Journal of Genetics (2020)99:1
https://doi.org/10.1007/s12041-019-1160-8

RESEARCH NOTE

Signatures of positive selection on the hepatic lipase gene in human populations

SERGIO V. FLORES[1,2]*, CARLA F. OLIVARI[1] and LUIS FLORES PRADO[3]

[1]*Laboratorio de Genética y Bioantropología, Departamento de Antropología, Facultad de Ciencias Sociales, Universidad de Chile, Avenida Ignacio Carrera Pinto 1045, Ñuñoa 7750000, Santiago, Chile*
[2]*Present Address: Labortorio de Genómica Evolutiva, Ñuñoa 7750000, Santiago, Chile*
[3]*Instituto de Entomología, Universidad Metropolitana de Ciencias de la Educación, Ñuñoa 7750000, Santiago, Chile*
*For correspondence. E-mail: sergiovladimirflores@gmail.com.

Received 13 June 2018; revised 20 June 2019; accepted 17 September 2019

Abstract. The hepatic lipase plays a central role in the lipid metabolism, catalyzing the hydrolysis of phospholipids, monoglycerides, diglycerides, and triglycerides, and acyl-CoA. It is also implied in the conversion of very low-density lipoprotein and intermediate density lipoprotein to low density lipoproteins. As a consequence, the gene encoding the hepatic lipase (LIPC) is associated with several diseases derived from the imbalance of lipids that are in general derived from the interaction between life styles and genetic architecture. Therefore, it is interesting to understand more about the characteristics of the microevolutionary processes affecting genes that, like LIPC, have a role in nutrition and lipid metabolism in human populations. We explored the selection signatures on LIPC in 26 populations, detecting three regions under recent positive selection.

Keywords. hepatic lipase; positive selection; Tajima's *D*; Fu and Li test; 1000 genomes.

Introduction

Hepatic lipase (LIPC) is a member of the lipase gene family (Asuka and Aoki 2006) and is a cluster of genes with extracellular phospholipase A1 activity. Holmes *et al.* (2011) reported that the origin of LIPC predated fish in the vertebrate evolution. The same authors also found a relatively high rate of evolution (2–3 times) affecting LIPC in comparison with lipoprotein lipase (LPL) in vertebrates, according to branch lengths from phylogenetic analysis using protein sequences. This observation could be interpreted as a balance between positive selection on LIPC favouring aminoacidic divergence, and/or negative selection on LPL. Thus, neutral evolution appears to be refuted and, therefore, natural selection seems to have had a role in the evolution of lipase genes.

Signals of positive selection on lipid catabolism genes have been found in contemporary Europeans from Neanderthal ancestry alleles. Khrameeva *et al.* (2014) found that variants shared between humans and Neanderthal are specifically enriched in genes involved in lipid catabolism in contemporary Europeans. In other study, Metspalu *et al.* (2011) scanned high haplotype homozygosity regions in Indian populations, founding signals of positive selection on MSTN and DOK5, two loci associated with lipid metabolism and type 2 diabetes. At least two other studies using genomic data have found evidence of positive selection on genes associated to lipid metabolism (reviewed in Luca *et al.* 2010). However, none of these studies have found signals of selection on the *LIPC* gene, or in other genes belonging to the lipase gene families, that in the case of array-based genotype data is perhaps due to low coverage data around LIPC.

By using the data publically available, here we explored the signatures of positive natural selection on the *LIPC* gene in human populations to understand the evolutionary trajectories of lipase genes that could contribute to explain the current diversification of lipid metabolism in human populations.

Methods

SNP data for the *LIPC* gene in 26 human populations (n = 2504) were collected from the 1000 genome phase 3 project, comprising the five 'super-populations'. Details on

Published online: 18 December 2019

samples are showed in table 1. SNPs were filtered to a minor allele frequency (MAF) to 0.001, using VCFTOOLS (Danecek *et al.* 2011). Variant call format (VCF) files were transformed to obtain phased haplotypes in FASTA format for each individual by using two scripts written in the R statistics package.

Population genetic estimates of F_{st} nucleotide diversity (pi) and heterozigosity (H) were computed for each population using VCFTOOLS. F_{ST} for populations were estimated comparing each population with the remaining samples for the respective super-population. F_{ST} for specific super-populations were obtained comparing each in relation to the remaining sample.

