

Diversidade alélica de dois membros da família das OMP de *Helicobacter pylori*, genes *jhp870* e *jhp649*

Allelic Diversity of two members of the Helicobacter pylori OMP family, jhp870 and jhp649 genes

Rita Cordeiro¹; Mónica Oleastro¹; Lurdes Monteiro¹

¹ U Unidade Helicobacter/Campylobacter, Centro de Bacteriologia e Micologia Prof. Doutor Arnaldo Sampaio, Instituto Nacional de Saúde Dr. Ricardo Jorge, Av. Padre Cruz, 1649-016 Lisboa, Portugal.

Resumo

Helicobacter pylori coloniza o estômago humano e está associada a várias gastropatologias.

Este trabalho teve como objectivo o estudo de dois membros da família das OMP de *H. pylori*, o gene *jhp870*, possível marcador da úlcera péptica, e o seu homólogo *jhp649*. A prevalência destes genes, a variabilidade no número de cópias, a sua localização genómica e o estudo dos polimorfismos, ao nível da sequência de bases e respectivas proteínas, foram avaliados numa amostra de estirpes clínicas de *H. pylori*, isoladas de doentes de diferentes regiões geográficas, apresentando diversas gastropatologias.

No geral, ambos os genes apresentam uma elevada similaridade quer ao nível de nucleótidos quer de aminoácidos, observando-se, no entanto, uma região média muito polimórfica. Nesta, foram identificadas seis variantes alélicas (alelo AI a AVI), com diferentes prevalências nos dois genes. O alelo AI é exclusivo e predominante do gene *jhp870*, e o AII predominante do gene *jhp649*. A observação da constância destes alelos em todas as estirpes, independentemente da sua origem geográfica, patologia, tipo de população, sugere a sua antiguidade, e evidencia a sua conservação ao longo da evolução. Mais, ao existir esta conservação numa região polimórfica, implica que estes alelos podem ter influência na função da proteína.

Palavras-Chave: *Helicobacter pylori*; OMP; *jhp870*; *jhp649*; variabilidade alélica.

Abstract

Helicobacter pylori colonizes the human stomach and is associated with various gastropathologies.

The present study aims to research two members of the *H. pylori* OMP family, *jhp870* gene, an ulcer marker candidate, and his homologue *jhp649*.

The prevalence of these genes, number of copies, genomic localization and sequence polymorphisms, were evaluated in a panel of clinical *H. pylori* strains isolated of patients from different geographic regions, presenting different gastropathologies.

Both genes displayed a high level of nucleotide and amino acid similarity. However, an extremely polymorphic region was observed, corresponding to the middle region of each gene, which shows six allelic variants (allele AI to AVI) with different prevalence in the two genes. Allele AI is exclusive and predominant of *jhp870* gene, and AII is predominant of *jhp649* gene.

The maintenance of the alleles in all strains, independently of the geographical origin, pathology and population, suggests their antiquity, and shows conservation up along the evolution. Furthermore, the existence of an allelic conservation in a polymorphic region, implicates that these alleles can influence the protein function.

Key-Words: *Helicobacter pylori*; OMP; *jhp870*; *jhp649*; allelic variability.

Recebido em 10/04/2006

Aceite em 10/05/2006

Rev. Lusófona de Ciências e Tecnologias da Saúde, 2006; (3) 1: 63-76

Versão electrónica: <http://revistasaude.ulusofona.pt>

Introdução

Helicobacter pylori (*H. pylori*) é uma bactéria gram negativa microaerófila de forma espiralada, que coloniza exclusivamente o estômago humano, localizando-se normalmente no antro e/ou no corpo, à superfície das células epiteliais^[1]. Esta bactéria está etiológicamente associada a várias gastropatologias como a dispepsia não ulcerosa, úlcera péptica, e é considerada um factor de risco no desenvolvimento do carcinoma gástrico e linfoma de MALT^[2,3].

Nos genomas, que se encontram totalmente sequenciados, das duas estirpes de referência de *H. pylori*, estirpe 26695 (isolada de um doente com gastrite)^[4] e estirpe J99 (isolada de um doente com úlcera duodenal)^[5], existe uma grande família de proteínas da membrana externa (OMP), composta por 32 membros, incluindo as proteínas BabA, HopZ e OipA, que estão implicadas na aderência da bactéria às células epiteliais gástricas^[6, 7, 8]. Outro membro desta família é codificado pelo gene *jhp870*. Estudos recentes sugerem que este gene poderá constituir um marcador da úlcera péptica, estando significativamente associado a estirpes de *H. pylori* isoladas de doentes pediátricos com úlcera gástrica e duodenal^[10].

O gene *jhp870* codifica para uma proteína (JHP870) que faz parte de um pequeno grupo de proteínas parálogas, possuindo sequências sinal típicas das OMP, denominado de família *hom*. No genoma da estirpe de referência 26695 estão presentes três genes desta família, havendo no genoma da outra estirpe de referência J99 um gene adicional, o gene *jhp870*. Todos os membros desta família apresentam os terminais C e N conservados, enquanto que o domínio central da molécula tem uma variabilidade significativa^[9].

O gene *jhp870*, específico da estirpe J99, é 90% semelhante ao gene *jhp649* também presente no genoma desta estirpe, e homólogo ao gene *HP710* presente no genoma da estirpe 26695, estando as diferenças entre eles confinadas ao interior do gene (resíduos 147 a 344). Isto sugere que o gene *jhp870* pode ter resultado de uma duplicação, relativamente recente, do gene *jhp649*^[9].

O gene *jhp870* é uma inserção no genoma da estirpe J99, sendo os genes flanqueantes do ponto de inserção nos dois genomas das estirpes de referência ortólogos entre si (*jhp869/HP935* e *jhp871/HP936*). A região intergénica entre os genes *HP935* e *HP936* no genoma da estirpe 26695 contém uma zona de 219 nucleótidos que é 96,8% semelhante ao gene *jhp870* (resíduos 496 a 569). A presença deste fragmento de DNA na estirpe 26695, nesta localização genómica, sugere que um ortólogo de *jhp870* já existiu nesta estirpe^[9].

Assim, o gene *jhp870*, bem como o seu homólogo *jhp649*, constituem o objecto de estudo deste trabalho. De modo a conhecer a diversidade destes genes, pretendeu-se estudar a sua prevalência, a variabilidade

Introduction

Helicobacter pylori (*H. pylori*) is a gram-negative, spiral shaped, microaerophilic bacteria that colonizes exclusively the human stomach and is normally found in the antrum and/or body at the surface of epithelial cells^[1]. This bacteria is etiologically associated to various gastropathologies such as non-ulcer dyspepsia or peptic ulcer and is considered a risk factor in the development of gastric carcinoma and MALT lymphoma^[2,3].

The totally sequenced genomes, from the two referenced *H. pylori* strains, strain 26695 (isolated from a patient with gastritis)^[4] and strain J99 (isolated from a patient with a duodenal ulcer)^[5], presented a large family of outer membrane proteins (OMP), composed by 32 members, including the BabA, HopZ and OipA proteins, which are implicated in the adhesion of the bacteria to the epithelial gastric cells^[6, 7, 8]. Another member of this family is codified by the *jhp870* gene. Recent studies showed a significant association between this gene and *H. pylori* strains isolated from paediatric patients with duodenal and gastric ulcers, suggesting that *jhp870* gene may constitute a peptic ulcer marker^[10].

The *jhp870* gene codifies a protein (JHP870) which is part of a small group of paralogous proteins, possessing typical OMP sequence signals, denominated as the *hom* family. In the genome of the reference strain 26695, there are three genes from this family, while in the genome of the other reference strain J99, there is an additional gene, the *jhp870*. All of the members of this family present the conserved C- and N-terminal, whereas the central domain of the molecule has a significant variability^[9].

