

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL



**Importância da variabilidade genética de
Helicobacter pylori na construção de uma vacina
de DNA**

Filipa Baltazar da Costa Vaz

MESTRADO EM MICROBIOLOGIA APLICADA

2009

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL



**Importância da variabilidade genética de
Helicobacter pylori na construção de uma vacina
de DNA**

Filipa Baltazar da Costa Vaz

MESTRADO EM MICROBIOLOGIA APLICADA

Sob orientação de

Professora Doutora Cecília Ribeiro da Cruz Calado

Professor Doutor Francisco André de Campos Pereira Dionísio

2009

Resumo

A bactéria gram-negativa *Helicobacter pylori* coloniza o estômago de cerca de 50% da população mundial e está associada ao desenvolvimento de patologias gástricas como a gastrite, úlcera péptica, neoplasia gástrica e linfoma de MALT. As estratégias de erradicação desta infecção baseiam-se em terapias com antimicrobianos, que apresentam limitações dado o aumento da resistência a antibióticos e a possibilidade de re-infecção após o tratamento. Nesta perspectiva, é relevante o desenvolvimento de uma vacina profiláctica e terapêutica que possa representar a variabilidade genética da bactéria. No presente trabalho, com base em dados de imunoproteômica e da relevância dos alvos antigénicos no processo de infecção, considerou-se para o desenho de uma vacina de DNA as seguintes cinco proteínas: NapA, HpaA, VacA, HomB e Omp9. De forma a desenhar uma vacina multi-antigénica que incluía estes cinco alvos, foi necessário definir para cada alvo uma região de menor dimensão conservada e rica em epítomos B, T ajudantes e se possível T citotóxicos, estimados por métodos *in silico*. Dada a ausência de informação quanto à variabilidade genética dos alvos NapA, HpaA e Omp9, foi necessário proceder à sua sequenciação em diversas estirpes de *H. pylori* associadas a patologias gástricas distintas, previamente amplificadas por PCR e PCR *touchdown*, respectivamente. Pretende-se desta forma contribuir para a construção de uma vacina de DNA multi-antigénica, baseada em fragmentos conservados e ricos em epítomos B, Th e Tc de cinco proteínas de *H. pylori*.

Palavras Chave

Helicobacter pylori, vacina de DNA, multi-antigénica, variabilidade genética, epítomos conservados.

Abstract

The gram-negative bacterium *Helicobacter pylori* colonizes the stomach of about 50% of the worldwide population and is associated with the development of gastric diseases, such as gastritis, peptic ulcer, gastric neoplasia and MALT lymphoma. Eradication strategies of this infection are based in antimicrobial drugs that present limitations due to the growing resistance to antibiotics and the possibility of re-infection after treatment. Therefore, it is relevant the development of a prophylactic and therapeutic vaccine that can represent the genetic variability of the bacteria. In the present work, based on immunoproteomic data and on the importance of the antigenic targets to the process of infection, it was considered for the construction of a DNA vaccine the following five proteins: NapA, HpaA, VacA, HomB and Omp9. To design a multi-antigenic vaccine that comprises all the five selected targets, it was necessary to define for each of them, a smaller region conserved and rich in B, T helper and ultimately T cytotoxic epitopes, estimated by *in silico* methods. Due to the lack of information on the genetic variability of the targets NapA, HomB and Omp9, it was necessary to proceed to its sequencing in several strains of *H. pylori* associated with different gastric pathologies, after PCR and PCR touchdown amplification. It is intended to contribute for the construction of a multi-antigenic DNA vaccine, based on conserved and rich B, T helper and T cytotoxic epitopes of five proteins of *H. pylori*.

Key Words

Helicobacter pylori, DNA vaccine, multi-antigenic, genetic variability, conserved epitopes.

Índice

Resumo	iii
Palavras Chave	iii
Abstract	iv
Key Words	iv
Índice	v
Abreviaturas	vi
Índice de Tabelas	vii
Índice de Figuras	vii
Agradecimentos	viii
Capítulo I. Introdução	1
Capítulo II. Selecção das proteínas alvo	8
Capítulo III. Variabilidade dos genes alvo <i>napA</i> , <i>hpaA</i> , <i>vacA</i> e <i>homb</i>	7
III.1. Amplificação e sequenciação dos genes alvo	7
III.1.1. Crescimento e extracção de DNA genómico de <i>Helicobacter pylori</i>	7
III.1.2. Amplificação dos genes alvo por reacção de polimerase em cadeia	8
III.1.3. Sequenciação dos genes alvo <i>napA</i> e <i>hpaA</i>	12
III.2. Análise da variabilidade dos genes alvo	13
III.2.1. Selecção e Recolha de Dados	13
III.2.2. Alinhamento das sequências nucleotídicas	14
III.2.3. Análise da conservação e variabilidade genética	14
III.2.4. Conclusões	21
III.3. Impacto do estudo da variabilidade genética na construção de uma vacina contra <i>Helicobacter pylori</i>	21
Capítulo IV. Selecção de epítomos	23
IV.1. Análise da antigenicidade e conservação dos epítomos	24
IV.2. Fragmentos alvo	25
Capítulo V. Conclusões e Perspectivas	32
Capítulo VI. Bibliografia	33
Anexos	37

Abreviaturas

µl	Micro litro
µM	Micro molar
A	Adenina
Aa	Aminoácido
ARB	<i>Average Relative Binding method</i>
C	Citosina
DNA	Ácido desoxirribonucleico
F	<i>Forward</i>
G	Guanina
HpaA	Hemaglutinina de ligação à neuraminil-lactose
IEDB	<i>Immune epitope database</i>
m/V	Massa/volume
mM	Mili molar
MALT	Tecido linfóide associado à mucosa
MHC	Complexo major de histocompatibilidade
Min	Minutos
MLST	<i>Multi Locus Sequence Typing</i>
NapA	Proteína de activação de neutrófilos
NCBI	<i>National Center for Biotechnology Information</i>
Ng	Nano grama
Nt	Nucleótido
OMP	<i>Outer membrane proteins</i>
p/v	peso/volume
Pb	Pares de base
PCR	Reacção de polimerase em cadeia
pmol	Pico mole
R	<i>Reverse</i>
Seg	Segundos
T	Timina
Ta	Temperatura de <i>annealing</i>
TAP	Transportador associado ao processamento de antigénio
Tc	Células T citotóxicas
Th	Células T <i>helper</i>
Tm	Temperatura de <i>melting</i>
U	Unidades
V	Volt
VacA	Citotoxina vacuolizante

Índice de Tabelas

Tabela III.1. Estirpes de <i>H. pylori</i> usadas para amplificação dos genes <i>napA</i> , <i>hpaA</i> e <i>omp9</i> .	8
Tabela III.2. Oligonucleótidos utilizados para a amplificação e sequenciação das regiões codificantes dos genes <i>napA</i> , <i>hpaA</i> e <i>omp9</i> .	9
Tabela III.3. Oligonucleótidos e temperaturas de <i>annealing</i> utilizadas nas reacções de sequenciação.	13
Tabela III.4. Diversidade dos genes alvo.	14
Tabela III.5. Análise para os genes <i>napA</i> , <i>hpaA</i> , <i>vacA</i> e <i>homB</i> da prevalência das substituições por transição e transversão.	18
Tabela IV.1. Características dos fragmentos seleccionados para cada alvo.	28

Índice de Figuras

Figura III.1. DNA genómico de <i>H. pylori</i> .	7
Figura III.2. Resultados da amplificação das sequências de interesse dos genes <i>napA</i> (A) e <i>hpaA</i> (B) por PCR em 29 estirpes de <i>H. pylori</i> .	10
Figura III.3. Géis dos PCRs do gene <i>omp9</i> .	11
Figura III.4. Géis dos resultados da amplificação de <i>omp9</i> por PCR <i>touchdown</i> .	12
Figura III.5. Análise filogenética dos genes alvo.	19
Figura IV.1. Análise pelo método de Emini e Bepipred da sequência consenso da proteína HomB, originada a partir de 120 sequências.	26
Figura IV.2. Antigenicidade para epítomos T dos fragmentos antigénicos HomB seleccionados.	27
Figura IV.3. Alinhamento das proteínas e análise de antigenicidade de epítomos B para a proteína NapA.	29
Figura IV.4. Comparação entre duas regiões da proteína NapA quanto à afinidade de ligação a moléculas MHC-II.	30
Figura IV.5. As regiões mais antigénicas são as que apresentam maior variabilidade na sequência proteica.	30

Agradecimentos

Agradeço ao Professor Doutor Rogério Tenreiro a oportunidade de integrar o Mestrado de Microbiologia Aplicada e pelo bom Professor que foi no passado, ao longo da minha Licenciatura em Microbiologia.

Muito obrigada às Professoras Doutoras Cecília Calado e Mónica Roxo-Rosa, por me terem dado a oportunidade de integrar no projecto e transformar o meu trabalho como bolsista de investigação na minha tese de mestrado. Agradeço ainda e não menos importante, a dedicação e a amizade expressas ao longo do tempo.

À Fundação para a Ciência e Tecnologia pelo apoio financeiro que efectuou ao projecto PTDC/Bio/69242/2006, ao abrigo do qual este trabalho foi efectuado.

Gostaria de agradecer à Faculdade de Engenharia da Universidade Católica Portuguesa por me ter recebido para efectuar trabalho de investigação no Laboratório de Engenharia Biomédica.

À Professora Doutora Cecília Calado, gostaria de demonstrar o meu apreço pela sua postura, profissional e pessoal, que muito gosto me deu em conhecer e desenvolver o meu trabalho com ela, bem como a sua compreensão e amizade expressas ao longo de um ano.

Um especial agradecimento à Professora Doutora Mónica Roxo-Rosa, pois foi ela quem me orientou no desenho de oligonucleótidos, na optimização de PCRs e nas sequenciações e abriu as portas à colaboração com o INSA, onde conheci a Doutora Sónia Pedro, a quem quero expressar a minha gratidão pela preocupação, profissionalismo e atenção que teve sempre para comigo e para com o meu trabalho.

À Professora Doutora Mónica Oleastro, quero agradecer as estirpes que me facultou e que compõem a maior parte da minha amostra de estudo, a disponibilidade para responder a questões e a oportunidade de ter passado uma manhã no seu laboratório com o Doutor João Benoliel que me foi extremamente útil e, ainda, a amabilidade em me fornecer artigos e documentos importantes para a escrita da minha tese.

À Inês e à Teresa, minhas colegas, agradeço toda a boa disposição, amizade, cumplicidade e companheirismo ao longo de um ano!

Aos meus Tios Rui e Tatyana pela amizade, compreensão, apoio e amor incondicional que me deram, não apenas ao longo do meu ano de mestrado, mas principalmente ao longo dos últimos dois anos da minha vida que foram cruciais para o momento de agora. Muito Obrigada!

À Rosa Maria Félix, porque me ensinou a viver e contornar os momentos mais difíceis da minha infância, adolescência e início de idade adulta. É a prova de que mesmo aqueles que partem por vezes deixam a sua marca para o resto das nossas vidas, pois para mim as suas palavras não perecem no tempo e cada dia que passa compreendo cada vez melhor tudo aquilo que ela me foi dizendo e ensinando, que se aplica a qualquer área da minha vida.

Muito Obrigada a estas três últimas pessoas mencionadas, pois ainda antes de eu própria acreditar em mim, já eles o faziam quando tudo e todos duvidaram!

A todos os meus amigos por todos os momentos – eles sabem quem são – porém com especial gratidão, à Cris, à Margarida e ao António.

Capítulo I. Introdução

Há mais de um século que existem referências de bactérias espiraladas no estômago do Homem, com indicação da associação entre a sua presença e o desenvolvimento de patologias gástricas. No entanto, esta associação nunca foi tida como muito relevante em termos clínicos, devido à existência de uma elevada prevalência da bactéria em indivíduos sem sintomatologia ou sinais clínicos de doença gástrica. Esta visão mudou drasticamente nos anos 80 com os trabalhos desenvolvidos por Barry Marshall e Robin Warren, que descreveram e isolaram estas bactérias espiraladas a partir de amostras gástricas, que mais tarde foram denominadas por *Helicobacter pylori*. Com a ingestão de suspensões da bactéria pelo próprio Marshall e por outros voluntários, provou-se os postulados de Koch para a gastrite (Marshall&Warren, 1984; Kusters *et al.*, 2006, Marshall *et al.*, 1985). Pela descoberta de *H. pylori* e o seu papel na gastrite e úlcera peptídica, Warren e Marshall foram galardoados em 2005 com o Nobel da Fisiologia e Medicina.

