J. Biochem. 28, 4523–4530.

Wang, X., Zhang, J., Zhang, Y.P., 2002. Erratic evolution of SRY in higher primates. Mol. Biol. Evol. 19, 582–584.

Yang, Z.H., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comp. Appl. Biosci. 13, 555–556.

Yang, Z., Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. 19, 908–917.

Yang, Y.A., Han, W.F., Morin, P.J., Chrest, F.J., Pizer, E.S., 2002. Activation of fatty acid synthesis during neoplastic transformation: role of mitogen-activated protein kinase and phosphatidylinositol 3-kinase. Exp. Cell Res. 279, 80–90.

Yang, Y.A., et al., 2003. Regulation of fatty acid synthase expression in breast cancer by sterol regulatory element binding protein-1c. Exp. Cell Res. 282, 132–137.

Zhang, J., 2004. Frequent false detection of positive selection by the likelihood method with branch-site models. Mol. Biol. Evol. 21 (7), 1332–1339.

Zhou, W., et al., 2003. Fatty acid synthase inhibition triggers apoptosis during S phase in human cancer cells. Cancer Res. 63, 7330–7337.