The *jhp870* gene, specific of the J99 strain, is 90% identical to the *jhp649* gene which is also present in the genome of that strain and is homologous to the *HP710* gene present in the 26695 strain. The differences between them are confined to the interior of the gene (residuals 147 and 344). This suggests that the *jhp870* gene may have resulted from a relatively recent duplication of the *jhp649* gene^[9].

The *jhp870* gene is an insertion in the genome of strain J99, while the genes flanking the insertion point in the two genomes of the reference strains are orthologous between themselves (*jhp869/HP935* and *jhp871/HP936*). The intergenic region between *HP935* and *HP936* genes in the genome of 26695 strain contains a zone of 219 nucleotides that is 96.8% identical to the *jhp870* gene (residuals 496 and 569). The presence of this DNA fragment in the 26695 strain, at this genomic localization, suggests that a *jhp870* orthologous once existed in this strain^[9].

Therefore gene *jhp870* as well as its homologous *jhp649* constitute the object of study in this paper. Polymorphisms at the level of nucleotide and predicted amino acid sequences, in a panel of *H. pylori* clinical strains isolated from patients of different

no número de cópias e sua localização genómica, e os polimorfismos ao nível da sequência de bases e respectivas proteínas, numa amostra de estirpes clínicas de *H. pylori*, isoladas de doentes de diferentes regiões geográficas, apresentando diversas patologias gastroduodenais (carcinoma gástrico, úlcera péptica e dispepsia não ulcerosa), e recorrendo às técnicas de PCR e sequenciação enzimática.

Materiais e Métodos

Estirpes de *H. pylori*. Foram seleccionadas 65 estirpes de *H. pylori* isoladas de biopsias gástricas de doentes submetidos a endoscopia digestiva alta, das quais: 32 isoladas de doentes adultos ($55,4 \pm 12,8$ anos) e 33 isoladas de crianças ($10,1 \pm 3,3$ anos). Na população adulta: 30 estirpes de origem Portuguesa [3 apresentando carcinoma gástrico; 16 úlcera péptica (gástrica ou duodenal); e 11 dispepsia não ulcerosa]; e 2 de origem Brasileira (carcinoma gástrico). Na população pediátrica: 26 estirpes de origem Portuguesa [13 apresentando úlcera péptica (gástrica ou duodenal) e 13 dispepsia não ulcerosa]; e 7 de origem Brasileira [3 com úlcera péptica (gástrica ou duodenal) e 4 com dispepsia não ulcerosa]. Foram ainda estudadas 2 estirpes de origem Etíope, mas não existem dados de patologia e população referentes a estas estirpes. Como referência foram utilizadas duas estirpes de *H. pylori*, cujos genomas já se encontram sequenciados, a estirpe 26695 (ATCC 700392)^[4], isolada de um doente com gastrite, e a estirpe J99 (ATCC 700824)^[5], isolada de um doente com úlcera duodenal.

Condições de cultura de *H. pylori*. As estirpes de *H. pylori* conservadas a -80°C em meio TSB com 15% de glicerol, foram semeadas em meio não selectivo Columbia e selectivo H.P., ambos suplementados com 10% de sangue de cavalo, e incubadas a 37°C numa atmosfera de microaerófilia (5% de O_2 ; 10% de CO_2 ; 85% de N_2), durante 48 horas.

Extracção de DNA pelo método de Tiocianato de guanidina. A extracção de DNA foi realizada pelo método de Tiocianato de guanidina, adaptado de Pitcher *et al.*^[11].

Avaliação da distribuição dos genes *jhp870* e *jhp649* na amostra de estirpes em estudo. Para o estudo da distribuição dos genes *jhp870* e *jhp649*, na amostra de estirpes em estudo, procedeu-se a uma reacção de amplificação, pela técnica de PCR.

A reacção de amplificação foi realizada num volume final de 25 μl contendo: 1 μl de DNA, Tampão 1X com 1,5 mM MgCl_2 , 200 μM de cada dNTP (Roche Diagnostics GmbH, Mannheim, Alemanha), 0,5 μM de cada iniciador (*jhp870F1B* e *jhp870R1B*), 1 U de Taq polimerase (Roche Diagnostics GmbH), e água

Geographic regions who presented diverse gastroduodenal pathologies (gastric carcinoma, peptic ulcer and non-ulcer dyspepsia), by PCR and sequencing.

Materials and Methods

***H. pylori* strains.** Sixty five *H. pylori* strains isolated from gastric biopsies of patients submitted to a endoscopy of which: 32 isolated from adult patients (55.4 ± 12.8 years) and 33 isolated from children (10.1 ± 3.3 years) were studied. In the adult population: 30 strains of Portuguese origin [3 presented gastric carcinoma; 16 peptic ulcer (gastric or duodenal); and 11 non-ulcer dyspepsia]; and 2 of Brazilian origin (gastric carcinoma). In the paediatric population: 26 strains of Portuguese origin [13 presented peptic ulcer (gastric or duodenal) and 13 non-ulcer dyspepsia]; and 7 of Brazilian origin [3 with peptic ulcer (gastric or duodenal) and 4 with non-ulcer dyspepsia]. Two strains of Ethiopian origin were studied, but there is no data regarding the pathology and population of these strains. The two *H. pylori* strains with completely sequenced genomes, 26695 strain (ATCC 700392)^[4] isolated from a patient with gastritis, and J99 strain (ATCC 700824)^[5], isolated from a patient presenting duodenal ulcer, were used as reference strains.

***H. pylori* culture conditions.** The *H. pylori* strains conserved at -80°C in TSB with 15% of glycerol, were grown in a non-selective Columbia and an H.P. selective media, both supplemented with 10% horse blood, and incubated at 37°C in a microaerophilic atmosphere (5% of O_2 ; 10% of CO_2 ; 85% of N_2), during 48 hours.

DNA extraction by the guanidinium thiocyanate method. The DNA extraction was carried out by the guanidinium thiocyanate method, adapted from Pitcher *et al.*^[11].

Evaluation of the *jhp870* and *jhp649* genes distribution in the sample of strains studied. For the study of the *jhp870* and *jhp649* genes distribution, in the sample of strains studied, we proceeded with an amplification reaction by using the PCR technique.

The amplification reaction took place in a final volume of 25 μl containing: 1 μl of DNA, 1X PCR buffer with 1,5 mM MgCl_2 , 200 μM of each dNTP (Roche Diagnostics GmbH, Mannheim, Germany), 0,5 μM of each primer (*jhp870F1B* e *jhp870R1B*), 1 U of Taq polymerase (Roche Diagnostics GmbH), and sterilized distilled water. In each assay a negative

destilada estéril. Em cada ensaio usou-se um controle negativo constituído por água e como controle positivo o DNA da estirpe J99, portadora de ambos os genes.

A reacção de amplificação consistiu num ciclo inicial de desnaturação da cadeia de DNA durante 30 segundos a 94°C, seguido de 35 ciclos constituídos por uma desnaturação de 30 segundos a 94°C, uma etapa de hibridação dos iniciadores a 60°C durante 30 segundos e uma extensão de 30 segundos a 72°C. A etapa de extensão final foi realizada a 72°C durante 7 minutos. Foi utilizado um Termociclador GeneAmp® PCR System 9700 (PE Applied BioSystems, Foster City, CA, EUA).

Os iniciadores utilizados neste estudo, descritos na tabela 1, foram desenhados de modo a amplificar uma região conservada no interior dos dois genes, flaqueando três inserções, de 27, 2 e 4 pb presentes no gene *jhp870* e ausentes no gene homólogo *jhp649* (Figura 1).