A *Helicobacter pylori* é uma bactéria gram-negativa do género *Helicobacter* da subdivisão ϵ das Proteobactérias, ordem *Campylobacteriales*, família *Helicobacteraceae*. Coloniza o estômago humano, é microaerófila e fastidiosa requerendo meios de cultura complexos para o crescimento, geralmente suplementados com sangue. A prevalência da infecção é variável em termos geográficos – nos países em vias de desenvolvimento atinge cerca de 80% da população e a taxa de infecção tem-se mantido constante; nos países desenvolvidos atinge cerca de 40%, observando-se um declínio da prevalência provavelmente devido às melhorias das condições de higiene e saneamento (Kusters *et al.*, 2006). O principal hospedeiro é o Homem podendo colonizar alguns primatas. Os mecanismos exactos de transmissão são ainda desconhecidos, mas pensa-se que as principais vias são a oral-oral e a fecal-oral e que a aquisição ocorre durante a infância por membros da família e pessoas próximas (Kusters *et al.*, 2006).

Este patógeno contribui para o desenvolvimento de gastrite aguda e crónica, úlcera péptica, metaplasia e neoplasia intestinal e linfoma de MALT (*mucosa-associated lymphoid tissue*). O risco de desenvolvimento destas patologias depende de diversos factores, tais como, a estirpe, o hospedeiro e o meio ambiente, observando-se na história natural da doença uma progressão gradual de gastrite até à neoplasia. Esta bactéria foi considerada pela Organização Mundial de Saúde como um carcinógeno classe 1. O diagnóstico da infecção pode ser realizado por amostras obtidas de exame endoscópico, da análise de ar expirado utilizando-se urea ^{13}C e pesquisa de antígenos em amostras fecais. O tratamento é geralmente efectuado a partir de terapias triplas combinadas contendo claritromicina e/ou metronidazol, e com um inibidor da bomba de prótons. No entanto, tem-se registado um aumento da resistência à claritromicina nomeadamente nos países do Sul da Europa (Malferttheiner *et al.*, 2007).

Dada a sua relevância em termos médicos, esta bactéria foi a primeira a ter dois genomas totalmente sequenciados, sendo que actualmente já existem 7 genomas totalmente sequenciados – 26695 (Tomb *et al.*, 1997), J99 (Alm *et al.*, 1999), HPAG1 (Oh *et al.*, 2006),

G27 (Baltrus *et al.*, 2009), P12, Shi470, B38. O *core* do genoma é constituído por 1111 genes determinantes para as características da espécie. Uma grande parte de genes auxiliares pertencem a genes específicos da estirpe e codificam, na sua maioria, para proteínas de função desconhecida. Outros exemplos de genes auxiliares são os que codificam para proteínas da ilha de patogenicidade *cag*, proteínas externas de membrana – OMP (*outer membrane proteins*) e para proteínas com função no sistema de restrição e modificação do DNA. O genoma apresenta uma zona de plasticidade que tem um conteúdo menor em GC do que o resto do genoma (Dong *et al.*, 2009). Esta bactéria apresenta uma elevada variabilidade genética devido a uma elevada taxa de mutação e recombinação, transferência de DNA e um sistema de reparação do DNA pouco eficiente. Por conseguinte, apresenta rearranjos a nível do cromossoma, elevada diversidade quanto ao número e conteúdo de genes e elevado número de alelos. A transmissão e infecção num novo hospedeiro não provocam alterações no genoma da bactéria, presumindo-se que o genoma actual tem capacidade suficiente para permitir a invasão bacteriana no homem e noutras espécies. Mas ao longo do curso da infecção num determinado hospedeiro, a bactéria pode adquirir genes auxiliares que lhe conferem vantagens na colonização, invasão e adaptação ao ambiente gástrico. Nesta perspectiva, a diversidade genética é relevante para a adaptação e sobrevivência da bactéria às condições dinâmicas e extremas do ambiente gástrico. Como suporte desta ideia acresce por exemplo a importância do mosaicismos do gene *vacA* na determinação da actividade da proteína – consoante o genótipo assim há produção de uma citotoxina activa e virulenta (exemplo genótipo s1/m1) ou não (genótipo s2/m2).

Recentemente tem-se levantado a hipótese de que é possível erradicar esta infecção a nível mundial dado que: i) este microrganismo não apresenta para além do estômago outros nichos ecológicos alternativos; ii) apenas é conhecido um único tipo de hospedeiro; iii) a infecção tende a diminuir com a melhoria das condições de higiene e sanitárias; iv) apesar de alguns insucessos da terapêutica considera-se que o seu uso continuado ao longo do tempo acabará por erradicar a bactéria. Apesar desta hipótese de erradicação mundial, o facto é que esta bactéria apresenta uma elevada capacidade de adaptação a um ambiente tão hostil e dinâmico como é o estômago humano. Esta elevada capacidade de adaptação, está na base da persistência desta bactéria ao longo da história evolutiva do Homem, sendo inclusive possível associar as migrações humanas às características genéticas da bactéria (Falush *et al.*, 2003). Para mais, sabe-se hoje que se trata de um patógeno com capacidade de invasão intracelular (Dubois&Borén, 2007), possuindo assim um nicho ecológico de fuga às agressões externas do ambiente gástrico. A infecção intracelular pode permitir um estadio de latência prolongado, sem sintomatologia associada, ocorrendo a transmissão e disseminação em períodos de reactivação. Por todos os motivos expostos, considera-se necessário a pesquisa de novos métodos terapêuticos, constituindo a vacina o método farmacológico que apresenta a maior razão benefício/ custo. Em termos de saúde pública a vacina apresenta diversas vantagens em relação aos antimicrobianos, como seja, financeiro, prevenção, terapêutica, evita a re-infecção que pode ocorrer no caso de terapêuticas à base de antimicrobianos, o aumento

de resistência a antimicrobianos deixa de ter relevância e pode prevenir patologias muito graves que por não apresentarem sintomatologia podem levar ao seu diagnóstico em estádios muito avançados

Já foram realizadas muitas tentativas de desenvolvimento de uma vacina contra esta infecção, mas todas sem sucesso, tal como revisto em Del Giudice *et al.* (2009). É opinião do grupo de investigação onde este projecto se insere que uma vacina eficiente contra esta bactéria deve ser constituída por diversos antígenos e não apenas por um (multi-antigénica). Isto porque, a bactéria apresenta uma variabilidade genética muito elevada e uma capacidade de adquirir novos antígenos ou de deixar de expressar outros de acordo com as suas necessidades de sobrevivência. Recentemente, Malfertheiner *et al.* (2008), uma vacina com as proteínas NapA, VacA e CagA, Zhou *et al.* (2009) construíram uma vacina com epítomos das proteínas UreB e HpaA, e Zhang *et al.* (2009) apenas com a proteína UreB. O objectivo final do projecto de investigação, onde o trabalho desta tese se insere, é o de desenvolver uma vacina de DNA multigénica contra *H. pylori* com um vector plasmídico. A grande vantagem das vacinas de DNA em relação às vacinas proteicas é a possibilidade de desencadear uma resposta equilibrada humoral (mediada por anticorpos produzidos por linfócitos B) e celular. Para desencadear uma resposta imunitária eficiente e de memória é essencial a activação de produção de linfócitos T ajudantes (T CD4+, também designadas por Th). As vacinas de DNA implicam a apresentação dos antígenos pelo complexo major de histocompatibilidade classe I (MHC – *major histocompatibility complex*) às células T CD8+ (também designados por Tc – T citotóxicos) e da apresentação de antígenos pelo complexo MHC-II às células T CD4+ (denominadas por linfócitos T ajudantes – Th do inglês “T helper”).

O objectivo da presente tese foi de definir um conjunto de alvos antigénicos relevantes para a construção da vacina multigénica. Também se pretendeu, para cada alvo, definir uma zona de menor dimensão que seja conservada e antigénica para epítomos B, Th e Tc. Nos casos em que não existia informação publicada sobre a diversidade genética dos alvos escolhidos, efectuou-se a sequenciação do alvo em diversas estirpes de *H. pylori* associadas a patologias gástricas distintas. Seleccionou-se um conjunto de fragmentos conservados e ricos em epítomos B, Th e Tc de alvos antigénicos relevantes, que constituem a base de construção de uma futura vacina de DNA eficiente contra a infecção por *H. pylori*.

Capítulo II. Selecção das proteínas alvo

Uma vacina eficiente contra *H. pylori* deve ser multi-antigénica dado que: i) a bactéria pode apresentar uma elevada variabilidade quanto à presença/ ausência de genes auxiliares; ii) a expressão de determinados genes depende do meio ambiente; iii) cada hospedeiro pode estar infectado por mais do que uma estirpe; iv) a bactéria pode adaptar-se ao meio ambiente através de mecanismos que implicam a manutenção de uma elevada variabilidade génica, ou seja o perfil antigénico pode variar de uma forma muito dinâmica ao longo do tempo.

A selecção dos alvos deve ser efectuada com base na importância do alvo no mecanismo de patogénese e interacção com o hospedeiro, na sua antigenicidade e se é uma proteína comum entre as diferentes estirpes de *H. pylori*. Os dados de imunoproteómica permitem seleccionar proteínas que sejam expressas na maioria das estirpes e que são antigénicas. A intensidade do sinal observado nos *western-blot*s obtidos a partir dos géis de electroforese 2-D, utilizando-se soro de pacientes, é proporcional à quantidade da proteína e da sua antigenicidade. Não se deve seleccionar alvos forçosamente muito antigénicos e/ou expressos em elevadas quantidades, uma vez que podem induzir fenómenos de tolerância antigénica ou mesmo de auto-imunidade ao descontrolarem a resposta do sistema imunitário (Jungblut *et al.*, 2000). Contudo, e dado que a resposta imunitária dependente da quantidade de antigénio (Del Giudice *et al.*, 2001), também não é aconselhável seleccionar alvos muito pouco expressos. Exemplo de uma consequência de aplicação destes critérios de selecção foi eliminação da Urease B como alvo antigénico, por ser expressa em elevadas quantidades e ser muito imunogénica. De forma a decidir entre diversas proteínas comuns e imunogénicas, considerou-se a relevância do alvo para o mecanismo de infecção e patogenicidade.

As proteínas alvo seleccionadas para a construção da vacina foram as seguintes: NapA (proteína de activação de neutrófilos), HpaA (hemaglutinina de ligação à neuraminil-lactose), VacA (citotoxina vacuolizante), HomB e Omp9/HopU. Estes alvos incluem proteínas de membrana antigénicas, dado constituírem um grupo de proteínas características da superfície exterior da bactéria, serem relativamente conservadas ao desempenharem funções estruturais essenciais e por desempenharem funções de relevo em processos de adesão, internalização, invasão, colonização e virulência (Carlshon *et al.*, 2006a). As seguintes proteínas foram identificadas por imunoproteómica como muito comuns a diversas estirpes e como antigénicas: NapA (McAtee *et al.*, 1998; JungBlut *et al.*, 2000; Kimmel *et al.*, 2000), VacA (JungBlut *et al.*, 2000; Sabarth *et al.*, 2002; Lock *et al.*, 2001) e HpaA (JungBlut *et al.*, 2000; Haas *et al.*, 2002; Lock *et al.*, 2001). Foram também realizados ensaios de imunização para a NapA (Rossi *et al.*, 2004; Malfertheiner *et al.*, 2008), para a VacA (Ghiara *et al.*, 1997, Rossi *et al.*, 2004, Malfertheiner *et al.*, 2008) e para a HpaA (Zhou *et al.*, 2009, Nystrom&Svennerholm, 2007). Os alvos HomB e Omp9 foram seleccionados por serem genes específicos da estirpe que conferem uma vantagem no decurso da infecção (Alm *et al.*, 2000). Para mais, a HomB constitui um co-marcador de estirpes associadas a úlcera péptica, induz respostas inflamatórias e participa na adesão da bactéria (Oleastro *et al.*, 2006, 2008 e 2009). Todas as

proteínas alvo seleccionadas são proteínas de superfície, com excepção da proteína VacA que é secretada para o exterior.

A proteína NapA constitui um marcador que permite discriminar entre úlcera e cancro (Khoder *et al.*, 2009), permite a adesão da bactéria à mucina, é um quimioatractor de monócitos, induz a produção pelos neutrófilos de espécies reactivas de oxigénio e estimula a produção de IL-12 e -13 contribuindo para a polarização da resposta imunitária para Th1. De facto, em biopsias de indivíduos infectados é vulgar encontrar infiltrados de neutrófilos e monócitos, que certamente contribuem para os danos epiteliais. Especula-se que esta proteína desempenhe uma função de protecção do DNA e armazenamento de ferro. *In vivo* há produção de anticorpos anti-NapA (Del Giudice *et al.*, 2001),

A proteína HpaA ou adesina A é um factor de colonização que induz a maturação e apresentação de antígenos por células dendríticas e é essencial à colonização e estabelecimento de infecção (Volland *et al.*, 2003, Carlsohn *et al.*, 2006b).

