P3-09. DUPLICATED GENES UNDER CONCERTED EVOLUTION IN *Helicobacter pylori*

Jerson Alexander García-Zea, Rafael Navajas-Pérez, Francisca Robles, Juan Felipe Zapata,Roberto de la Herrán, Carmelo Ruiz Rejón

Department of Genetics, University of Granada, Spain, Faculty of Forensic Science, Technology of Antioquia, Colombia. Corresponding Author: Jerson Alexander García-Zea (alexander7719@correo.ugr.es)

Helicobacter pylori, a Gram-negative bacillus, is present in more than half of the worldpopulation. Colonizes the human stomach inducing the development of superficialgastritis, gastric adenocarcinoma, and mucosa-associated lymphoid tissue lymphoma.
H. pylori is highly variable thanks to a great genomic dynamic promoted by mutationsand high recombination rates. Also characterized by exhibiting panmixia, genetic mosaicism, and great genomic plasticity, which indicates its ability to adapt at a micro-evolutionary level to its environment (1).

This dynamic influences in the gene duplication events and in the evolution of them. Concerted evolution (CE) between duplicate genes can occur through a genetic exchange called gene conversion (2). Few studies have analyzed the CE and the affected genes in this bacterium (3). Implementing a computational framework (including IseeCe program) (4) in 53 complete *H. pylori* genomes of five different geographical origins, 8 genes (*HopJ/Hopk, 3'-5 'exonuclease, fucosyltransferase, glycosyltransferase, oipA, RMs, 50S ribosome and cagA*) were detected showing patterns of CE. The gen *HopJ/Hopk* shows CE in the five geographical region, *3'-5 'exonuclease, fucosyltransferase* in Asia and Amerindian, *glycosyltransferase, oipA, RMs* in populations of Asia only, and finally *50S ribosome and cagA* showed CE in hybrid strains. Different test by RDP5 software evidence recombination events in the majority of the genes analyzed.

Additionally, we realized analysis of selective pressures on the groups of genes underCE by calculating the non-synonym/synonym mutations ratio (dN/dS) over the complete sequences. Proteins of these genes play crucial roles in the pathogenic life cycle of *H. pylori* such as virulence, pathogenicity, colonization of the environment, adaptation to the host.

1. Garcia-Zea et al. (2019) PeerJ 11,7:e6221
2. Wang S et al. (2018) Commun Biol. 8, 1-12
3. Pride DT et al. (2020) J Mol Biol. 316, 629-42
4. Meinersmann RJ, et al. (2010) Lett Appl Microbiol. 51, 539-545

DUPLICATED GENES UNDER CONCERTED EVOLUTION IN *Helicobacter pylori*

Jerson Alexander García-Zea, Rafael Navajas-Pérez, Francisca Robles, Juan Felipe Zapata, Roberto de la Herrán, Carmelo Ruiz Rejón
Department of Genetics, University of Granada, Spain,
Faculty of Forensic Science, Technology of Antioquia, Colombia
Corresponding author: Jerson Alexander García-Zea (alexander7719@correo.ugr.es)

Facultad de Ciencias

UNIVERSIDAD DE GRANADA

BACKGROUND

Helicobacter pylori, a gram-negative bacillus, is present in more than half of the world's population. It colonizes the human stomach, inducing the development of superficial gastritis, gastric adenocarcinoma, and mucosa-associated lymphoid tissue lymphoma. *H. pylori* is highly variable thanks to great genomic dynamics promoted by mutations and high rates of recombination. It is also characterized by exhibiting panmixia, genetic mosaicism and great genomic plasticity, which indicates its ability to adapt at a microevolutionary level to its environment **(1)**. This dynamics influences gene duplication events and their evolution, as is already evident in the genes that encode the membrane proteins babA / babB, and HomA / HomB in *H. pylori*. Concerted evolution (CE) between duplicate genes can occur through a genetic exchange called gene conversion **(2)**. Few studies have analyzed EC and affected genes in this bacterium **(3)**.

MATERIALS AND METHODS

A computational framework that included IseeCe was implemented to detect CE for gene conversion (GC) based on phylogeny and genomic synthesis (Mauve) in 53 complete *H. pylori* genomes, from five different geographic origins. Recombination was studied by RDP5 (RDP, Maxchi, Geneconv, Bootscan, Chimera, Siscan, 3seq, Lard and Phylopro). Synonymous / non-synonymous substitutions was calculated by ETE3 evol-PAML (models M0, M1, M2, M3, M7 and M8) and natural selection by MegaX (Tajima's tests of neutrality)

RESULTS and DISCUSSION

In total, eight genes were identified in 16 clusters that exhibited CE patterns in different geographical regions, belonging all of them to variable genome, except 50S ribosome-binding GTPase which is part of the central genome. The HopJ / HopK gene exhibited CE patterns in all five geographic groups (Africa, Europe, Asia, Amerinds, and hybrids). In the Amerindian strains, a duplication of the cluster HopJ / HopK can be observed **(Table 1)**. The fucosyltransferase gene showed CE in the Asian, Amerindian and hybrid strains, 3'-5 'exonuclease in the Asia and Amerindian groups. The genes for glycosyltransferase, oipA, and restriction endonuclease (hsdS) only showed CE for the specific Asian group and the oncogenic effector gene of the type IV secretion system (cagA) only in Amerindians. Finally, GTPase gene that binds to the 50S ribosome shows EC in the hybrid group **(Table 1)**.