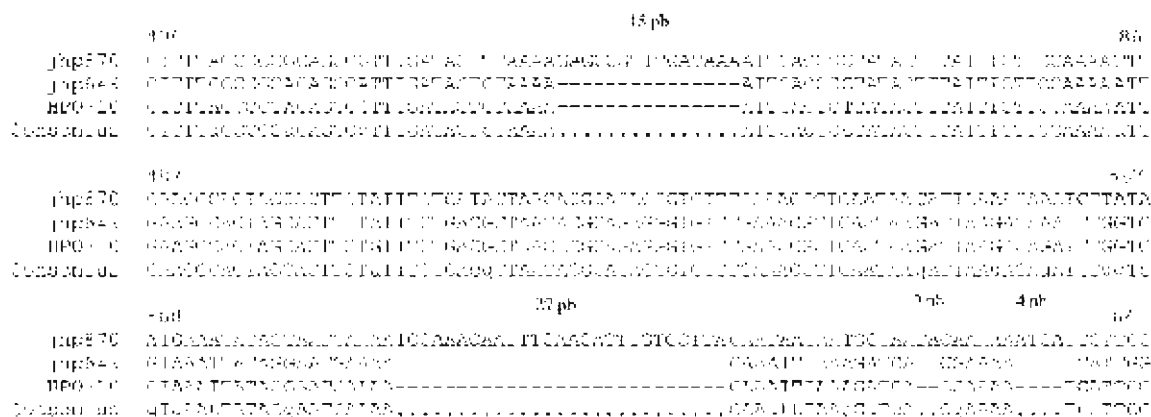


Figura 1: Alinhamento representativo das sequências de nucleótidos dos genes *jhp870* e *jhp649*, das estirpes de referência, na região entre 406 e 648 nucleótidos.

Figure 1: Alignment representing the nucleotide sequences for the *jhp870* e *jhp649* genes, of the reference strains, in the region between 406 and 648 nucleotides.

Os produtos amplificados foram detectados por electroforese em gel de agarose de 2,5%, corado com brometo de etídeo (0,5 µg/ml), em tampão de corrida TBE 0,5X (0,045 M Ácido Bórico; 0,01 M EDTA; pH 8). Os produtos esperados são de 161 pb para o gene *jhp870* e 128 pb para o *jhp649*. Foi utilizado o marcador de peso molecular IX (Fragmentos de DNA ΦX 174 RF digeridos com Hae III de 72 pb a 1353 pb) (Invitrogen, Carlsbad, CA, EUA).

Determinação da sequência de bases dos genes *jhp870* e *jhp649* em estirpes clínicas de *H. pylori*. Nesta parte do trabalho pretendeu-se, em primeiro lugar, estudar a variação no número de cópias dos genes *jhp870* e *jhp649* bem como a sua localização no genoma das estirpes clínicas de *H. pylori*. Em segundo

control constituted by water, and a positive control, the DNA of J99 strain, which carried both genes, were used.

The amplification reaction consisted of an initial denaturation cycle of the DNA during 30 s at 94°C, followed by 35 cycles constituted by a 30s denaturation at 94°C, a annealing of the primers at 60°C during 30 s and an extension at 72°C for 30 s. The final extension cycle occurred at 72°C for 7 min. A Thermocycler GeneAmp® PCR System 9700 (PE Applied BioSystems, Foster City, CA, USA) was used.

The primers used in this study, described in table 1, were designed in order to amplify a conserved region within the two genes, flanking three insertions of 27, 2 and 4 bp present in the *jhp870* gene and absent in its homologous *jhp649* (Figure 1).

The amplified products were detected by electrophoresis in a 2.5% agarose gel, stained with ethidium bromide (0,5 µg/ml) in a 0,5X TBE buffer (0,045 M Boric Acid; 0,01 M EDTA; pH 8). The expected products were 161 bp for the *jhp870* gene and 128 bp for the *jhp649*. The molecular weight marker IX (Fragments of DNA ΦX 174 RF digest of Hae III from 72 bp to 1353 bp) (Invitrogen, Carlsbad, CA, USA) was used.

Determination of the *jhp870* and *jhp649* genes nucleotide sequences in *H. pylori* clinical strains. In this part of the paper, the first aim was study the variation in the number of copies of the genes *jhp870* and *jhp649* as well as its localization in the genome of the *H. pylori* clinical strains. The *jhp870* or *jhp649*

lugar, foi determinada a sequência de bases dos genes *jhp870* ou *jhp649* nestas mesmas estirpes. Para tal foram utilizadas técnicas de PCR e sequenciação enzimática. A reacção de amplificação foi realizada num volume final de 25 µl contendo: 1 µl de DNA, Tampão 1X, 2,8 mM de MgCl₂, 200 µM de cada dNTP (Roche Diagnostics GmbH), 0,5 µM de cada iniciador, 1 U de Bio-X-Act Short polimerase (Bioline, Londres, Reino Unido), e água destilada estéril. Em cada ensaio usou-se um controlo negativo constituído por água destilada, e como controlos positivos o DNA das estirpes de referência J99 e 26695.

A reacção de amplificação consistiu num ciclo inicial de 5 minutos a 95°C e 1 minuto a 56°C, seguido de 35 ciclos constituídos por uma extensão de 3 minutos a 68°C, uma desnaturação a 95°C durante 30 segundos e uma hibridação dos iniciadores a 56°C durante 30 segundos, com um ciclo de extensão final de 10 minutos a 68°C, realizada num Termociclador GeneAmp® PCR System 9700 (PE Applied Biosystems).

De modo a obter um fragmento contendo cada gene completo, foram desenhados iniciadores nas regiões genómicas conservadas, adjacentes aos genes *jhp870* e *jhp649*, tendo como referência a sua localização genómica nas duas estirpes de referência (Tabela 1).

Genes nucleotide sequences in these strains, was subsequently determined, using PCR technique and enzymatic sequencing. Amplification reaction was carried out in a final volume of 25 µl containing: 1 µl of DNA, 1X PCR buffer, 2,8 mM of MgCl₂, 200 µM of each dNTP (Roche Diagnostics GmbH), 0,5 µM of each primer, 1 U of Bio-X-Act Short polymerase (Bioline, London, United Kingdom) and sterilized distilled water. A negative control, consisting of distilled water, and two positive controls, consisting of the reference strains J99 and 26695 DNA, were used.

The amplification reaction started with an initial cycle of 5 min at 95°C and 1 min at 56°C, followed by 35 cycles constituted by an extension of 3 min at 68°C, a denaturation at 95°C during 30 s and annealing of the primers at 56°C during 30 s, with a final extension cycle of 10 min at 68°C, and was carried out in a Thermocycler GeneAmp® PCR System 9700 (PE Applied Biosystems).

In order to obtain a fragment containing each complete gene, primers were designed in the conserved genomic regions adjacent to the *jhp870* e *jhp649* genes, maintaining, as a reference, the genomic localization in the two reference strains (Table 1).

Tabela 1: Iniciadores utilizados nas reacções de amplificação.
Table 1: PCR primers used in this study.