A VacA está implicada em diversos mecanismos de virulência, como por exemplo: inibe a fagocitose nos macrófagos, forma poros na membrana das células epiteliais, induz a formação de vacúolos nas células, induz apoptose e modula as respostas das células do sistema imunitário. Apesar de todas as estirpes possuírem o gene *vacA*, a expressão e actividade da proteína depende da heterogeneidade genética da região sinal (s) que codifica o péptido sinal e da região média (m) que codifica o domínio de ligação à p58. O genótipo da região s pode ser classificado em s1 ou s2 e a região m em m1 ou m2. O genótipo s1/m1 está associado a uma elevada actividade da proteína, s1/m2 a uma actividade intermédia e no genótipo s2/m2 a actividade é ausente. Por conseguinte, o genótipo s1/m1 encontra-se associado a úlcera péptica e carcinoma gástrico, estando também relacionado com o genótipo *cagA* positivo. A expressão proteica é variável ao longo do curso da infecção, presumindo-se que a melhoria e agravamento das úlceras possa estar associada com alterações de expressão da proteína (Kusters *et al.*, 2006).

No capítulo seguinte é analisado a variabilidade genética dos genes codificantes destas proteínas.

Capítulo III. Variabilidade dos genes alvo *napA*, *hpaA*, *vacA* e *homb*

III.1. Amplificação e sequenciação dos genes alvo

III.1.1. Crescimento e extracção de DNA genómico de *Helicobacter pylori*

Seleccção e crescimento das estirpes

Dada a ausência de informação quanto à variabilidade genética dos genes alvo *napA*, *hpaA* e *omp9*, foi necessário proceder à sua sequenciação em 29 estirpes clinicamente isoladas de indivíduos que sofriam de diferentes patologias gástricas, indicadas na tabela III.1. Considerou-se que este número de estirpes com características *vacA/cagA* tão diferentes dentro das patologias, a acrescentar as já existentes nas bases de dados, seria suficiente para atingir o objectivo do trabalho. A estirpe J99 foi utilizada como controlo positivo para optimização dos protocolos.

As estirpes foram semeadas em meio de Mueller Hinton (Mueller Hinton Broth, Difco, Becton Dickinson and Company Sparks, USA) suplementado com 1,5% de agar (Bacto™ Agar, Difco), 10% de sangue desfibrinado de cavalo (Probiologica, Belas, Portugal) e suplemento de crescimento *Helicobacter pylori Selective Supplement Dent* (Oxoid, Hampshire, Inglaterra). De seguida, foram incubadas a 37°C durante 48h em atmosfera de microaerofilia por meio de saquetas *Campygen* (Oxoid) dentro de uma jarra de anaerobiose.

Extracção e quantificação do DNA genómico

O DNA genómico foi extraído utilizando o *kit Wizard® SV Genomic DNA Purification System* (Promega Corporation, USA) com um protocolo elaborado por adaptação dos diferentes protocolos que o *kit* disponibilizava, dado que o aconselhado para extracção em bactérias gram-negativas não resultou. O DNA genómico obtido para as várias estirpes foi analisado em gel de agarose (Seakem® LE agarose BioWhittaker Molecular Applications, Rockland, USA) 0,8% (p/v) tampão TBE 1x concentrado (Tris-Borate-EDTA buffer 10x concentrate, Sigma-Aldrich, Inc., EUA). Como podemos observar na figura III.1 (nos poços 1 e 2 do gel) o DNA genómico de algumas estirpes quando extraído desta forma aparecia invariavelmente degradado, tendo-se optado nestes casos por uma extracção por lise em GES (isocianato de guanidina, EDTA e sarcosil). O DNA extraído foi quantificado com o leitor de placas Synergy™2 (BioTek® Instruments, Inc., USA) usando uma placa de 96 poços (Corning®, EUA).

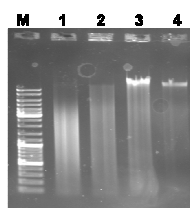


Figura III.1. DNA genómico de *H. pylori*. 1 – estirpe 26695 (não incluída neste estudo); 2 – estirpe 228/99; 3 – estirpe B23/99; 4 – estirpe JP26/95. M – marcador de pesos moleculares *GeneRuler™ 1 kb Plus DNA Ladder, ready-to-use* (Fermentas, Vilnius, Lituânia) (conf. Anexo I).

Estirpes	Patologia	<i>cagA</i>	<i>vacA</i>
J99	Úlcera duodenal	+	+
3432A/08	Úlcera duodenal	+	+
927/03	Úlcera duodenal	+	+
93/00	Úlcera duodenal	+	+
147/00	Úlcera duodenal	-	+
SS1	Úlcera gástrica	+	+
207/99	dispepsia não ulcerada	-	-
B1/99	dispepsia não ulcerada	-	-
839/99	dispepsia não ulcerada	-	-
2/99	dispepsia não ulcerada	-	-
228/99	dispepsia não ulcerada	-	-
600/99	dispepsia não ulcerada	-	-
440/02	dispepsia não ulcerada	-	-
506/02	dispepsia não ulcerada	+	+
1622/05	dispepsia não ulcerada	+	+
1094/03	dispepsia não ulcerada	-	-
1713/05	dispepsia não ulcerada	-	-
514/02	dispepsia não ulcerada	-	-
375/01	dispepsia não ulcerada	+	+
P9/92	Neoplasia gástrica	+	+
JP26/95	Neoplasia gástrica	-	+
B23/99	Neoplasia gástrica	+	+
JP22/96*	Neoplasia gástrica	+	+
B22/96*	Neoplasia gástrica	+	?
B56/98	Neoplasia gástrica	?	?
P3/92	Neoplasia gástrica	-	-
B46/95	Neoplasia gástrica	?	?
JP1/95	Neoplasia gástrica	-	-
A3/90	Neoplasia gástrica	+	+

Tabela III.1. Estirpes de *H. pylori* usadas para amplificação dos genes *napA*, *hpaA* e *omp9*. Das 29 estirpes, 13 foram gentilmente facultadas pela Professora Mónica Oleastro do INSA e 16 já existiam no laboratório, provenientes da Faculdade de Farmácia da Universidade de Lisboa. 5 estirpes são associadas a úlcera gástrica, 14 são estirpes associadas a dispepsia não ulcerada e 10 a neoplasia gástrica. Exceptuando as estirpes J99 (Americana), SS1 (Australiana), 514/02 e 375/01 (indivíduos portugueses de etnia africana), as restantes foram todas clinicamente isoladas de indivíduos portugueses. As estirpes assinaladas com * são dois clones isolados do mesmo indivíduo no mesmo ano com intervalo de meses. O símbolo ? indica que se desconhece a informação. Os sinais + e - indicam genótipo *vacA* s1 e s2, respectivamente e também a presença ou ausência de *cagA*, respectivamente.

III.1.2. Amplificação dos genes alvo por reacção de polimerase em cadeia

Desenho dos oligonucleótidos

Os pares de oligonucleótidos (ou *primers*) utilizados na amplificação dos genes de interesse foram desenhados manualmente de acordo com os seguintes critérios, utilizando-se as estirpes J99 e 26695 como referência: dimensão entre 18-30 nucleótidos (nt) de preferência idêntico para os dois *primers*; 40-60% de conteúdo GC, preferencialmente idêntico para os dois *primers*; nucleótidos de guanina ou citosina na extremidade 3' (extremidade onde a DNA polimerase inicia a sua actividade) para tornar a ligação à cadeia molde mais forte; evitar sequências complementares dentro do *primer* de modo a não haver formação de estruturas em *hairpin* e dentro de cada par de *primers* para se evitar a formação de dímeros; temperaturas de *melting* (Tm) dos dois *primers* o mais próximas possível. Cada par incluiu um *primer* para a extremidade 5' da região a amplificar (*primer forward* ou F) e um para a extremidade 3' da mesma região (*primer reverse* ou R). Os pares de *primers*: *napF/napR*, *hpaF/hpaR*, *ompF/ompR*, para amplificação dos genes *napA*, *hpaA* e *omp9* respectivamente, estão representados na tabela III.2.

Pretendeu-se amplificar e sequenciar as regiões codificantes de cada gene por inteiro. Assim, e apesar de aumentar a probabilidade de inespecificidade e, portanto, de insucesso, mas preferindo não perder informação, os oligonucleótidos foram, por vezes, desenhados alguns pares de bases ainda dentro da região promotora e já na região do terminador do gene, como se pode observar na tabela III.2. Relativamente aos genes *napA*, *hpaA* e *omp9* para além das estirpes com o genoma totalmente sequenciado, na altura de desenho dos oligonucleótidos, e do nosso conhecimento, não existia qualquer trabalho publicado com referência à amplificação de toda a região codificante destes genes.

Gene	Sequência 5' -> 3'	Características	Localização	PCR (pb)
<i>napA</i>	NapF - GGT GTA GGA TAG CGA TCA AGG	Tm: 50,6°C %GC:52,4 21nt	início 69 nt a montante do codão de iniciação	493
	NapR - GGG CTT GCA GCA TCC	Tm: 46,7°C %GC:66,7 15nt	410-424 nt	
	NapASeqF - GCA AGC GGA TGC GAT CGT GTT ATT TAT GAA AGT G	Tm: 70,2°C %GC:44,1 34nt	30-63 nt	
	NapASeqR - CCA AAT GGA TTT TTG CAA CTT GGC TAA TTG ATC	Tm: 66,8°C %GC:36,4 33nt	379-411 nt	
<i>hpaA</i>	HpaF - GGA TAG AAC GAT GAA AGC	Tm: 41,3°C %GC:44,4 18nt	início 10 nt a montante do codão de iniciação	777
	HpaR - CCT TTT AAT TCC TTG GC	Tm: 41,9°C %GC:41,2 17nt	751-767 nt	
	HpaASeqF - GCC TTT TAG GTG CGA GCG TGG TGG CTT TGT TAG TGG	Tm: 75,5°C %GC:55,6 36nt	44-79 nt	
	HpaASeqR - CCT TGG CGT CTT TTT GAT AAG ATT CTA AAT TC	Tm: 61,7°C %GC:34,4 32nt	726-757 nt	
<i>omp9</i>	OmpF - CTT TCA TTA GCT TTA GGC	Tm: 39,3°C %GC:38,9 18nt	16-33	2225
	OmpR - GCT TAG TAA GCG AAC AC	Tm: 36,8°C %GC:47,1 17nt	início 2 nt a jusante do codão stop	
	Omp9F - GCT TTA GGC TCG CTT TTA GTT	Tm: 49,9°C %GC:42,9 21nt	25-45 nt	
	Omp9R - GAG CCT TCT GTA TTT GAG TTC	Tm: 49,9°C %GC:42,9 21nt	2182-2202 nt	
	Omp9SeqF - AAA CCA TTC AAC CCT TAC AC	Tm: 45,7°C %GC:40,0 20nt	1075-1094 nt	
	Omp9SeqR - GTG TAA GGG TTG AAT GGT TT	Tm: 45,7°C %GC:40,0 20nt	1075-1094 nt	

Tabela III.2. Oligonucleótidos utilizados para a amplificação e sequenciação das regiões codificantes dos genes *napA*, *hpaA* e *omp9*. F significa *forward*, R *reverse*, Pb pares de base e aqueles que foram desenhados especificamente para sequenciação são denominados com a terminação Seq. Todos os *primers* foram sintetizados pela Thermo Fisher Scientific Biopolymers (Sedanstrasse Ulm, Alemanha).

Reacção de Polimerase em Cadeia

Todas as reacções de polimerase em cadeia (PCR do inglês *Polymerase Chain Reaction*) foram efectuadas num termociclador Biometra® Tpersonal (Göttingen, Alemanha) usando o kit de PCR *DreamTaq™ DNA Polymerase* e a solução de deoxinucleótidos *dNTPMix*

10mM each (Fermentas). 5 µl de produto de cada PCR foram analisados por electroforese em gel de agarose (Seakem® LE agarose) 1% (p/v) para o gene *omp9* e 1,5% (p/v) para os genes *napA* e *hpaA* em tampão TBE 1X concentrado (Tris-Borate-EDTA buffer 10x concentrate, Sigma-Aldrich) com 3µl do marcador *GeneRuler™ 1 kb Plus DNA Ladder, ready-to-use* (Fermentas) (conf. Anexo I). Nesta fase, consideramos que a nossa sequência alvo foi correctamente amplificada por identificação de um produto da reacção de PCR com o peso molecular esperado.

Após vários passos de optimização dos protocolos de PCR, as sequências alvo dos genes *napA* e *hpaA* foram conseguidas utilizando as seguintes condições: primeiro passo de desnaturação 95 °C durante 10 minutos (min), seguido de 40 ciclos de desnaturação a 95 °C durante 1 min, *annealing* a 47 °C durante 30 segundos (seg), e extensão a 72 °C durante 30 seg. A mistura de reacção continha 100 ng DNA genómico de *H. pylori*, 10 pmol de cada oligonucleótido, 200 µM de mixdNTPs, 3 mM de MgCl₂, 1U de Taq Polimerase e o respectivo tampão 1x concentrado, tudo isto para um volume final de 50 µl. Tal como se pode observar na figura III.2, a região de interesse destes dois genes foi amplificada em todas as estirpes incluídas no estudo (tabela III.1), que correspondem às bandas de ~500 pares de base (pb) e de ~800 pb, para o gene *napA* e *hpaA*, respectivamente. Podemos contudo observar para algumas estirpes a presença de algumas bandas para além da que tem o peso molecular esperado, que provavelmente resultam de amplificações inespecíficas. O padrão de bandas tidas como inespecíficas não é o mesmo de estirpe para estirpe.