To test the neutral molecular evolution, we used different 'site frequency spectrum' summary statistics: Tajima's D (Tajima 1989), and Fu and Li (1993) tests were estimated in DNAsp (Rozas *et al.* 2003), over the final 831 SNPs output from the filtering, using window-size and window-step of 20 and five SNPs, respectively. Finally, the extended haplotype homozygosity (EHH) was estimated using the REHH software for the candidate loci to detect signals of positive selection.

Results and discussion

Population parameters for each population are summarized in table 1. These results indicate moderate genetic differentiation (F_{ST}), as compared in human populations by previous studies (Nelis *et al.* 2009). However, high differentiation was found in the Peruvian, Puerto Rican, Finland and Japanese populations, in relation to their respective super-populations.

Nucleotide divergence (pi) and heterozygosity (H), showed significant low values as compared with the estimates from a 2.1 Mb region overlapping LIPC (chr.15: 59568578-57410802); $H = 0.26$ (CI 95% = 0.26–0.26), pi = 0.20 (CI 95% = 0.19–0.20). In general, the admixed American (AMR) and South Asian population (SAS) showed heterozygosity significantly lower than the other super-populations ($P_{kruskall–Wallis} < 0.001$).

Results of the neutrality test are summarized in figure 1a, where the Fu & Li D* test values are presented. The results accounting for positive selection (i.e. negative values) here were also statistically significant by the Fu & Li F*, Tajima's D and the Way and Fu tests.

A general pattern of positive selection across the *LIPC* gene was obtained for the Peruvian population (PEL) as compared with all the other populations ($P_{Kruskall–Wallis-} < 0.01$). Also, a similar result was obtained for the Spaniard population (IBS) ($P_{Kruskall–Wallis} < 0.05$). Our results suggest that positive selection along the *LIPC* gene was intensified in the Peruvian and Spaniard populations, probably the interaction with demographic events in the past (i.e. population expansion). We did not address demographic factors in this work and further analyses, like those based on haplotype network structure are necessary to test the selection/demography interaction.

Population-specific patterns of positive selection

A gene region showing low and significant values ($P < 0.01$) for the Fu and Li* test was found in the intron 3 (chr.15: 58468445–58477563), observed in the South Asia (GWD) and African (YRY) populations. That signal overlaps a promoter flanking region (ID ENSR00001452587). The EEH analysis showed evidence of positive selection covering almost 80 kb, flanking the derived allele (A) on the SNP rs1869136 (figure 1b).

Another signal on intron 3 (chr.15: 58525201–58529000), was found in Han Chinese in Beijing, China (CHB) and Kinh in Ho Chi Minh city, Vietnam (KHV) (Asian populations), the IBS European population, and the PEL admixed American population. This signal overlaps an enhancer, which was verified by analysing this region in the Promoter 2.0 tool (www.cbs.dtu.dk/services/Promoter). An extended homozygosity region, comprising ~20 kb, flanking the ancestral allele (G), was found around the SNP rs58829990 (figure 1b).

Finally, the stronger signal of positive selection was observed between introns 8 and 9, in the ITU and PEL populations (Fu & Li D* = − 4.49 and − 3.12 respectively; $P < 0.01$). The minimum value (Fu & Li D* = − 7.47) was found at the centre of the window, containing the two SNPs sampled in exon 9 (ENSE00003549295); rs3829462 and rs3829461. From these, rs3829462 is a biallelic (C/A), missense variant (TT**A**/ TT**C**), determining the presence of phenylalanine (ancestral) or leucine (derived) in LIPC, at the position 355 in four transcripts (ENST00000299022.9, ENST00000356113.10, ENST00000414170.7 and ENST00000433326.2).

The ancestral variant (C) is the minor allele frequency (MAF) allele for this SNP, showing a global frequency of 6%; 1% in SAS, EUR and AMR); 17% in AFR; and 6% in EAS. In the ITU population, the frequency of this allele is 0.5%, corresponding to the unique allele from the total chromosome sample ($N = 204$).

The EEH analysis showed evidence of positive selection covering almost 25 kb flanking the derived allele (A) on the SNP rs3829462 (figure 1b).

Interestingly, the derived allele (A) corresponds to the risk allele, associated with hyperlipidaemia due to hepatic triglyceride lipase deficiency (Carlquist *et al.* 2011). The change of the ancestral by the derived aminoacid was predicted to be benign, with a score of 0.006 (sensitivity: 0.97; specificity: 0.75), according to PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/index.shtml).