Região Genómica <i>Genomic Region</i>	Iniciador <i>Primer</i>	Sequência (5' 3') <i>Sequence (5' 3')</i>	Tm (°C) <i>Tm (°C)</i>
Região Interna <i>Internal to genes</i>	jhp870B F1	AAAGGGTGTGTTGAAACGCTCAATA	66,2
	jhp870B R1	GGTGAATTCTTCTGCGGTTTG	65,5
Região Externa <i>External to genes</i>	jhp869 F1	AAGAGGATTGCGTGGTGGAGTTG	62,4
	jhp871 R1	GGGTTGCCTTTGGGCTTGGA	61,4
	jhp648 F1	TAATTTGCGCAAAAACATC	51,2
	jhp650 R1	ATTCCAGCGCCTAATGGA	56,7

Os produtos de amplificação obtidos foram purificados com o kit Jetquick (Genomed GmbH, Löhne, Alemanha), segundo instruções do fabricante. Estes produtos purificados foram sujeitos a sequenciação enzimática utilizando o Big Dye® Terminator (PE Applied Biosystems Chemistry, Foster City, CA, EUA) e o Sequenciador automático Genetic Analyser Abi-Prism 3130 xl (PE Applied Biosystems).

The amplification products obtained were purified with a Jetquick kit (Genomed GmbH, Löhne, Germany) in accordance with the manufacturer's instructions. These purified products were subjected to enzymatic sequencing using the Big Dye® Terminator (PE Applied Biosystems Chemistry, Foster City, CA, USA) and the Automatic Sequencer Genetic Analyser Abi-Prism 3130 xl (PE Applied Biosystems).

A reacção de sequenciação foi realizada num volume final de 10 µl contendo: 2 µl de Big Dye® V1.1 (PE Applied Biosystems Chemistry), 0,5 µM de iniciador, 0,5 a 7,5 µl de DNA purificado e água destilada estéril. O programa de sequenciação consistiu num ciclo inicial de desnaturação de 30 segundos a 96°C, seguido de 25 ciclos constituídos por uma desnaturação de 10 segundos a 96°C, uma hibridação de 5 segundos a 50°C e uma extensão de 4 minutos a 60°C. Estas reacções foram realizadas num Termociclador GeneAmp® PCR System 9700 (PE Applied BioSystems).

As características dos iniciadores de sequenciação encontram-se descritas na tabela 2. A sequenciação foi realizada pela técnica de gene *walking* de modo a sequenciar o gene na sua totalidade.

As sequências de nucleótidos obtidas foram visualizadas no programa *Chromas* (Version 1.45) e foram analisadas recorrendo ao programa *MultAlin* (<http://prodes.toulouse.inra.fr/multalin/multalin.html>). Em seguida foram traduzidas em sequências de aminoácidos correspondentes recorrendo ao programa *EMBOSS transeq* (www.ebi.ac.uk/emboss/transeq/).

The sequencing reaction occurred in a final volume of 10 µl containing: 2 µl of Big Dye® V1.1 (PE Applied Biosystems Chemistry), 0,5 µM of primer, 0,5 to 7,5 µl of purified DNA and distilled sterile water. The sequencing program consisted in an initial denaturation at 96°C for 30 s, followed by 25 cycles constituted by a 10 s denaturation at 96°C, a 5 s hybridization at 50°C and an extension of 4 min at 60°C. These reactions were carried out in a Thermocycler GeneAmp® PCR System 9700 (PE Applied BioSystems).

The characteristics of the sequencing primers are described in table 2. The sequencing was carried out with the gene walking technique, in order to sequence the gene in its totality.

The nucleotide sequences obtained were visualized in the *Chromas* (Version 1.45) program and were analysed by resorting to the *MultAlin* program (<http://prodes.toulouse.inra.fr/multalin/multalin.html>). Then, the corresponding amino acid sequences were translated using the *EMBOSS transeq* program (www.ebi.ac.uk/emboss/transeq/).

Tabela 2: Iniciadores utilizados na reacção de amplificação da sequenciação enzimática.
Table 2: Sequencing primers used in this study.

Região Genómica <i>Genomic Region</i>	Iniciador <i>Primer</i>	Sequência (5' 3') <i>Sequence (5' 3')</i>	Tm (°C) <i>Tm (°C)</i>
jhp869/jhp871	jhp869 F1	AAGAGGATTGCGTGGTGGAGTTG	62,4
	jhp870 F2	TTTACAGACGCTCAAGGCAA	51
	jhp870 F4	CATCTTGCATGAATTCAGC	52,4
	jhp870 R3	GCGTTGGTGGATTTTGGT	53,7
jhp648/jhp650	jhp648 F1	TAATTTTCGCGCAAAAACATC	51,2
	jhp870 F2	TTTACAGACGCTCAAGGCAA	51
	jhp870 F4	CATCTTGCATGAATTCAGC	52,4
	jhp650 R1	ATTCCAGCGCCTAATGGA	56,7

Análise bioinformática das sequências obtidas. Gráficos de similaridade, utilizando alinhamentos gerados pelo programa *BioEdit Sequence Alignment Editor* (Version 7.0.1), foram obtidos com recurso à aplicação do programa *SimPlot* (Version 3.5.1).

Bioinformatic Sequences analysis. Similarity plots, using the alignments generated by *BioEdit Sequence Alignment Editor* (Version 7.0.1) program, were obtained with the application of the *SimPlot* (Version 3.5.1) program.

Resultados

O estudo da distribuição e da diversidade genómica dos genes *jhp870* e *jhp649* foi realizado, por PCR e sequenciação, em 67 estirpes clínicas de *H. pylori*, seleccionadas ao acaso de um painel mais alargado de estirpes, obtidas em três diferentes países, Portugal, Brasil e Etiópia.

Avaliação da distribuição dos genes *jhp870* e *jhp649* no genoma de estirpes clínicas de *H. pylori*.

Para o estudo da distribuição dos genes *jhp870* e *jhp649*, na amostra de estirpes em estudo, procedeu-se a uma reacção de amplificação com os iniciadores internos, pela técnica de PCR, de modo a amplificar uma região conservada nos dois genes, flaqueando três inserções, de 27, 2 e 4 pb presentes no gene *jhp870* e ausentes no gene homólogo *jhp649*, distinguindo assim os dois genes pelo tamanho do fragmento obtido.

A partir dos resultados obtidos, analisou-se a presença destes genes nas estirpes clínicas, e para cada patologia foi avaliada a distribuição de cada gene.

Nas estirpes isoladas de doentes com carcinoma gástrico, obtiveram-se 2 (40%) estirpes só com o gene *jhp870*, 1 (20%) com o gene *jhp649*, e outras 2 (40%) com os dois genes. Nas estirpes de doentes com úlcera, foram obtidas 15 (46,9%) estirpes com o gene *jhp870*, 9 (28,1%) com o gene *jhp649*, e 8 (25%) com ambos os genes. E nas estirpes isoladas de dispepsia não ulcerosa obtiveram-se 14 (50%) estirpes com o gene *jhp870*, 9 (32,14%) com o gene *jhp649*, e 5 (17,86%) estirpes com ambos os genes. Nas duas estirpes de origem Etíope, observou-se apenas o gene *jhp870*.

Tendo como base a localização dos genes *jhp870* e *jhp649* na estirpe J99, foram denominadas duas posições onde estes genes se podem encontrar: a região *jhp869/jhp871* (posição 1) e a região *jhp648/jhp651* (posição 2).

Na estirpe de referência J99 ambos os genes estão presentes, encontrando-se o *jhp870* na posição 1 e o *jhp649* na posição 2, situação identificada como *jhp870/jhp649*. Por outro lado, na estirpe 26695 apenas existe o gene ortólogo do gene *jhp649*, denominado *HP0710*, e que se encontra na posição 2, enquanto que na posição 1 existe uma região intergénica, situação identificada como intergénica/*jhp649*, não existindo nesta estirpe o ortólogo do gene *jhp870*, tal como já referido anteriormente.