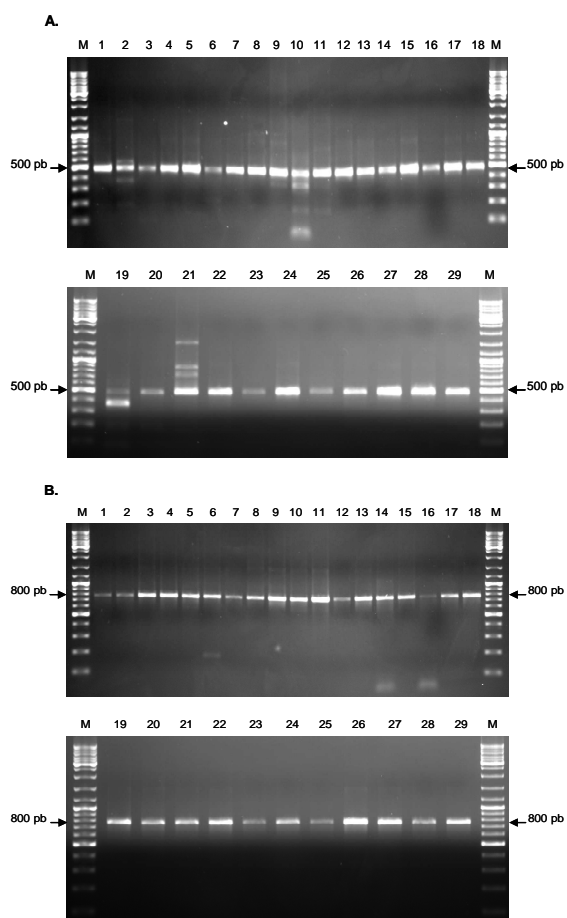


Figura III.2. Resultados da amplificação das sequências de interesse dos genes *napA* (A) e *hpaA* (B) por PCR em 29 estirpes de *H. pylori*. 5µl dos produtos de PCR foram analisados por electroforese em gel de agarose 1,5% (p/v), em tampão TBE 1x concentrado. 1 – J99; 2 – 3432A/08; 3 – 927/03; 4 – 93/00; 5 – 147/00; 6 – SS1; 7 – 207/99; 8 – B1/99; 9 – 839/99; 10 – 2/99; 11 – 228/99; 12 – 600/99; 13 – 440/02; 14 – 506/02; 15 – 1622/05; 16 – 1094/03; 17 – 1713/05; 18 – 514/02; 19 – 375/01; 20 – P9/92; 21 – JP26/95; 22 – B23/99; 23 – JP22/96; 24 – B22/96; 25 – B56/98; 26 – P3/92; 27 – B46/95; 28 – JP1/95; 29 – A3/90. M – marcador de pesos moleculares *GeneRuler™ 1 kb Plus DNA Ladder* (Fermentas).

A otimização do protocolo de PCR para amplificação da região codificante do gene *omp9* não foi tão bem sucedida. Além da dificuldade em amplificar o alvo pretendido, não foi possível eliminar as inespecificidades como se pode observar na figura III.3.A. Isto pode em parte ser devido a pelo menos um dos *primers* corresponder à sequência já fora da região codificante (terminador) (ver tabelalIII.2). Nesta perspectiva, foram desenhados novos oligonucleótidos, através da ferramenta PrimerSelect do programa Lasergene v8 (DNASTAR, EUA), designados por Omp9F e Omp9R (tabela III.2.). Procurou-se amplificar o gene inteiro usando o par Omp9F/Omp9R, contudo a inespecificidade manteve-se. Consequentemente, dividiu-se a amplificação da região de interesse em duas metades, a primeira amplificada pelo par Omp9F/Omp9SeqR e a segunda amplificada pelo par Omp9SeqF/Omp9R. Porém, os resultados foram bastante semelhantes aos obtidos para o par Omp9F/omp9R (figura III.3.B). Por conseguinte, optou-se pela estratégia de PCR *touchdown* (Don *et al.*, 1991), utilizando para o efeito o seguinte protocolo: primeiro passo de desnaturação 95 °C durante 10 min, seguido de 1 ciclo de desnaturação a 95 °C durante 1 min, *annealing* a 56 °C durante 30 seg, e extensão a 72°C durante 1 min, seguido de uma série de ciclos idênticos nos quais se foi progressivamente reduzindo a Ta em 2 °C até atingir os 46 °C; seguido de 25 ciclos com a Ta de 46°C; e por fim um passo de extensão a 72 °C durante 7 min.

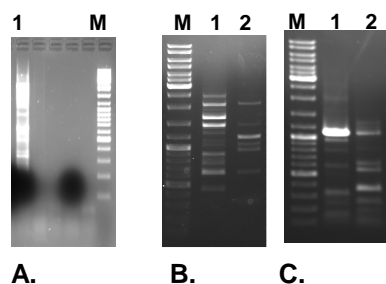


Figura III.3. Géis dos PCRs do gene *omp9*. A. Poço 1 – estirpe MP26/95; poços entre a estirpe e o marcador correspondem a controles negativos de várias mix que foram corridas no mesmo gel. **B.** Amplificação com o par omp9F/omp9R (1 – J99; 2 – SS1). **C.** par omp9F/omp9SeqR (poço 1) e omp9SeqF/omp9R (poço 2) (estirpe J99). Marcador 1Kb como indicado para os géis *napA* e *hpaA* da figura III.2.

A mistura de reacção continha 100 ng DNA genómico de *H. pylori*, 8 pmol de cada oligonucleótido, 200 µM de mixdNTPs, 3 mM de MgCl₂, 1,25 U de Taq Polimerase e o respectivo tampão 1x concentrado, tudo isto para um volume final de 20 µl. Deste modo, amplificou-se a primeira metade do gene com aproximadamente 1070 pb com o par Omp9F/Omp9SeqR e a segunda metade do gene de aproximadamente 1166 pb com o par Omp9SeqF/OmpR. Os resultados são apresentados na figura III.4., onde se podem observar ainda algumas inespecificidades. No entanto, observa-se também a presença das bandas que por terem o peso molecular esperado são provavelmente as regiões de interesse do gene *omp9*. Estas bandas foram extraídas e purificadas através do *Agarose Gel DNA Extraction kit* (Roche Diagnostics, Mannheim, Alemanha) estando prontas para serem sequenciadas, o que será feito num futuro próximo. Deve ser salientado que em 5 estirpes não foi possível amplificar as duas metades do gene (figura III.4, poços 4, 8, 12, 18 e 25) e que não se amplificou nenhuma das duas metades em duas outras estirpes (figura III.4, poços 9 e 16). Nestas últimas pode-se colocar a hipótese de se tratarem de estirpes que não têm o gene em questão no genoma. Observamos ainda que nalgumas estirpes (figura III.4, poços 2, 5, 15, 23 e 24), o

produto de PCR obtido por amplificação com o par de *primers* Omp9F/Omp9SeqR (primeira metade do gene) tem um peso molecular ligeiramente superior ao esperado, levantando a dúvida se se tratará do gene ou se haverá alguma inserção de nucleótidos na sua sequência.

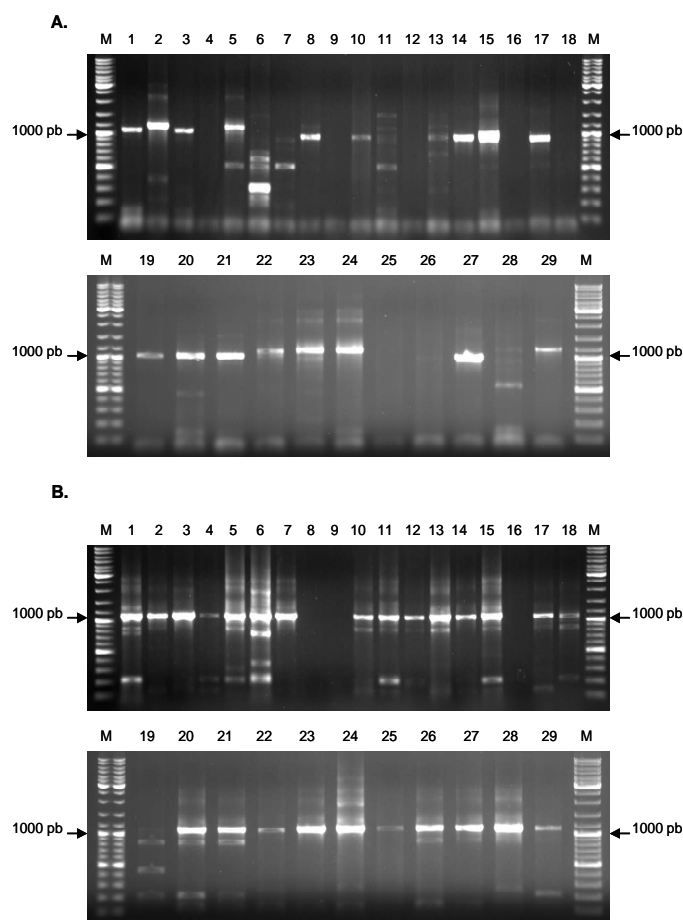


Figura III.4. Géis dos resultados da amplificação de *omp9* por PCR touchdown.

A. Primeira metade do gene com 1070 pb. **B.** Segunda metade do gene com 1166 pb de tamanho esperado. Amplificou-se as duas metades do gene em 22 estirpes: 1 – J99; 2 – 3432/08; 3 – 927/03; 5 – 147/00; 6 – SS1; 7 – 207/99; 10 – 2/99; 11 – 228/99; 13 – 440/02; 14 – 506/02; 15 – 1622/05; 17 – 1713/05; 19 – 375/01; 20 – P9/92; 21 – JP26/95; 22 – B23/99; 23 – JP22/96; 24 – B22/96; 26 – P3/92; 27 – B46/95; 28 – JP1/95; 29 – A3/90. Em 5 estirpes não foi possível amplificar as duas metades do gene: 4 – 93/00; 8 – B1/99; 12 – 600/99; 18 – 514/02; 25 – B56/98. Não se amplificou nenhuma das duas metades nas estirpes 9 – 839/99 e 16 – 1094/03. Na primeira metade do gene o tamanho da banda nas estirpes 3432/08, 147/00, 1622/05, JP22/96 e B22/96 parece ser maior do que o esperado.

III.1.3. Sequenciação dos genes alvo *napA* e *hpaA*

Purificação e quantificação dos produtos de PCR

Os produtos de PCR dos genes *napA* e *hpaA* foram purificados usando o *High Pure PCR Product Purification Kit* (Roche Diagnostics) e posteriormente quantificados pelo QubitTM fluorimeter usando Quant-iTTM dsDNA BR Assay Kit (Invitrogen, EUA).

Reacção de Sequenciação Cíclica

Ao todo foram sequenciados os produtos de PCR obtidos pela amplificação do gene *napA* de 22 estirpes (1622/05, 1713/05, 1094/03, 228/99 B1/99 600/99 506/02 207/99 440/02, 839/99, 93/00, 147/00, J99, 3432A/08, P3/92, JP1/95, JP26/95, B46/95, P9/92, A3/90, B22/96, JP22/96, e B23/99). Foram sequenciados os produtos de PCR obtidos pela amplificação do gene *hpaA* de 12 estirpes (1622/05, 1713/05, 514/02, B1/99, 2/99, 93/00, 147/00, 927/03, JP26/95, P9/92, A3/90 e B23/99). Inicialmente foram utilizados na sequenciação os oligonucleótidos da amplificação dos vários fragmentos, no entanto, mais tarde recorreu-se a

novos *primers* (tabela III.2) para se atingir este objectivo. Por motivos de tempo não foi possível a sequenciação de todos os genes, devido a diversas dificuldades técnicas assim como provavelmente devido à existência de um produto inespecífico em ambos os genes *napA* e *hpaA* que interfere com o sucesso da reacção. No entanto, as inespecificidades observadas nos resultados de PCR não deverão constituir a total razão do insucesso, visto que as estirpes 3432/08 e JP26/95 para o gene *napA* com bastante amplificações inespecíficas foram sequenciadas com sucesso, no entanto, talvez para as estirpes 2/99 e 375/01 as amplificações inespecíficas interfiram com a reacção de sequenciação. O tratamento dos resultados das sequenciações foi feito usando o software AB Sequence Scanner v1.0 (Applied Biosystems™, EUA).