A challenging question is, why this risk allele has been the target of positive selection in specific populations, and increasing its frequency globally. One hypothesis is that this allele has/had other beneficial consequences, still not revealed, operating either in the past or currently in human populations, perhaps in interaction with changes in dietary or lifestyle patterns.

Table 1. Sample information and statistics of populations and super populations used in this study.

Super population (code)	Population description	Population code	Number of chromosomes	F_{ST}*	Pi	H
Africa (AFR)	African Caribbeans in Barbados	ACB	192	0.00033	0.001073	0.236
	Americans of African Ancestry in SW, USA	ASW	122	0.00482	0.001093	0.278
	Esan in Nigeria	ESN	198	0.00101	0.001063	0.217
	Gambian in western divisions of Gambia	GWD	226	0.00519	0.001045	0.197
	Luhya in Webuye, Kenya	LWK	198	0.00600	0.001075	0.285
	Mende in Sierra Leone	MSL	170	0.00182	0.001069	0.256
	Yoruba in Ibadan, Nigeria	YRI	216	0.00068	0.001075	0.285
	Total		1322	0.00499	0.001070	0.251
America (AMR)	Colombians from Medellin, Colombia	CLM	188	0.01030	0.000934	0.185
	Mexican Ancestry from Los Angeles, USA	MXL	128	0.00411	0.000920	0.226
	Peruvians from Lima, Peru	PEL	170	0.04387	0.000884	0.193
	Puerto Ricans from Puerto Rico	PUR	208	0.02147	0.000958	0.199
	Total		694	0.03089	0.000924	0.206
East Asian (EAS)	Chinese Dai in Xishuangbanna, China	CDX	186	0.00184	0.000962	0.252
	Han Chinese in Beijing, China	CHB	206	0.00004	0.000989	0.199
	Southern Han Chinese	CHS	210	0.00282	0.000922	0.231
	Japanese in Tokyo, Japan	JPT	208	0.01472	0.000950	0.219
	Kinh in Ho Chi Minh city, Vietnam	KHV	198	0.00187	0.000930	0.220
	Total		1008	0.00697	0.000950	0.224
Europe (EUR)	Utah residents (CEPH) with northern and western European ancestry	CEU	198	0.00056	0.000924	0.207
	Finnish in Finland	FIN	198	0.00619	0.000966	0.245
	British in England and Scotland	GBR	182	0.00021	0.000917	0.196
	Iberian population in Spain	IBS	214	0.00152	0.000930	0.211
	Toscani in Italia	TSI	214	0.00196	0.000915	0.205
	Total		1006	0.00326	0.000930	0.213
South Asians (SAS)	Bengali from Bangladesh	BEB	172	0.00381	0.000994	0.155
	Gujarati Indian from Houston, Texas	GIH	206	0.00323	0.001045	0.201
	Indian Telugu from the UK	ITU	204	− 0.00002	0.000972	0.178
	Punjabi from Lahore, Pakistan	PJL	192	0.00582	0.000922	0.183
	Sri Lankan Tamil from the UK	STU	204	0.00220	0.000965	0.150
	Total		978	0.00480	0.000980	0.178

Population genetic estimates of F_{ST} (Weir and Cockerham); pi, nucleotide diversity; H, heterocigosity. F_{ST} for populations were estimated comparing each population with the remaining samples for the respective super population. F_{st} for super populations were obtained comparing each super population with the other four.