Pela amplificação destas regiões utilizando os iniciadores externos, verificou-se que nas estirpes clínicas estes genes ocupam as mesmas posições genómicas das estirpes de referência, no entanto as suas posições podem estar trocadas e podem existir em mais do que uma cópia. Assim, pode-se encontrar no mesmo genoma uma destas combinações: 2 cópias do gene *jhp870* ou do gene *jhp649*, ocupando as posições 1 e 2 (*jhp870/jhp870*; *jhp649/jhp649*); 1 cópia do gene

Results

The study of the genomic distribution and diversity of the *jhp870* and *jhp649* genes was carried out through PCR sequencing in 67 *H. pylori* clinical strains, which were randomly selected from a wider panel of strains obtained in three different countries: Portugal, Brazil and Ethiopia.

Evaluation of the *jhp870* and *jhp649* genes distribution in *H. pylori* clinical strains genome.

To study the *jhp870* and *jhp649* genes distribution in the panel of strains considered, an amplification reaction was performed using the PCR technique, with the internal primers. This was done in order to amplify a conserved region of the two genes, flanking three insertions of 27, 2 and 4 bp present in the *jhp870* gene and absent in its homologous *jhp649*. Thus the two genes were distinguished by the size of the PCR fragment obtained.

Based on the results of this PCR, the distribution of these genes was evaluated in the clinical strains, as well as within each pathology.

Among strains from patients with gastric carcinoma, 2 (40%) had only the *jhp870* gene, 1 (20%) the *jhp649* gene and another 2 (40%) had both genes. In strains isolated from patients with peptic ulcer, 15 strains (46.9%) presented the *jhp870* gene, 9 (28.1%) the *jhp649* gene and 8 (25%) had both genes. Finally, in non-ulcer dyspepsia strains, 14 strains (50%) harbour the *jhp870* gene, 9 (32.14%) the *jhp649* gene, and 5 strains (17.86%) had both genes. In the two strains of Ethiopian origin, we only observed the *jhp870* gene.

Based on the localization of the *jhp870* and *jhp649* genes on the J99 strain, two positions where these genes could be found were identified: the *jhp869/jhp871* region (position 1) and the *jhp648/jhp651* region (position 2).

In the J99 reference strain, both genes are present, with the *jhp870* gene in position 1 and the *jhp649* in position 2, which was designated *jhp870/jhp649*. On the other hand, in the 26695 strain there is only an orthologous gene to *jhp649*, denominated as *HP0710*, which is found in position 2, while in position 1 there is an intergenic region, which was designated intergenic/*jhp649*. As previously mentioned, no homologous gene of *jhp870* exists in this strain.

Using the external primers targeting the two genomic positions, we verified that in the clinical strains these genes occupy the same positions as in the reference strains. However, their positions may be switched and they may exist in more than one copy. Thus, in the same genome one of these combinations can be found: 2 copies of the *jhp870* gene or the *jhp649* gene, occupying positions 1 and 2 (*jhp870/jhp870*; *jhp649/jhp649*); 1 copy of the *jhp870* gene or the *jhp649* always occupying position 2, while in position 1 there is an intergenic region (intergenic/*jhp870*;

jhp870 ou do *jhp649* ocupando sempre a posição 2, enquanto que na posição 1 existe uma região intergênica (intergênica/*jhp870*; intergênica/*jhp649*); ou 1 cópia de cada gene, ocupando indiferentemente uma das posições (*jhp870/jhp649*; *jhp649/jhp870*). Foi então avaliada a distribuição genômica por posição e número de cópias, para cada patologia.

Em estirpes isoladas de doentes apresentando carcinoma gástrico, encontrou-se apenas uma estirpe para cada combinação apresentada: região intergênica/*jhp870*, *jhp870/jhp870*, *jhp870/jhp649*, *jhp649/jhp870* e *jhp649/jhp649*.

Nas estirpes com origem ulcerosa encontraram-se 8 (25%) estirpes apresentando a combinação intergênica/*jhp870*, 7 (21,8%) *jhp870/jhp870*, 5 (15,6%) intergênica/*jhp649*, e 4 (12,5%) estirpes para cada uma destas combinações *jhp870/jhp649*, *jhp649/jhp870* e *jhp649/jhp649*. Por último, nas estirpes isoladas de doentes com dispepsia não ulcerosa, obtiveram-se 8 (28,6%) com a combinação intergênica/*jhp870*, 7 (25%) intergênica/*jhp649*, 6 (21,4%) *jhp870/jhp870*, 5 (17,8%) *jhp649/jhp870* e por fim 2 (7,14%) estirpes com *jhp649/jhp649*.

Nas estirpes da Etiópia, para as quais não se conhece a proviniência clínica, foi encontrada a combinação intergênica/*jhp870* para uma estirpe, e *jhp870/jhp870* para outra.

Análise das sequências de nucleótidos dos genes *jhp870* e *jhp649*. Para todas as estirpes estudadas, foi obtida uma sequência completa codificante do gene ou dos genes presentes, compatível com a sua funcionalidade, e com tamanhos bastante variáveis. Os fragmentos correspondentes ao gene *jhp870* variaram entre 1974 e 2007 pb, a maioria apresentando os tamanhos 1986 e 2007 pb. Para o gene *jhp649*, os fragmentos apresentaram um tamanho inferior, com maior variabilidade, encontrando-se tamanhos entre 1941 e 1986 pb, sendo o tamanho mais frequente de 1962 pb. No total das estirpes clínicas, foram obtidas 64 sequências de nucleótidos e de aminoácidos para o gene *jhp870*, e 41 para o gene *jhp649*.

Avaliação da homologia entre as sequências dos genes *jhp870* e *jhp649*. Foi avaliado o grau de homologia entre as sequências de nucleótidos dos genes *jhp870* e *jhp649*, de todas as estirpes, utilizando alinhamentos múltiplos gerados no programa *BioEdit Sequence Alignment Editor*, pelo método *ClustalW* e utilizando *Bootstrap NJ Tree* com 1000 réplicas.

A partir desses alinhamentos foram gerados gráficos de similaridade pelo método *Jukes-Cantor (1-parameter)*^[12], através do programa *Simplot*.

Os gráficos de similaridade sugerem a existência de três regiões distintas em ambos os genes, que foram denominadas por segmentos 1, 2 e 3, e que correspondem, respectivamente, às regiões 5', média e 3' dos genes (Figura 2).

intergênica/*jhp649*); or 1 copy of each gene, occupying indifferently one of the positions (*jhp870/jhp649*; *jhp649/jhp870*). The genomic distribution was then evaluated according to the position and number of copies for each pathology.

In strains isolated from patients with gastric carcinoma, only one strain for each combination was observed: intergênica/*jhp870*, *jhp870/jhp870*, *jhp870/jhp649*, *jhp649/jhp870* and *jhp649/jhp649*.

In strains of ulcerous origin, we found 8 strains (25%) that presented the combination intergênica/*jhp870*, 7 (21.8%) *jhp870/jhp870*, 5 (15.6%) intergênica/*jhp649* and 4 strains (12.5%) for each of these combinations: *jhp870/jhp649*, *jhp649/jhp870* and *jhp649/jhp649*. Finally, in strains isolated from patients with non-ulcer dyspepsia, 8 (28.6%) had the combination intergênica/*jhp870*, 7 (25%) intergênica/*jhp649*, 6 (21.4%) *jhp870/jhp870*, 5 (17.8%) *jhp649/jhp870* and 2 strains (7.14%) *jhp649/jhp649*.

In the Ethiopian strains, whose clinical source is unknown, the combination intergênica/*jhp870* was observed for one strain, and *jhp870/jhp870* for the other.

Analysis of the *jhp870* and *jhp649* genes nucleotide sequences. For all of the strains studied, a complete codifying sequence of the gene or genes present was observed, with significant size variations.