Gene	Oligonucleótidos	T <i>annealing</i>
<i>napA</i>	napF	47°C
	napR	47°C
	napASeqF	70°C
	napASeqR	60°C
<i>hpaA</i>	hpaF	45°C
	hpaR	45°C
	hpaASeqF	75°C
	hpaASeqR	60°C

Tabela III.3. Oligonucleótidos e temperaturas de *annealing* utilizadas nas reacções de sequenciação. Todas as reacções foram feitas num volume de 10 µl com 0,8 µl de *Big Dye Terminator Ready Reaction v1.1* (Applied Biosystems™), 15 ng de DNA e 2 pmol de oligonucleótidos. Programa: 25 ciclos de desnaturação a 96 °C durante 10 seg, Ta 5 seg e extensão a 60 °C durante 4 min. As amostras foram analisadas com um ABI3130xl Genetic Analyzer (Applied Biosystems™).

III.2. Análise da variabilidade dos genes alvo

III.2.1. Selecção e Recolha de Dados

Para cada alvo antigénico (*NapA*, *HpaA*, *VacA*, *HomB* e *Omp9/HopU*) foram consideradas as sequências retiradas da base de dados UnitProtKB, as sequências das estirpes de genoma totalmente sequenciado pela página da internet¹ onde foram disponibilizadas e as obtidas da sequenciação dos genes apresentadas no capítulo III.1.3. Para o gene *napA* foram consideradas 29 sequências, das quais 22 foram obtidas neste trabalho. Para o gene *hpaA* foram consideradas 25 sequências, das quais 11 foram obtidas neste trabalho. Considerou-se 56 sequências do gene *vacA*, das quais 28 são de estirpes Ocidentais e 28 de estirpes Asiáticas. A estirpe Shi470 foi considerada como Asiática visto que de acordo com os dados publicados em relação à *VacA* está mais próxima destas do que das Ocidentais [www.expasy.ch/sprot/hamap/HELPS.html]); Foram consideradas 120 sequências do gene *homb*, sendo que 17 das estirpes apresentam duas cópias do gene. Devido ao número muito

1

Estirpes J99 e 26695 <http://genolist.pasteur.fr/PyloriGene/>

Estirpe P12 http://gib.genes.nig.ac.jp/single/main.php?spid=Hpyl_P12

Estirpe G27 http://hpylori.ucsc.edu/cgi-bin/hgGateway?db=heliPylo_G27

Estirpe Shi470 http://gib.genes.nig.ac.jp/single/main.php?spid=Hpyl_Shi470

Estirpe HPAG1 http://cmr.jcvi.org/cgi-bin/CMR/shared/AllGeneList.cgi?sub_org_val=nthp03

Estirpe B38 http://gib.genes.nig.ac.jp/single/index.php?spid=Hpyl_B38

reduzido de sequências do gene *omp9*, este foi descartado para efeitos de estudo de análise da variabilidade genética.

III.2.2. Alinhamento das sequências nucleotídicas

O alinhamento das sequências foi efectuado pelo programa MAFFT v6 (<http://align.bmr.kyushu-u.ac.jp/mafft/online/server/>), dado ser o programa com maior exactidão e rapidez para alinhamento de sequências muito semelhantes nomeadamente quando se desconhece o tipo de dados (Golubchik *et al.*, 2007). No programa MAFFT utilizou-se o método L-INS-i por ser o mais adequado para amostras com dimensão inferior a 200 sequências e que podem apresentar regiões conservadas únicas com *gaps* longos. De forma a corrigir alguns alinhamentos, estes foram visualizados através do programa Jalview (Clamp *et al.*, 2004).

III.2.3. Análise da conservação e variabilidade genética

A análise da variabilidade genética baseada nos alinhamentos efectuados no ponto anterior foi realizada através do programa DnaSP v5 (Librado&Rozas, 2009). A análise de substituições, transições, transversões e das relações filogenéticas foi efectuada com o programa MEGA v4 (Tamura *et al.*, 2007). As árvores filogenéticas foram desenhadas através do método *Neighbour-Joining* (Saitou&Nei 1987). Os resultados obtidos encontram-se na tabela III.4, tabela III.5 e figura III.5. Nestes estudos não foram consideradas as zonas de *gaps* ou com ausência de dados. A região codificante foi definida por utilização do código genético universal.

	<i>napA</i> n=29	<i>hpaA</i> n=25	<i>vacA</i> n=56	<i>homb</i> n=120
Nº locais polimórficos, S	38	36	778	836
Nº alelos	25	22	55	118
Diversidade do haplotipo, Hd	0,9877	0,987	0,999	0,9997
Diversidade nucleotídica, π	0,0329	0,04225	0,09787	0,07834
Teste de Watterson, θ	0,03612	0,04414	0,09225	0,10464
Teste de Tajima <i>D</i>	-0,33150*	-0,16203*	0,2195*	-0,84256*
Recombinação, R	44	75,9	13,5	65,9
R entre locais adjacentes, r	0,1066	0,1188	0,0035	0,0331
Razão dN/dS	3,77	4,03	3,36	3,42

Tabela III.4. Diversidade dos genes alvo. Análise genética efectuada aos genes alvo em relação ao número de locais polimórficos (S); do número de alelos; da diversidade de haplotipo (Hd) determinado pela fórmula de Nei (1987) que implica a substituição de 2n por n; da diversidade nucleotídica (π) que representa a média do número de nucleótidos diferentes por local entre duas sequências (Nei, 1987); teste de Watterson ($\theta = 4N\mu$, em que N representa a dimensão da população e μ a taxa de mutação em cada nucleótido por geração) (Watterson, 1975); teste de Tajima (*D*) para avaliar a hipótese nula de que as mutações são neutras (* indica que os resultados não são significativos: $P > 0,10$) (Tajima, 1989); taxa de recombinação (R) (Hudson, 1987); taxa de recombinação entre locais adjacentes ($r = R/L$, em que L representa a média da distância em nucleótidos entre os locais mais distantes sob recombinação); razão entre o número de substituições não sinónimas (dN) e o número de substituições sinónimas (dS).

Da análise efectuada, observou-se que todos os genes apresentam um elevado número de alelos, obtendo-se quase um número de alelos iguais ao número de sequências

analisadas: o gene *napA* apresentou 25 alelos em 29 sequências analisadas, o gene *hpaA* 22 alelos em 25 sequências, o gene *vacA* 55 alelos em 56 sequências e o gene *homb* apresentou 118 alelos em 120 sequências. Logo estamos perante um elevado polimorfismo não existindo, portanto, nenhum alelo predominante na população, inferindo-se assim ausência de clonalidade em *Helicobacter pylori*. Estes resultados estão de acordo com o observado por Go *et al.* (1996), que registaram 73 perfis alélicos através de electroforese enzimática multi-locus (MLST – *Multi-Locus Sequence Typing*) aplicada a 74 estirpes. Suerbaum *et al.* (1998) reportaram que *H. pylori* apresenta uma estrutura clonal apenas durante um curto período após transmissão, tendo identificado um elevado número de alelos para os genes *vacA*, *flaA* e *flaB*, tal como Kansau *et al.* (1996) para o gene *ureC/glmM* e Pan *et al.* (1997) para o gene *cagA*.

O teste de Watterson (1975) foi efectuado de modo a estimar a taxa de mutação da população. Observou-se que todos os genes apresentaram uma taxa de mutação (θ) próxima do da diversidade nucleotídica (π). Os genes *napA* e *hpaA* apresentaram uma taxa de recombinação entre locais adjacentes (r) superior à taxa de mutação (θ), indicando que a variabilidade observada é devida maioritariamente a fenómenos de recombinação. Pelo contrário, os genes *vacA* e *homb* apresentaram uma taxa de mutação superior à taxa de recombinação entre locais adjacentes, indicando que nestes casos a variabilidade genética depende maioritariamente de mutações espontâneas.

O teste estatístico D de Tajima (1989) permite avaliar se está a ocorrer uma evolução neutra (isto é resultante do acaso). Os resultados do teste D não foram estatisticamente significativos para nenhum dos genes. No entanto, para o gene *vacA* o teste D tem um valor positivo, indicando a ocorrência de *balancing selection*. Os genes *napA*, *hpaA* e *homb* apresentaram um teste D negativo que deixa em aberto as possibilidades de estar provavelmente a ocorrer uma selecção estabilizadora ou uma selecção direccionada que pode implicar uma subdivisão da população. Para os quatro genes a razão dN/dS é muito superior a 1, pelo que podemos concluir que está a ocorrer uma diversificação genética sob uma selecção positiva. Os resultados do teste D de Tajima não são estatisticamente significativos para nenhum dos genes. Estes resultados foram comparados com os resultados do teste D de Fu&Li (1993), em relação ao qual os resultados também não apresentaram significado estatístico excepto para o gene *homb* ($P < 0,05$). Esta ausência de significado estatístico, poderá prender-se com o facto de as fórmulas não entrarem em conta com o número de alelos, que é um parâmetro que caracteriza uma amostra. Entram apenas em conta com a dimensão da amostra e o número de locais polimórficos. Este último, por sua vez, é maior quanto maior a amostra, dado que maior a probabilidade de encontrar sequências de DNA diferentes sendo ainda um parâmetro que define o número de alelos da população. Em suma, numa amostra em que o número de alelos é extremamente elevado, talvez seja preponderante entrar com este factor intrínseco da amostra nos testes de neutralidade, atendendo ainda ao facto de que se trata de uma característica que nos permite inferir sobre a estrutura da população e, por conseguinte, constitui um factor de impacto na determinação do tipo de forças selectivas que se sobrepõem e determinam o futuro evolutivo da população.

Tendo em conta os resultados obtidos pode-se indicar que está provavelmente a ocorrer uma selecção equilibrada para todos os genes avaliados e para os quais existe a manutenção de diversidade genética, típica de populações panmíticas, como a de *Helicobacter pylori*. É de realçar que estas conclusões são ainda mais limitadas no caso dos genes *napA* e *hpaA*, em que se considerou uma amostra de pequena dimensão. No que respeita *homb* e provavelmente *napA* e *hpaA*, está em curso uma selecção direccionada. Pode-se especular que esta selecção seja devida à alteração recente de hábitos de higiene na população humana, que pode tornar alelos até então mais vantajosos em menos vantajosos, redireccionando a selecção de variantes antigénicas. Este tipo de selecção poderá já ter ocorrido para o gene *vacA*, justificando assim a diferença observada em termos geográficos (Suerbaum *et al.*, 1999). De facto o número de alelos detectados para todos os genes está bastante acima da taxa de mutação observada, apontando para a hipótese de uma *balancing selection*, que vai ao encontro com uma estrutura populacional panmítica. Para mais, o número elevado de substituições não-sinónimas observado em todos os genes, indica que a selecção positiva e direccionada esteja a ocorrer, provavelmente, em todos os antigénios. Wang *et al.* (1999) reportam que o elevado polimorfismo se prende com uma elevada taxa de mutação. No entanto, a elevada frequência de eventos de recombinação parecer ser a principal causa para a variabilidade em *H. pylori* (Achtmann&Suerbaum, 2000; Suerbaum *et al.*, 1998; Falush *et al.*, 2001). Achtmann&Suerbaum (2000) publicaram um comentário em resposta à publicação de Wang *et al.* (1999) argumentando que *H. pylori* deverá ser comparado a *Neisseria meningitidis* já que colonizam o mesmo hospedeiro, são transformáveis e apresentam variação de fase, pelo que na sua publicação de 1998, Suerbaum *et al.*, efectuaram um estudo de recombinação em *Helicobacter pylori* onde compararam os resultados com *N. meningitidis*. Neste estudo, conclui-se que *H. pylori* é panmítico, apresenta uma elevada taxa de recombinação e encontra-se em equilíbrio de ligação (*linkage equilibrium*). Também Go *et al.* (1996) referem a ausência de *linkage disequilibrium*. Mais tarde em 1999, estes autores concluem que não só ocorre recombinação quando um indivíduo está infectado por mais do que uma estirpe, como várias rondas de recombinação (Suerbaum *et al.*, 1999). Foi devida à elevada variabilidade por recombinação que conduz à impossibilidade de detecção de clonalidade entre estirpes próximas, que Kuipers *et al.* (2000) introduziram o conceito de quasiespécie em *H. pylori*, típico de vírus de RNA.