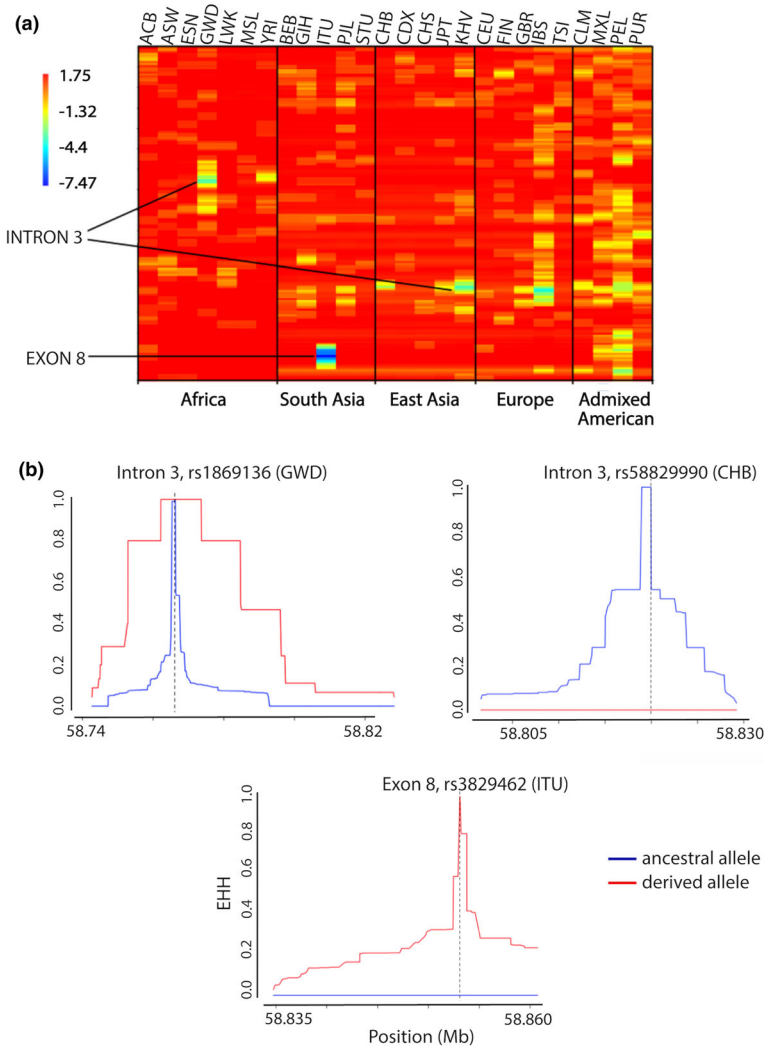

Figure 1. Signatures of positive selection on LIPC. (a) Matrix plot showing the Fu and Li* test values for each population. Genomic regions with significant values, rejecting neutral evolution, are indicated. (b) Plots of the EHH test using candidate SNPs as focal variants, showing moderate signals of positive selection for rs1869136 and rs3829462, but not for rs58829990. Populations are indicated between parenthesis.

The results obtained here show that the *LIPC* gene has been under independent positive selection events operating differentially among populations, and evidence does not support continental patterns. Further studies could reveal the physiological consequences associated to the regions under positive selection on LIPC.

Acknowledgement

This work was supported by the Grant FPCI 17-0716, VID, Universidad de Chile.

References

Asuka I. and Aoki J. 2006 Phospholipase A1: structure, distribution and function. *Future Lipidol.* **1**, 687–700.

Carlquist J. F., McKinney J. T., Horne B. D., Camp N. J., Cannon-Albright L., Muhlestein J. B. *et al*. 2011 Common variants in 6 lipid-related genes discovered by high-resolution DNA melting analysis and their association with plasma lipids. *J. Clin. Exp. Cardiolog*. **2**, 138.

Danecek P., Auton A., Abecasis G., Albers C., Banks E., DePristo M. *et al.* 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158.

Fu Y. and Li W. 1993 Statistical tests of neutrality of mutations. *Genetics* **133,** 693–709.

Holmes R. S., Vandeberg J. L. and Cox L. A. 2011 Vertebrate hepatic lipase genes and proteins: a review supported by bioinformatic studies. *Bioinformatics* **3**, 85–95.

Khrameeva E., Bozek K., He L., Yan Z., Jiang X., Wei Y. *et al.* 2014 Neanderthal ancestry drives evolution of lipid catabolism in contemporary Europeans. *Nat. Commun*. **5**, 3584.

Luca F., Perry G. H. and Di Rienzo A. 2010 Evolutionary adaptations to dietary changes. *Annu. Rev. Nutr.* **30,** 291–314.

Metspalu M., Romero I. G., Yunusbayev B., Chaubey G., Mallick C. B., Hudjashov G. *et al.* 2011 Shared and unique components of human population structure and genome-wide signals of

positive selection in South Asia. *Am. J. Hum. Genet.* **89**, 731–744.

Nelis M., Esko T., Magi R., Zimprich F., Zimprich A., Toncheva D. *et al.* 2009 Genetic structure of Europeans: a view from the north-east. *PLoS One* **4**, e5472.

Rozas J., Sanchez-del Barrio J. C., Messeguer X. and Rozas R. 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.

Tajima F. 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Corresponding editor: T. N. C. VIDYA