The corresponding fragments of the *jhp870* gene varied between 1974 and 2007 bp, with the majority presenting a size between 1986 and 2007 bp. For the *jhp649* gene, the fragments presented a size varying between 1941 and 1986 bp. Overall, 64 nucleotide and amino acid sequences for the *jhp870* gene and 41 for the *jhp649* gene were obtained.

Homology evaluation of the *jhp870* and *jhp649* genes sequences. The degree of homology between the nucleotide sequences of the *jhp870* and *jhp649* genes were evaluated in all of the strains by using multiple alignments which were generated in the *BioEdit Sequence Alignment Editor* program with the *ClustalW* method and using the *Bootstrap NJ Tree* with 1000 replications.

From these alignments similarity plots were generated with the *Jukes-Cantor (1-parameter)* correction^[12], using the *Simplot* program.

The similarity plots suggest the existence of three distinct regions in both genes which were denominated as segments 1, 2 and 3, and that correspond, respectively, to 5', middle and 3' regions of the genes (Figure 2).

By observing the zone corresponding to segment 1, represented in the figure 2 plot, we verified that there are differences among the nucleotide sequences of the

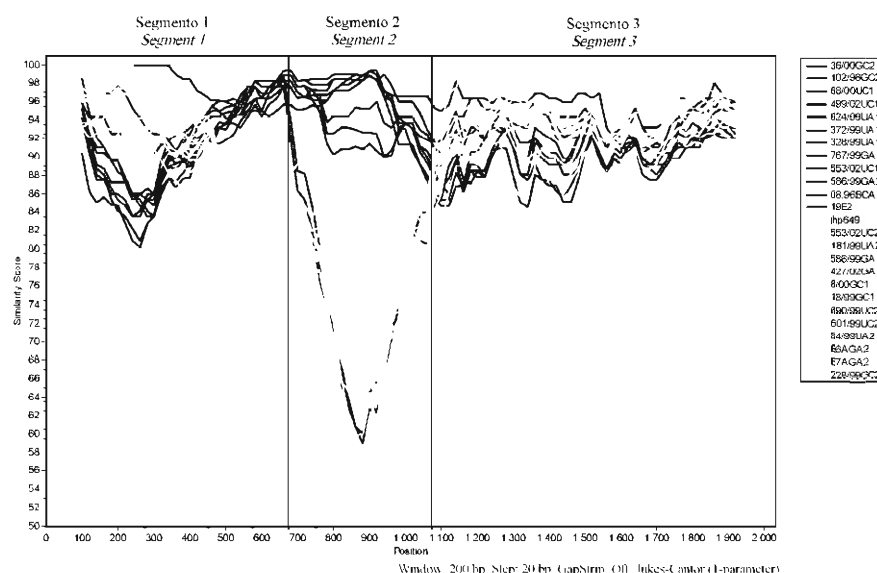


Figura 2: Representação gráfica da similaridade entre os genes *jhp870* (linha preta) e *jhp649* (linha cinzenta), em várias estirpes de *H. pylori* em estudo. Este gráfico foi gerado usando 16 estirpes representativas de cada gene, com o modelo *Jukes-Cantor* (*1-parameter*), uma janela de 200 pb, um passo de 20 pb, sem *Gap Strip* e como referência a sequência do gene *jhp870*.

Figure 2: Similarity plot representation between the *jhp870* genes (black line) and *jhp649* (grey line), in various *H. pylori* strains under study. This plot was generated by using 16 strains representative of each gene, with the *Jukes-Cantor* correction (*1-parameter*), a 200 bp window, a 20 bp step, without *Gap Strip* and as reference the *jhp870* gene sequence.

Pela observação da zona correspondente ao segmento 1, representado no gráfico da figura 2, verifica-se que existem diferenças entre as sequências nucleotídicas dos dois genes, sendo a zona de maior diferença entre 500 e 700 pb, onde a similaridade entre os dois genes é apenas cerca de 50%. Esta diferença traduz-se, a nível da sequência de bases, em quatro inserções de 15, 27, 2 e 4 nucleótidos, presentes no gene *jhp870* e ausentes no seu homólogo *jhp649* (Figura 1). Em relação ao segmento 3, não há diferenças muito demarcadas, demonstrando que a região 3' é bastante conservada nos dois genes e entre eles (Figura 2).

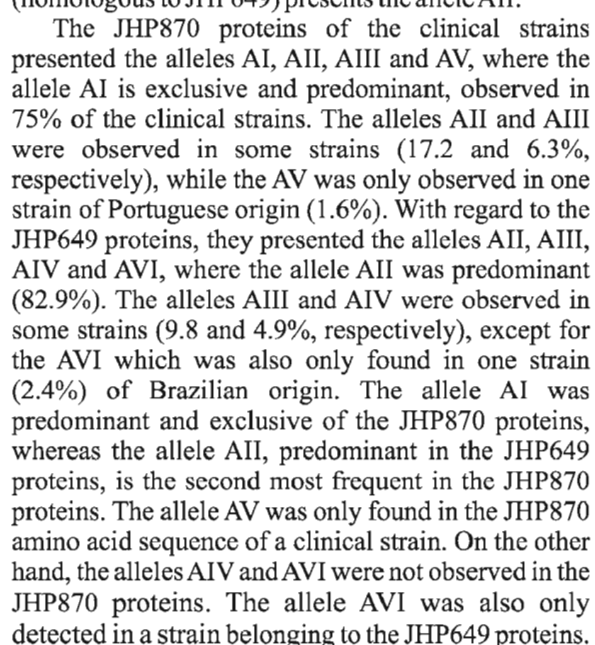
Ao contrário das regiões anteriores, a região genómica que corresponde ao segmento 2 de ambos os genes é extremamente polimórfica, tendo sido detectadas diferenças notórias entre os dois genes e em sequências do mesmo gene em diferentes estirpes (Figura 2).

Devido ao elevado grau de polimorfismo observado na região média de cada gene, procedeu-se a uma análise mais detalhada desta região, incluindo quer a análise das sequências de bases, quer das respectivas sequências proteicas. Assim, a partir de um alinhamento gerado com as proteínas JHP870 e JHP649 (Figura 3), observou-se que nas estirpes clínicas e de referência estudadas, existem seis variantes alélicas distintas. Estes alelos foram denominados por AI, AII, AIII, AIV, AV e AVI, e identificados em todas as estirpes presentes neste estudo, embora apresentando uma distribuição diferente consoante o gene.

two genes, the region of highest difference being between 500 and 700 bp, where the similarity between the two genes is only 50%. This difference corresponds to the four insertions of 15, 27, 2 and 4 nucleotides present in the *jhp870* gene and absent in *jhp649* (Figure 1). Regarding segment 3, there are not many relevant differences, demonstrating that 3' region is significantly conserved within and between the two genes (Figure 2).

On the other hand, the genomic region that corresponds to segment 2 for both genes is extremely polymorphic. There were notorious differences detected between the two genes and in sequences of the same gene in different strains (Figure 2).

Due to the high level of polymorphism observed in the middle region of each gene, a more detailed analysis of this region was performed, including the analysis of nucleotide sequences and predicted amino acid sequences. Thus, from the alignment generated from the JHP870 and JHP649 proteins (Figure 3), six allelic variants were distinguished. These alleles were denominated as AI, AII, AIII, AIV, AV and AVI, and identified in all of the strains present in this study, although they presented different distributions in each gene.



proteínas JHP649. Por fim, o alelo AIII foi detectado em proteínas JHP870 e JHP649, sendo embora pouco prevalente (< 10%).

O facto de o alelo AI ser exclusivo e predominante das proteínas JHP870, permite fazer a distinção entre os genes *jhp870* e *jhp649*.