Transições e Transversões nos genes alvo

gene *napA*

	A	T	C	G
A	-	2.11	1.74	9.96
T	3.17	-	23.61	1.53
C	3.17	28.66	-	1.53
G	20.68	2.11	1.74	-

Frequência dos nucleótidos: A: 0.371; T: 0.247; C: 0.203; G: 0.179. Razão entre as taxas de transição em relação às taxas de transversão para as purinas $k_1 = 6.521$, e para as pirimidinas de $k_2 = 13.594$. Taxa global de transição/transversão **$R = 3.494$**

gene *hpaA*

	A	T	C	G
A	-	1.27	0.75	18.69
T	1.52	-	14.37	0.84
C	1.52	24.4	-	0.84
G	33.79	1.27	0.75	-

Frequência dos nucleótidos: A: 0.348; T: 0.29 (T/U), 0.17 (C), 0.192 (G). Razão entre as taxas de transição em relação às taxas de transversão para as purinas $k_1 = 22.22$ e para as pirimidinas $k_2 = 19.267$. Taxa global de transição/transversão **$R = 7.143$**

gene *vacA*

	A	T	C	G
A	-	4.29	3.12	13.12
T	5	-	15.58	3.71
C	5	21.41	-	3.71
G	17.65	4.29	3.12	-

Frequência dos nucleótidos: A: 0.31; T: 0.266; C: 0.194; G: 0.23. Razão entre as taxas de transição em relação às taxas de transversão para as purinas $k_1 = 3.533$ e para as pirimidinas $k_2 = 4.992$. Taxa global de transição/transversão **$R = 1.706$**

gene *vacA* estirpes Asiáticas

	A	T	C	G
A	-	3.22	2.35	13.03
T	3.85	-	18.75	2.76
C	3.85	25.68	-	2.76
G	18.19	3.22	2.35	-

Frequência dos nucleótidos: A: 0.316; T: 0.264; C: 0.193; G: 0.226. Razão entre as taxas de transição em relação às taxas de transversão para as purinas $k_1 = 4.723$ e para as pirimidinas $k_2 = 7.975$. Taxa global de transição/transversão **$R = 2.481$**

gene *vacA* estirpes Ocidentais

	A	T	C	G
A	-	4.45	3.31	13.65
T	5.04	-	14.82	3.69
C	5.04	19.90	-	3.69
G	18.63	4.45	3.31	-

Frequência dos nucleótidos: A: 0.306; T: 0.27; C: 0.201; G: 0.224. Razão entre as taxas de transição em relação às taxas de transversão para as purinas $k_1 = 3.696$ e para as pirimidinas $k_2 = 4.47$. Taxa global de transição/transversão **$R = 1.666$**

gene *homb*

	A	T	C	G
A	-	3.66	2.45	13.12
T	4.05	-	15.79	2.37
C	4.05	23.6	-	2.37
G	22.44	3.66	2.45	-

Frequência dos nucleótidos: A: 0.323; T: 0.292; C: 0.195; G: 0.189. Razão entre as taxas de transição em relação às taxas de transversão para as purinas $k_1 = 5.544$ e para as pirimidinas $k_2 = 6.449$. Taxa global de transição/transversão **$R = 2.239$**

Tabela III.5. Prevalência das substituições por transição e transversão para os genes *napA*, *hpaA*, *vacA* e *homb*. Também estão apresentadas as frequências de ocorrência de cada nucleótido e as razões das taxas de transição e transversão. Os valores são referentes à probabilidade de substituição de uma base por outra. $R = [A^*G^*k_1 + T^*C^*k_2]/[(A+G)^*(T+C)]$. Todas as regiões que apresentam *gaps* foram excluídas da análise. A substituição de nucleótidos foi estimada pelo método de *Maximum Composite Likelihood*.

Observou-se que para todos os genes a taxa de transição é superior à de transversão, de acordo com o registado por Wang *et al.* (1999). Para os genes *napA* e *hpaA*, a taxa de transição entre pirimidinas é maior do que entre purinas, pelo contrário para o gene *vacA* as transições entre purinas ocorrem em muito maior número que as transições entre as pirimidinas. Já o gene *homb* apresentou uma taxa de transições entre pirimidinas e purinas muito semelhante. Nos genes *vacA* e *homb* a taxa de mutação é mais preponderante para a variabilidade do que a recombinação, pelo contrário os genes *napA* e *hpaA* a recombinação é mais preponderante para a variabilidade. Deste modo, e acordo com os dados publicados referidos acima, pode-se especular que a recombinação gera maior variabilidade que a mutação, sendo portanto o principal factor de manutenção da diversidade alélica e da estrutura panmítica da população. De todos os genes, o *hpaA* é o que apresenta maior valor de *R*. Relativamente ao gene *vacA*, observa-se que as estirpes Asiáticas apresentam maior variabilidade genotípica que as estirpes Ocidentais e uma predominância das transições das pirimidinas relativamente às purinas, sendo que as transições ocorrem muito mais comumente do que as transversões. Pelo contrário, nas estirpes Ocidentais, o *R* é bastante inferior, sendo que não se observa uma predominância tão elevada das transições relativamente às transversões e, ainda, que a probabilidade de ocorrência das transições entre pirimidinas é muito próxima à das purinas, semelhante ao observado para o gene *homb*.

Análise Filogenética

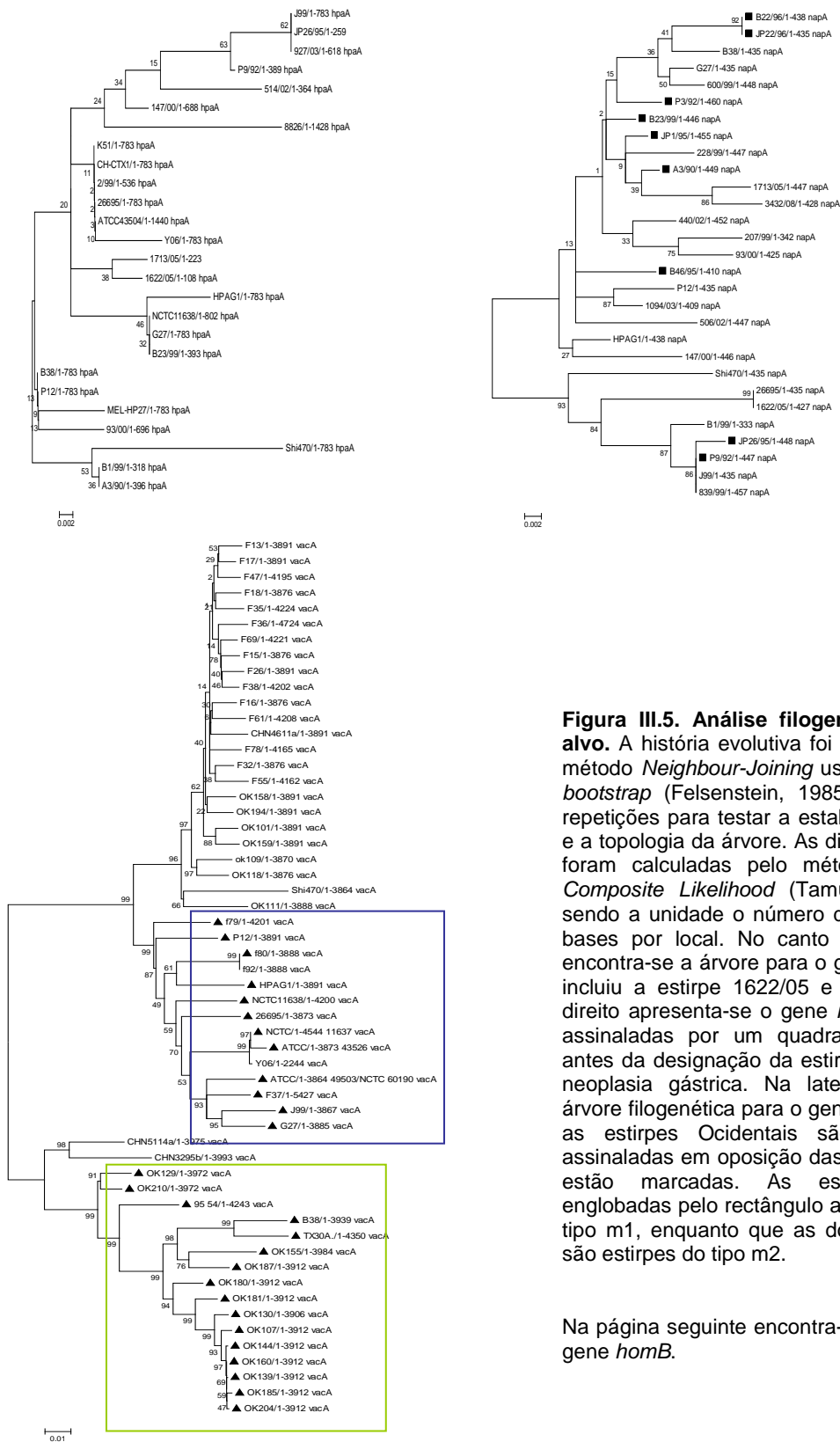
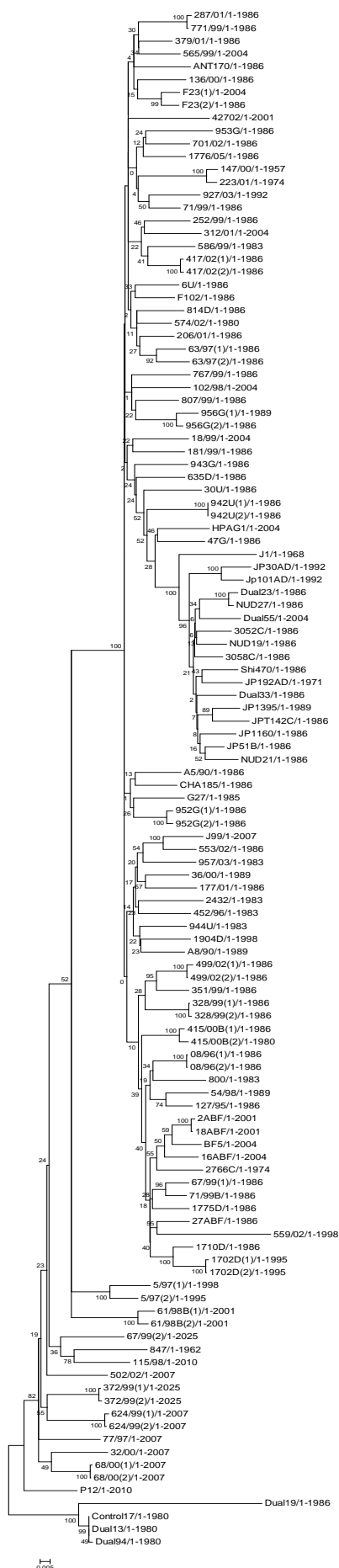


Figura III.5. Análise filogenética dos genes alvo. A história evolutiva foi inferida através do método *Neighbour-Joining* usando o método de *bootstrap* (Felsenstein, 1985) através de 500 repetições para testar a estabilidade dos ramos e a topologia da árvore. As distâncias evolutivas foram calculadas pelo método de *Maximum Composite Likelihood* (Tamura *et al.*, 2004), sendo a unidade o número de substituições de bases por local. No canto superior esquerdo encontra-se a árvore para o gene *hpaA* onde se incluiu a estirpe 1622/05 e no canto superior direito apresenta-se o gene *napA* – as estirpes assinaladas por um quadrado imediatamente antes da designação da estirpe são estirpes de neoplasia gástrica. Na lateral encontra-se a árvore filogenética para o gene *vacA*, sendo que as estirpes Ocidentais são as que estão assinaladas em oposição das Asiáticas que não estão marcadas. As estirpes Ocidentais englobadas pelo rectângulo azul são estirpes do tipo m1, enquanto que as do rectângulo verde são estirpes do tipo m2.

Na página seguinte encontra-se a árvore para o gene *homb*.



A avaliação da associação entre as características dos genes alvo e o tipo de patologia é dificultada pelas estirpes em causa não abrangerem todas as patologias de uma forma equitativa e pela limitação da dimensão de algumas amostras. Contudo, no gene *napA*, num conjunto albergando desde a estirpe B22/96 até à A3/90 (leitura na vertical), a maior parte são de neoplasia gástrica. Khoder *et al.* (2009) reportaram a proteína NapA como discriminadora entre úlcera e cancro. Porém, como a amostra apresenta uma dimensão pequena e não existe igual número de estirpes de cada patologia, não se pode concluir sobre esta relação. No gene *hpaA*, existem vários agrupamentos de estirpes, mas devido à reduzida dimensão da amostra, não é possível especular a razão deste agrupamento. Os genes *napA* e *hombB* apresentam ainda menor clonalidade que o gene anterior. No que respeita ao gene *vacA*, existe uma notória divisão entre as estirpes Asiáticas e as estirpes Ocidentais tal como esperado (Pan *et al.*, 1999), havendo dentro deste último grupo uma subdivisão de estirpes com base no tipo m1/m2. Algumas estirpes Asiáticas foram agrupadas juntamente com as estirpes Ocidentais. Não foi possível obter informação sob a origem da estirpe F92, que no entanto apresenta o gene *vacA* com características semelhantes às estirpes ocidentais. A análise das seqüências das estirpes F92 e F80 mostram que são idênticas. No entanto, Yamazaki *et al.* (2005) apresentam-nas com origem geográfica diferente e genótipos *vacA* e *cagA* diferentes e na base de dados localizam-se com uma entrada de gene distinta para cada estirpe mas apenas uma entrada comum para a proteína. Suerbaum *et al.* (1999) comentaram que os grupos clonais entre estirpes Asiáticas e Ocidentais em *vacA* poderão ser o resultado de uma selecção de variantes antigénicas que apresentam maior vantagem em determinadas regiões geográficas. Esta hipótese está de acordo com os dados obtidos: está em curso uma selecção positiva numa estrutura panmítica mantida por *balancing selection*. Aparentemente esta selecção já ocorreu para *vacA* e talvez com a evolução e a migração do Homem continue a ocorrer.