Comparative sequence analyzes provided evidence of recombination between paralogs in 13 of the 16 clusters, evidence of balancing selection in 4 clusters for the gene hopJ/HopK and in one cluster for the gene alpha-(1,3)-fucosyltransferase and evidence of selective sweep in the rest of the clusters, supporting concerted evolution result **(Table 1)**. Both mechanisms stimule the homogenization of sequences and the functionality of the paraloges gene.
In the case of duplicate cluster, hopJ/HopK in Amerind region, we can observed different identity between clusters? in the paraloges genes, being the more identity (95%) under selection sweep and the other cluster with low identity (54%) under balancing seletion. This diference in this duplicate cluster could indicate a pseudogenization process in the lattercluster.

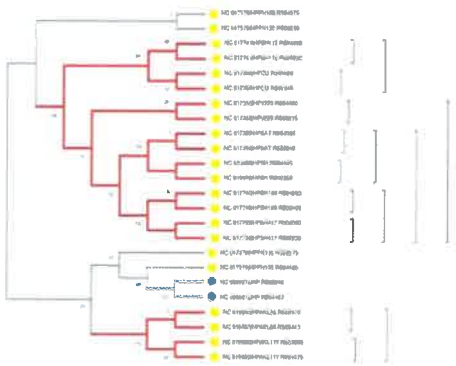

Figure 1. CE for 195_HopJ / HopK detected from IseeCe. In red duplicate genes that have evolved under CE evolution.

REFERENCES:
1. Garcia-Zea et al. (2019) PeerJ 11,7:e6221
2. Wang S et al. (2018) Commun Biol. 8, 1-12
3. Pride DT et al. (2020) J Mol Biol. 316, 629-42
4. Meinersmann RJ, et al. (2010) Lett Appl Microbiol. 51, 539-545

Table 1. Statistical summary. I%: Percentage of identity, dS / dN: synonymous substitutions and non-synonymous substitutions, Tajima's test: D = 0 (neutral equilibrium), D> (Balancing selection), D <0 (Recent selective sweep), R = recombination.

	Gene	I%	dS/dN	Tajima's Neutrality Test θ	π	D	evolutive event	R
1	15_glycosyltransferase (Asia)	75	0,40492	0.075699	0.062275	-0.668616	Recent selective sweep	yes
2	107_50S ribosome-binding GTPase (Hybrids)	89	0,47053	0.079129	0.030735	-2.959.527	Recent selective sweep	yes
3	90_restriction endonuclease subunit S (Asia)	49	1,48858	0.089179	0.044717	-1.916.201	Recent selective sweep	yes
4	30_type IV secretion system oncogenic effector cagA (Amerind)	81	1,38924	0.056251	0.036553	-1.559.445	Recent selective sweep	yes
5	54_outer inflammatory protein OipA (Asia)	98	0,23353	0.029876	0.024109	-0.735569	Recent selective sweep	yes
6	68_3'-5' exonuclease (Asia)	96	0,16825	0.032017	0.020114	-1.577.265	Recen tselective sweep	Not
	124_3'-5' exonuclease (Amerind)	97	0,16694	0.041696	0.028454	-1.257.399	Recent selective sweep	Not
7	88_alpha-(1,3)-fucosyltransferase (Asia)	82	0,22449	0.080567	0.066501	-0.691739	Recen tselective sweep	yes
	89_alpha-(1,3)-fucosyltransferase (Amerind)	88	0,98992	0.078631	0.066207	-0.647933	Recent selective sweep	yes
	106_alpha-(1,3)-fucosyltransferase Hybrids)	86	0,27713	0.086694	0.089457	0.150304	Balancing selection	yes
8	136_hopJ/HopK (Africa)	92	0,37239	0.064528	0.076976	0.962707	Balancing selection	Not
	177_hopJ/HopK (Europe)	94	0,23891	0.048842	0.060201	0.921328	Balancing selection	yes
	152_hopJ/HopK (Asia)	94	0,33026	0.053488	0.054862	0.098210	Balancing selection	yes
	156_hopJ/Hopk (Amerind)	54	0,49336	0.088041	0.092698	0.199842	Balancing selection	yes
	195_hopJ/HopK (Amerind)	95	0,25624	0.048580	0.047484	-0.090622	Recent selective sweep	yes
	134_hopJ/HopK (Hybrids)	92	0,24218	0.064525	0.076667	0.847879	Balancing selection	yes