Em geral, os resultados obtidos tendo em conta o tipo de população (adultos e crianças), as diferentes patologias e a origem geográfica, mostram que a distribuição dos alelos é homogênea, não existindo diferenças importantes entre os diferentes grupos estudados.

De forma a analisar as semelhanças e diferenças entre os vários alelos, foram gerados gráficos de similaridade a partir de sequências nucleotídicas representativas de cada alelo (dados não mostrados). Numa primeira análise, verifica-se que os alelos que mais divergem entre si são os alelos AI e AII, e que curiosamente são os predominantes nas proteínas JHP870 e JHP649, respectivamente.

Globalmente, e tendo em conta os alelos predominantes, os alelos AV e AVI apresentam uma maior similaridade entre si e são os mais próximos do alelo AI (AI e AV são exclusivos das proteínas JHP870 e o AVI exclusivo de JHP649). O mesmo se observa para os alelos AIII e AIV, que também apresentam uma maior similaridade com o alelo AII (AIV é exclusivo das proteínas JHP649).

Estas semelhanças também podem ser observadas nas sequências de aminoácidos mostradas no alinhamento da figura 3.

A presença dos alelos consoante a posição genómica não é propriamente aleatória, verificando-se que para algumas distribuições genómicas, existe uma tendência para a presença do mesmo alelo ou da mesma combinação de alelos. Nas estirpes que têm a combinação *jhp870/jhp870*, ambos os genes codificam para duas proteínas que apresentam sempre o mesmo alelo, à excepção de uma estirpe, e que pode ser AI, AII ou AIII. As que têm presente apenas uma cópia do gene *jhp870* (intergénica/*jhp870*), este codifica para proteínas que apresentam na grande maioria o alelo AI e, algumas, o alelo AII. O alelo AII só está presente nas proteínas JHP870 nas duas situações referidas anteriormente, e nunca aparece nas combinações *jhp870/jhp649* ou *jhp649/jhp870*. As proteínas JHP649 resultantes da combinação intergénica/*jhp649* apresentam sempre o alelo AII, existindo um único caso excepcional, que apresenta o alelo AIII. As proteínas resultantes da combinação *jhp649/jhp649* também apresentam na maioria dos casos o alelo AII, existindo apenas dois casos em que uma das cópias apresenta outro alelo diferente (AIII ou AVI).

Nas estirpes que possuem os dois genes, independentemente da posição, os alelos são sempre diferentes, em que as proteínas JHP870 podem apresentar os alelos AI ou AV, e as proteínas JHP649 variam entre os alelos AII, AIII ou AIV.

Finally, the allele AIII was detected in the JHP870 and JHP649 proteins, although with a low prevalence (< 10%).

Being the allele AI exclusive and predominant in JHP870 proteins, this fact can be used to make a distinction between the *jhp870* and *jhp649* genes.

The results obtained, in general, show that the distribution of alleles is homogeneous, and there are no important differences among the different groups studied (type of population, adults or children, kind of pathology and the geographic region).

In order to analyze the similarities and differences among the various alleles, we generated similarity plots, based on the nucleotide sequences that represented each allele (data not shown). In a first analysis, we verified that the alleles which diverges the most between themselves are the alleles AI and AII, and curiously these are predominant in the JHP870 and JHP649 proteins, respectively.

Globally, and taking into account the predominant alleles, the alleles AV and AVI present the higher similarity between themselves and are closer to the allele AI (AI and AV are exclusive to the JHP870 proteins and the AVI is exclusive to JHP649). The same is observed for the alleles AII and AIV that also present a higher similarity with the allele AII (AVI is exclusive to the JHP649 proteins).

These similarities can also be observed in the amino acid sequences shown in the alignment of figure 3.

The presence of the alleles according to the genomic position is not exactly random, and we verified in some genomic distributions that there is a tendency for the presence of the same allele or the same combination of alleles. In the strains that have the *jhp870/jhp870* combination, both genes codify for two proteins that always present the same allele, with the exception of one strain, and it can be AI, AII or AIII. Those that have only one *jhp870* gene copy (intergenic/*jhp870*) codify for proteins which present in most cases the allele AI, and some the allele AII. The allele AII is only present in the JHP870 proteins in the two previous referred situations and never appears in the *jhp870/jhp649* or *jhp649/jhp870* combinations. The JHP649 proteins resulting from the combination intergenic/*jhp649* always present the allele AII. There is only one exceptional case that presents the allele AIII. The proteins resulting from the *jhp649/jhp649* combination also present in the majority of cases the allele AII. There are only two cases where one of the copies presents a different allele (AIII or AVI).

In strains that have the two genes the alleles are always different, independently of the position, where the JHP870 proteins present the allelic variants AI or AV, and the JHP649 proteins vary among the alleles AII, AIII or AIV.

Discussão

Este trabalho teve como objectivo o estudo genotípico de um novo potencial marcador de virulência em *H. pylori*, o gene *jhp870*, e do seu gene homólogo *jhp649*, utilizando-se uma amostra de 67 estirpes clínicas de *H. pylori*, isoladas de doentes de diferentes regiões geográficas e apresentando diversas patologias gastroduodenais (carcinoma gástrico, úlcera péptica e dispepsia não ulcerosa).

Um estudo recente da distribuição destes genes num painel alargado de estirpes clínicas de *H. pylori* isoladas de doentes pediátricos, sugerem que o gene *jhp870* poderá constituir um novo marcador da doença ulcerosa péptica, pois foi observado em 78,3% de doentes com úlcera e 35,3% com dispepsia não ulcerosa, enquanto que o gene *jhp649* parece estar mais relacionado com a dispepsia não ulcerosa, tendo sido observado em 41,7% de doentes com úlcera e 74,5% com dispepsia não ulcerosa^[10].

No presente trabalho, foi analisada a prevalência, número de cópias, distribuição e diversidade genómica destes genes em estirpes clínicas de *H. pylori*.

Em relação à associação entre a prevalência dos genes e a patologia dos doentes, não serão feitas considerações, dada a reduzida dimensão da amostra e consequentemente o número reduzido de doentes para cada patologia, e porque esta associação já foi descrita em detalhe para a população em estudo^[10].

A diversidade nas sequências dos genes *jhp870* e *jhp649* foi avaliada na amostra de estirpes clínicas de *H. pylori* através de alinhamentos e representações gráficas de similaridade, com base nas sequências de bases e nas respectivas sequências de aminoácidos. No geral, e para ambos os genes, *jhp870* e *jhp649*, observou-se uma elevada conservação das sequências de bases em estirpes clínicas, expressa nos gráficos de similaridade.

No entanto, no gene *jhp870* foi identificada uma região, correspondente à região média e designada por segmento 2, que apresenta uma grande diversidade alélica onde se observaram quatro variantes distintas, alelos AI, AII, AIII e AV. O alelo predominante e exclusivo das proteínas JHP870 é o alelo AI, estando presente em 75% das estirpes.

Também para o gene *jhp649* se observou grande diversidade alélica na sua região média, onde estão presentes variantes comuns (AII e AIII) e distintas (AIV e AVI), das encontradas no gene *jhp870*. O alelo predominante nas proteínas JHP649 é o alelo AII, estando presente em 82,9% das estirpes, e é o segundo alelo mais frequente nas proteínas JHP870 (17,2%).

A conclusão mais importante deste trabalho é a definição das variantes alélicas, e a observação da sua constância em todas as estirpes, independentemente da sua origem geográfica, patologia, tipo de população, o que sugere a antiguidade destes alelos, e evidencia a sua conservação ao longo da evolução. Mais, ao existir esta conservação numa região polimórfica, implica

Discussion

The aim of this research was the genotypic study of a *H. pylori* potential new virulence marker, the *jhp870* gene, and its homologous the *jhp649*, using a sample of 67 *H. pylori* clinical strains, isolated from patients from different geographical regions and presenting diverse gastroduodenal pathologies (gastric carcinoma, peptic ulcer and non-ulcer dyspepsia).