III.2.4. Conclusões

Em relação aos estudos efectuados para os genes alvo *napA*, *hpaA*, *vacA* e *hombB*, observou-se que o principal mecanismo que está na origem de variabilidade genética depende do tipo de gene, sendo que para os genes *napA* e *hpaA* a recombinação é o factor que mais contribui para a variabilidade, enquanto que para o gene *vacA* e *hombB* é a mutação. Todos os genes analisados apresentaram uma elevada diversidade nucleotídica, apresentando um número de alelos quase igual ao número de sequências analisadas, de acordo com uma estrutura populacional panmítica. Esta estrutura é mantida por uma *balancing selection*, em que para os genes *napA*, *hpaA* e *hombB* está provavelmente a ocorrer uma nova selecção de alelos mais vantajosos devido a modificações nos hábitos de higiene do hospedeiro. Esta dedução é baseada na observação do elevado número de substituições não sinónimas registadas em todos os genes e dos testes *D* que apontarem para uma selecção direccionada. No caso de *H. pylori*, esta selecção direccionada tende a gerar alelos diferentes sem tender para a clonalidade. O facto dos testes *D* não apresentarem significado estatístico, poderá prender-se com a limitação das fórmulas por não entrarem em conta com o parâmetro número de alelos.

III.3. Impacto do estudo da variabilidade genética na construção de uma vacina contra *Helicobacter pylori*

A panmicticidade em *Helicobacter pylori* prende-se essencialmente a três motivos: o nicho ecológico que apresenta condições variáveis de hospedeiro para hospedeiro e as diferentes respostas imunitárias observadas ao longo da vida do hospedeiro e entre hospedeiros (Kuipers *et al.*, 2000). Assim sendo, a manutenção de uma elevada variabilidade genética é crucial à sobrevivência e perpetuação desta bactéria.

No capítulo anterior observou-se que os genes considerados para o desenho da vacina apresentam uma elevada taxa de substituições não-sinónimas e portanto grande variabilidade na sequência em aminoácidos. Nesta perspectiva, depreende-se que houve tempo evolutivo suficiente para ocorrer selecção e manutenção de uma elevada variabilidade genética nos genes alvo considerados, que levou a proteínas antigénicas diferentes de estirpe para estirpe mas funcionais.

Recentemente ocorreram três alterações drásticas na população humana com consequências para *H. pylori* – a melhoria das condições higiénicas e sanitárias, a modificação de hábitos e tipo de alimentação e a introdução de terapêuticas antimicrobianas. Por conseguinte, para o patógeno, menor é a probabilidade de sobrevivência e portanto, menor a probabilidade de transmissão. Observou-se neste trabalho que para os genes alvo está em curso uma selecção positiva direccionada. Isto significa que está a ocorrer nova selecção de variantes antigénicas mais vantajosas às condições com que o microrganismo se depara actualmente. Analisando os codões onde ocorrem substituições não-sinónimas, observa-se por vezes que nesses mesmos codões ocorrem substituições sinónimas. Neste sentido, é passível

de se especular que existe a probabilidade de eventualmente, com o decorrer da evolução, estas alterações se transformarem em substituições não sinónimas.

Não existindo um alelo mais predominante na natureza, para a construção de uma vacina infere-se que de modo a cobrir a grande variabilidade, será preferível a selecção de epítomos a partir de regiões conservadas do DNA o quanto possível. Tendo sido determinadas as regiões nos genes onde ocorrem mutações sinónimas, pode-se também o quanto possível excluir os aminoácidos que são codificados por codões onde é comum a ocorrência de substituições sinónimas, pois apesar de não se traduzirem numa alteração do aminoácido codificado, sabe-se que se trata de uma região variável do gene.

O capítulo seguinte trata da selecção de epítomos das proteínas codificadas por estes genes.

Capítulo IV. Selecção de epítomos

Os métodos *in silico* são extremamente úteis e cada vez mais cruciais no desenho de novos fármacos nomeadamente de vacinas. No entanto, e embora todos os avanços observados nos últimos tempos, estes métodos continuam a apresentar limitações. Enquanto que os métodos de estimação de epítomos B e Tc apresentam estimativas mais próximas da realidade *in vivo*, os métodos de estimação de epítomos Th apresentam constrangimentos relativos à elevada degenerescência dos motivos de ligação entre os epítomos e as moléculas de MHC-II (Wang *et al.*, 2008). Outro constrangimento deste método é a dimensão do péptido que se liga à molécula de MHC e que irá ser apresentado e reconhecido pelas células T. Enquanto que os epítomos que se ligam a MHC-I não devem exceder os 9 aminoácidos, os epítomos MHC-II podem assumir dimensões variáveis dado o local de ligação poder acomodar epítomos superiores a 9 aminoácidos. Neste estudo, considerou-se epítomos de 9 aminoácidos, dado que o *core binding motif* em ambas as moléculas de MHC (I e II) ser de 9 aminoácidos. À parte das suas limitações, as técnicas *in silico* são extremamente úteis e adequadas para comparação de epítomos e fragmentos, orientando a escolha do(s) melhor(es) epítomo(s) para testes futuros *in vitro* e *in vivo* de imunogenicidade e antigenicidade, poupando tempo, economizando recursos, permitindo saber quais os resultados esperados o que facilita e promove a qualidade do trabalho de investigação.

Uma vacina tem por objectivo a prevenção de doença mas pode também actuar como terapêutica. Actualmente, considera-se que uma vacinação profiláctica e terapêutica eficiente contra *Helicobacter pylori* está dependente de uma resposta tipo Th2 (Kupsters *et al.*, 2006). Nesta perspectiva, a ligação dos epítomos a moléculas MHC-II foi tida como critério crucial e a afinidade dos epítomos para moléculas MHC-I menos relevante.

Através das sequências proteicas da base de dados UnitProtKB referentes às mesmas estirpes utilizadas para análise da variabilidade genética, juntou-se as sequências proteicas das estirpes sequenciadas dos genes *napA* e *hpaA*, traduzidas através do programa ORF FINDER (*Open Reading Frame Finder*) do NCBI, usando o código genético bacteriano. Efectuou-se o alinhamento das sequências proteicas pelo mesmo método que os alinhamentos dos genes referido no capítulo anterior e extraiu-se a sequência consenso de cada proteína. A partir das sequências consensos, avaliou-se a antigenicidade para epítomos B, definindo-se uma região ou regiões de interesse, ou seja, as mais antigénicas. Também através da sequência consenso, foram avaliadas as regiões mais antigénicas para epítomos Th. Definiu-se então as regiões mais antigénicas para epítomos B e simultaneamente mais antigénicas para epítomos Th. Seguidamente, estas regiões foram seleccionadas a partir do seu grau de conservação da proteína – portanto, as regiões mais antigénicas B e Th que fossem as mais conservadas, constituíram os fragmentos seleccionados. Por fim, a partir dos fragmentos seleccionados, avaliou-se a afinidade para ligação a moléculas MHC-I. Após definição das regiões conservadas e ricas em epítomos B e Th, avaliou-se a afinidade dos epítomos desses fragmentos para MHC-I (ou seja epítomos Tc).

Em relação à análise de conservação foi considerado quer a conservação em diversas estirpes da sequência em aminoácidos da proteína, quer a conservação na sequência nucleotídica do gene analisada no capítulo anterior. Ou seja, além da análise da conservação da proteína foram consideradas as regiões do gene, na tentativa de seleccionar fragmentos com ausência de substituições não-sinónimas e com o menor número de locais de substituições sinónimas possíveis.

IV.1. Análise da antigenicidade e conservação dos epítomos

A análise da antigenicidade e conservação foi realizada através das ferramentas do “IEDB Analysis Resource” disponibilizadas pelo *Immune Epitope Database* (IEDB).

Estimativa de epítomos B

Os epítomos B foram estimados por determinação da acessibilidade dos fragmentos à superfície da proteína pelo algoritmo de Emini (Emini *et al.*, 1985) através da ferramenta “Emini Surface Accessibility Prediction”, e pelo método de Bepipred (Larsen *et al.*, 2006) através da ferramenta “Bepipred Linear Epitope Prediction”.

Antigenicidade para epítomos T

Para estimar epítomos T consideraram-se métodos que avaliam o processamento do péptido, o seu transporte pelos respectivos transportadores e a afinidade de ligação às moléculas MHC classe I e classe II. Na aplicação dos métodos foi considerado uma dimensão de epítomos de 9 aminoácidos, dado ser esta a dimensão do local central de ligação tanto das moléculas MHC-I como MHC-II. É de referir que devido à conformação do centro de ligação das moléculas MHC, as moléculas MHC-I raramente se ligam a péptidos maiores que 9 aminoácidos, enquanto que as moléculas MHC-II podem acomodar péptidos de maiores dimensões, entre 13 e 15 aminoácidos (Wang *et al.*, 2008).

Os epítomos Th foram determinados pelo método *consensus* (Wang *et al.*, 2008), em que se usou o alelo MHC-II mais prevalente na Europa, de acordo com a base de dados dbMHC – HLA DRB1*0401. Os epítomos T citotóxicos foram estimados pelo método NetCTL (Larsen *et al.*, 2005, 2007), que considera a afinidade de ligação às moléculas MHC-I, clivagem pelo proteossoma e a eficiência do transporte pela TAP (*transporter associated with antigen processing*) (conf. Anexo II).

Conservação e especificidade dos epítomos

O grau de conservação dos fragmentos para cada proteína no conjunto das sequências usadas foi avaliado através da ferramenta “Epitope Conservancy Analysis” pelo método de análise “Epitope linear sequence conservancy”. É também relevante que os alvos antigénicos seleccionados sejam específicos da *H. pylori*, de forma a evitar desencadear respostas imunitárias contra microrganismos não patogénicos e que podem fazer parte da flora

microbiana normal do hospedeiro. A especificidade dos epítomos para *H. pylori* foi avaliada por análise de homologia com proteínas de todos os outros microrganismos, usando a base de dados “Microbial DataBase” através da ferramenta “genomic BLAST” do programa BLASTP do NCBI (*National Center for Biotechnology Information*). Também é relevante que não exista homologia entre as proteínas antigénicas seleccionadas e as proteínas humanas de forma a evitar o desencadear de respostas auto-imunes. Efectuou-se uma análise para cada um dos fragmentos antigénicos seleccionados pela ferramenta “BLAST Human Sequences” do NCBI por BLASTP com a base de dados “RefSeq protein”.

IV.2. Fragmentos alvo

Tome-se como exemplo a proteína HomB, em relação às metodologias utilizadas para seleccionar zonas peptídicas ricas em epítomos B, Th e Tc conservados: a partir da sequência consenso obtida por alinhamento das proteínas, efectuou-se os seguintes passos: análise pelos métodos de Emini e Bepipred a partir da qual se definiu uma região rica em epítomos B entre o aminoácido 150 e 500 relativamente à sequência consenso (figura IV.1.). Em relação a esta região, seleccionou-se dois fragmentos que apresentam maior grau de conservação e concomitantemente mais ricos em epítomos Th, isto é, com elevada afinidade para ligação a moléculas MHC-II. Com base nesta metodologia foram definidos os seguintes fragmentos: um de 21 aminoácidos WQNFTPQTAEFTNLMLNMIA e um segundo de 28 aminoácidos YANLGSQTNYNLNASLNTQDLANSMLST (figura IV.2.A.). Os fragmentos seleccionados foram submetidos a análise de antigenicidade para epítomos Tc, por pesquisa de todos os supertipos MHC-I disponíveis na base de dados (figura IV.2.B). Para cada fragmento, confirmou-se o seu grau de conservação no conjunto das 120 sequências através da ferramenta “Epitope Conservancy Analysis”. A especificidade para *H. pylori* foi confirmada através da realização de BLASTP contra outros microrganismos e contra o hospedeiro humano.