A recent study of the distribution of these genes in a wide sample of *H. pylori* clinical strains, isolated from paediatric patients, suggested that the *jhp870* gene may constitute a new marker of the peptic ulcer disease, as it was observed in 78.3% of the patients with an ulcer and 35.3% of those with non-ulcer dyspepsia, while the *jhp649* gene seems to be more related with the non-ulcer dyspepsia, as it was observed in 41.7% of the patients with ulcer and 74.5% of those with non-ulcer dyspepsia^[10].

In the present study, the prevalence, number of copies, distribution and genomic diversity of these genes were analysed in *H. pylori* clinical strains.

Regarding the association between the prevalence of the genes and the pathology of the patients, no considerations are to be made, given the reduced dimension of the sample and the reduced number of patients for each pathology. This association has already been made in detail for the population studied^[10].

The diversity of *jhp870* and *jhp649* sequences was evaluated in a sample of the *H. pylori* clinical strains through alignments and similarity plot representations, based on the nucleotide and predicted amino acid sequences. In general, the *jhp870* and *jhp649* genes displayed a high level of conservation among the gene, shown in the similarity plots.

However, in the *jhp870* gene, a region was identified, corresponding to the middle region designated by segment 2, which presents a high allelic diversity, where four distinct allelic variants, AI, AII, AIII and AV, were observed. The predominant and exclusive allele for the JHP870 proteins is the allele AI which is present in 75% of the strains.

For *jhp649* gene an extremely allelic diversity in the middle region was also detected, where common variants (AII and AIII), as well as distinct ones (AIV and AVI), from those found in the *jhp870* gene, were present. The predominant allele in the JHP649 proteins is the allele AII, which is Present in 82.9% of the strains, and it is the second most frequent allele in the JHP870 proteins (17.2%).

The most important conclusion of this study is the definition of allelic variants and the observations of their maintenance in all of the strains, independently of the geographic origin, pathology or type of population, which suggests their antiquity, and shows conservation up along the evolution. Furthermore, the existence of an allelic conservation in a polymorphic region implies that these alleles can influence the

que estes alelos podem ter alguma influência na função da proteína.

Esta diversidade alélica é também observada noutros membros da família das OMP de *H. pylori*, como é o caso das proteínas BabA, BabB e HopZ, que partilham um perfil conservado de segmentação do gene, traduzido na existência de pelo menos duas variantes alélicas, altamente conservadas^[13,7].

A maior parte dos membros da família das OMP possui uma forte similaridade nos terminais C e N sugerindo a sua necessidade para funções conservadas, enquanto que as regiões centrais variáveis poderão codificar funções únicas específicas de cada proteína^[13].

Este estudo vem demonstrar a existência de variabilidade alélica nestes dois membros (*jhp870* e *jhp649*) da família das OMP de *H. pylori*, e a observação de alelos exclusivos de cada gene, que poderá constituir uma base para a sua diferenciação genotípica. Também a genotipagem de outras estirpes clínicas, oriundas de diferentes regiões geográficas, contribuirá para o conhecimento da distribuição alélica e para a confirmação da sua conservação.

Por fim, e de modo a elucidar a importância destas variantes alélicas na patogenia de *H. pylori*, pretende-se estudar a associação entre alelos e outros factores de virulência já descritos para esta bactéria.

protein function.

This allelic diversity is also observed in other members of the *H. pylori* OMP family, as is the case of the BabA, BabB and HopZ proteins that share a conserved profile of gene segmentation, with the existence of at least two highly conserved allelic variants^[13,7].

The majority of the members in the OMP family possess a strong similarity in the C- and N- terminal, suggesting a need for conserved functions, while the varying central regions may codify unique functions that are specific to each protein^[13].

This study shows the existence of allelic diversity in these two members (*jhp870* and *jhp649*) of the *H. pylori* OMP family, and the observation of exclusive alleles for each gene, which may constitute a base for a genotypic differentiation. The study between genes of other clinical strains, from different geographic regions, can also contribute to the knowledge of allelic distribution and to confirm their conservation.

Finally, and in order to elucidate the importance of these allelic variants in the pathogenicity of *H. pylori*, we intended to study the association between alleles and other virulent factors already described for this bacteria.

Referências / References

- [1] Vicent P. Transmission and acquisition of *Helicobacter pylori* infection: evidences and hypothesis. Biomed Pharmacother 1995; 49:11-8.
- [2] Blaser MJ, Perez-Perez GI, Kleanthous H, Cover TL, Peek RM, Chyou PH, Stemmermann GN, Nomura A. Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. Cancer Res. 1995; 55:2111-2115.
- [3] Roggero E, Zucca E, Pinotti G, Pascarella A, Capella C, Savio A, Pedrinis E, Paterlini A, Venco A, Cavalli F. Eradication of *Helicobacter pylori* infection in primary low-grade gastric lymphoma of mucosa-associated lymphoid tissue. Ann Intern Med 1995; 122:767-769.
- [4] Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk IIP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzgerald LM, Lee N, Adams MD, Hickey EK, Berg DE, Gocayne JD, Utterback TR, Peterson JD, Kelley JM, Cotton MD, Weidman JM, Fujii C, Bowman C, Watthey L, Wallin E, Hayes WS, Borodovsky M, Karp PD, Smith HO, Fraser CM, Venter JC. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature 1997; 388:539-47.
- [5] Alm RA, Ling LS, Moir DT, King BL, Brown ED, Doig PC, Smith DR, Noonan B, Guild BC, deJonge BL, Carmel G, Tummino PJ, Caruso A, Uria-Nickelsen M, Mills DM, Ives C, Gibson R, Merberg D, Mills SD, Jiang Q, Taylor DE, Vovis GF, Trust TJ. Genomic sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature 1999; 397:176-180.
- [6] Boren T, Falk P, Roth KA, Larson G, Normark S. Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. Science 1993; 262: 1892-5.
- [7] Peck B, Ortkamp M, Diehl KD, Hundt E, Knapp B. Conservation, localization and expression of HopZ, a protein involve in adhesion of *Helicobacter pylori*. Nucleic Acids Research 1999; 27:3325-3333.
- [8] Dossumbekova A, Reindl W, Sarryeva A, Backert S, Kusters J, Prinz CK, Rad R. Evaluating proinflammatory and adherence properties of the *Helicobacter pylori* outer membrane protein 13 (Omp 13). *Helicobacter* 2005; 10:03.08A.
- [9] Alm RA, Bina J, Andrews BM, Doig P, Hancock RE, Trust TJ. Comparative genomics of *Helicobacter pylori*: Analysis of the outer membrane protein families. Infect Immun 2000; 68:4155-68.

- [10] Oleastro M, Carvalho A, Nunes B, Ménard A, Queiroz D, Monteiro L, Mégraud F. New pathogenic markers for peptic ulcer disease found in *Helicobacter pylori* strains isolated from Portuguese children. *Helicobacter* 2005; 10: A03.15.
- [11] Pitcher DG, Saunders NA, Owen RJ. Rapid extraction of bacterial genomic DNA with guanidinium thiocyanate. *Lett Appl Microbiol* 1989; 8:151-156.
- [12] Jukes TH, Cantor CR. Evaluation of protein molecules. Mammalian protein metabolism. Academic Press 1969; 21-132.
- [13] Pride DT & Blaser MJ. Concerted evolution between duplicated genetic elements in *Helicobacter Pylori*. *J Mol Biol* 2002; 316:629-642.