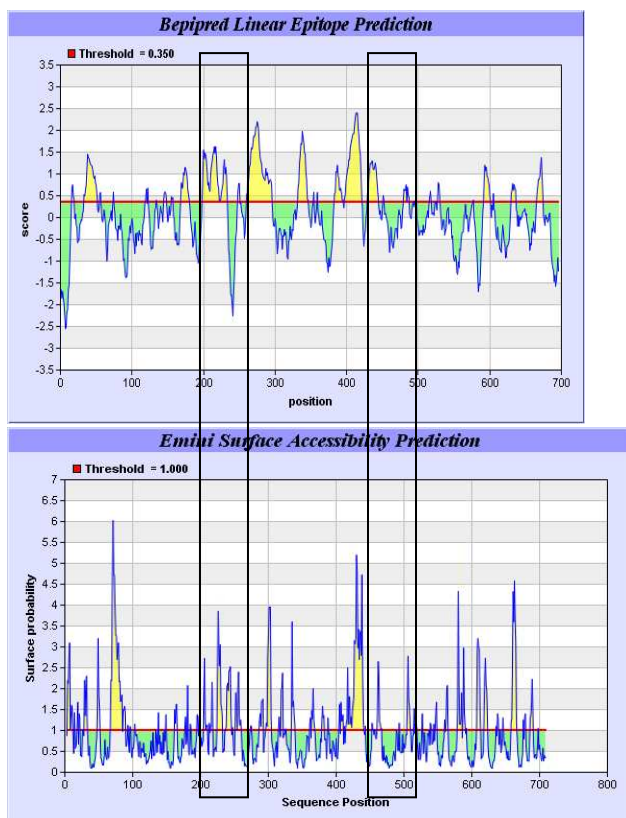


Figura IV.1. Análise pelo método de Emini e Bepipred da sequência consenso da proteína HomB, originada a partir de 120 sequências. As regiões englobadas pelos rectângulos destacam as regiões dos fragmentos antigénicos seleccionados, pois embora não sejam as regiões mais antigénicas para B da proteína são que concomitantemente apresentam epítomos B, Th e são o mais conservadas possível. No eixo das abcissas está representada a posição do aminoácido na sequência consenso. A linha a vermelho indica o limiar entre apresentar epítomos B (zona indicada a amarelo) e não apresentar (zona indicada a verde)

A

Fragmento antigénico WQNFTPQTAEFTNLMMLN												
Allele	Position	Sequence	ARB Core	ARB Score	ARB Percentile Rank	SMM_align Core	SMM_align Score	SMM_align Percentile Rank	Sturmiolo Core	Sturmiolo Score	Sturmiolo Percentile Rank	Consensus Percentile Rank
HLA DRB1*0401	1:1-15	WQNFTPQTAEFTNLM	WQNFTPQTA	232.11	32.82	FTPQTAEFF	947.00	10.21	WQNFTPQTA	0.80	5.77	10.21
HLA DRB1*0401	1:2-16	QNFTPQTAEFTNLM	TAEFTNLM	4241.77	68.52	FTPQTAEFF	1127.00	11.98	FTPQTAEFF	0.30	7.51	11.98
HLA DRB1*0401	1:3-17	NFTPQTAEFTNLM	TAEFTNLM	4241.77	68.52	TAEFTNLM	1999.00	19.39	FTPQTAEFF	0.30	7.51	19.39
HLA DRB1*0401	1:4-18	FTPQTAEFTNLM	TAEFTNLM	4241.77	68.52	TAEFTNLM	2024.00	19.58	FTPQTAEFF	0.30	7.51	19.58
HLA DRB1*0401	1:5-19	TPQTAEFTNLM	EFTNLM	2874.77	63.95	TAEFTNLM	3176.00	27.15	TPQTAEFF	-12.50	34.99	34.99
HLA DRB1*0401	1:6-20	PQTAEFTNLM	FTNLM	36.80	15.14	TAEFTNLM	1509.00	15.46	FTNLM	0.50	7.01	15.14
HLA DRB1*0401	1:7-21	QTAEFTNLM	NLM	2.77	3.20	FTNLM	914.00	9.88	FTNLM	0.50	7.01	7.01

Fragmento antigénico YANLGSQTNYNLNASLNTQDLANSMLST												
Allele	Position	Sequence	ARB Core	ARB Score	ARB Percentile Rank	SMM_align Core	SMM_align Score	SMM_align Percentile Rank	Sturmiolo Core	Sturmiolo Score	Sturmiolo Percentile Rank	Consensus Percentile Rank
HLA DRB1*0401	1:1-15	YANLGSQTNYNLNAS	YANLGSQTN	3.89	4.06	YANLGSQTN	268.00	2.79	YANLGSQTN	0.90	5.42	4.06
HLA DRB1*0401	1:2-16	ANLGSQTNYNLNAS	SQTNYNLNA	276.06	34.81	LGSQTNYNL	904.00	9.78	LGSQTNYNL	-0.40	11.01	11.01
HLA DRB1*0401	1:3-17	NLGSQTNYNLNAS	YNLNASLN	6.35	5.60	LGSQTNYNL	1683.00	16.92	LGSQTNYNL	-0.40	11.01	11.01
HLA DRB1*0401	1:4-18	LGSQTNYNLNASLNT	YNLNASLNT	0.16	0.28	LGSQTNYNL	1.6e+02	1.5	YNLNASLNT	3.2	0.62	0.62
HLA DRB1*0401	1:5-19	GSQTNYNLNASLNTQ	YNLNASLNT	0.16	0.28	YNLNASLNT	1.6e+02	1.5	YNLNASLNT	3.2	0.62	0.62
HLA DRB1*0401	1:6-20	SQTNYNLNASLNTQDL	YNLNASLNT	0.16	0.28	YNLNASLNT	1.6e+02	1.5	YNLNASLNT	3.2	0.62	0.62
HLA DRB1*0401	1:7-21	QTNYNLNASLNTQDL	YNLNASLNT	0.16	0.28	YNLNASLNT	1.5e+02	1.4	YNLNASLNT	3.2	0.62	0.62
HLA DRB1*0401	1:8-22	TNYLNASLNTQDL	YNLNASLNT	0.16	0.28	YNLNASLNT	1.5e+02	1.4	YNLNASLNT	3.2	0.62	0.62
HLA DRB1*0401	1:9-23	NYLNASLNTQDLANS	YNLNASLNT	0.16	0.28	YNLNASLNT	4.3e+02	4.7	YNLNASLNT	3.2	0.62	0.62
HLA DRB1*0401	1:10-24	YNLNASLNTQDLANS	YNLNASLNT	0.16	0.28	YNLNASLNT	4.4e+02	4.8	YNLNASLNT	3.2	0.62	0.62
HLA DRB1*0401	1:11-25	NLNASLNTQDLANS	LNASLNTQD	9.31	7.11	LNTQDLANS	1284.00	13.47	LNTQDLANS	0.20	8.36	8.36
HLA DRB1*0401	1:12-26	LNASLNTQDLANS	LNASLNTQD	9.31	7.11	NTQDLANS	1220.00	12.88	LNTQDLANS	0.20	8.36	8.36
HLA DRB1*0401	1:13-27	NASLNTQDLANS	QDLANS	12.23	8.36	NTQDLANS	942.00	10.17	LNTQDLANS	0.20	8.36	8.36
HLA DRB1*0401	1:14-28	ASLNTQDLANS	QDLANS	12.23	8.36	NTQDLANS	936.00	10.11	LNTQDLANS	0.20	8.36	8.36

B

Residue Number	Peptide Sequence	Predicted MHC Binding Affinity	Rescale Binding Affinity	C Terminal Cleavage Affinity	TAP Transport Efficiency	Prediction Score
4	FTPQTAEFF	0.6711	0.9182	0.1014	2.5440	1.0606

Afinidade do fragmento WQNFTPQTAEFTNLMLNMIA para o supertipo A24.

Residue Number	Peptide Sequence	Predicted MHC Binding Affinity	Rescale Binding Affinity	C Terminal Cleavage Affinity	TAP Transport Efficiency	Prediction Score
2	ANLGSQTNY	0.1649	0.4786	0.8609	2.8820	0.7519
17	NTQDLANSM	0.2392	0.6942	0.8445	0.3360	0.8377

Figura IV.2. Antigenicidade para epítomos T dos fragmentos antigénicos HomB seleccionados.

A. Análise da afinidade de ligação dos epítomos existentes no fragmento indicado ao complexo MHC-II. A posição indicada na tabela é referente à posição no fragmento peptídico do 1º aminoácido do epítomo. Na coluna "Sequence" encontra-se a sequência de 15 aa considerada para avaliação da afinidade dos epítomos de 9 aa às moléculas MHC-II. A análise de afinidade de ligação foi efectuada pelo método ARB – *Average Relative Binding method* (Bui *et al.*, 2005), pelo método SMM-align (Nielsen *et al.*, 2007) e pelo método de Sturniolo (Sturniolo *et al.*, 1999). Este último método foi usado como *raw score*, em que quanto maior o valor observado maior é a afinidade de ligação. Os métodos de ARB e SMM-align são dados em unidades de IC₅₀nM, pelo que quanto menor o valor, maior a afinidade. Regra geral, um IC₅₀nM <50 nM tem elevada afinidade, <500 nM afinidade intermédia e <5000 nM tem baixa afinidade. Por definição, um epítomo Th não tem um valor superior a 5000. O "Percentile Rank" constitui a comparação do valor de cada método, com os valores de 5 milhões de 15 aa ao acaso da base de dados SWISSPROT. Quanto menor este valor, maior é a afinidade. O *rank* do método *consensus* representa o valor médio de todos os métodos.

B. Análise da afinidade de ligação dos epítomos existentes no fragmento indicado ao complexo MHC-II, a afinidade de cada fragmento para um supertipo. "Residue number" indica a posição do primeiro aminoácido do epítomo ("Peptide Sequence") relativamente à sequência proteica introduzida.

O fragmento YANLGSQTNYNLNASLNTQDLANSMLST é mais antigénico para Th do que o fragmento WQNFTPQTAEFTNLMLNMIA, no entanto é menos conservado. No que toca à antigenicidade para epítomos Tc, são reconhecidos por 3 e 4 supertipos, respectivamente. Dado estas características de semelhança, a selecção de apenas um ou dos dois fragmentos, dependerá de testes futuros.

As metodologias descritas para a proteína HomB, e utilizadas para seleccionar fragmentos da proteína ricos em epítomos B, Th e Tc conservados, foram aplicadas aos restantes alvos. Com base nesta análise foram seleccionados diversos fragmentos tal como está apresentado na tabela IV.1.

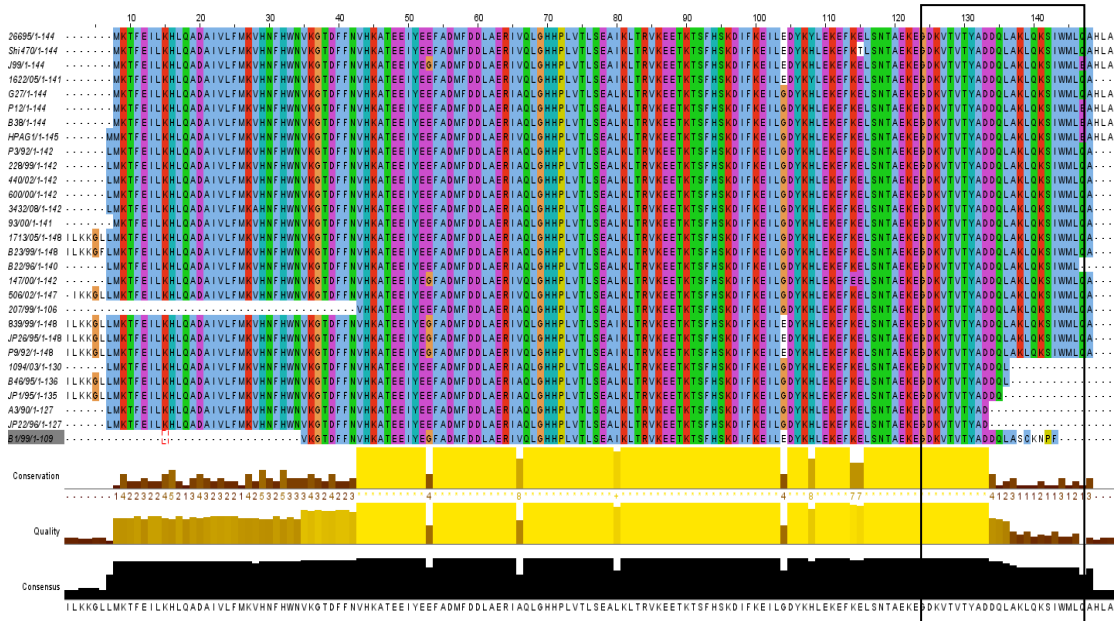
Em relação à proteína NapA, foi difícil obter regiões antigénicas em B e Th simultaneamente, como se pode observar na figura IV.3 e figura IV.4, uma vez que as regiões mais ricas em epítomos B apresentam menores valores de afinidade de ligação a moléculas MHC-II e vice-versa. Por isso, optou-se pela selecção de uma região com valores menos elevados de antigenicidade B mas que apresentassem boa antigenicidade em epítomos Th. Para as proteínas HpaA, HomB e VacA, observou-se que as regiões antigénicas eram as regiões mais variáveis – tal como observado na figura IV.5. Para a proteína HpaA conseguiu-se obter um fragmento o rico em epítomos B e Th conservados, no entanto, para as proteínas VacA e HomB, seleccionaram-se 3 e 2 fragmentos respectivamente, com menor percentagem de conservação. Visto que um bom imunogénico deve apresentar uma boa antigenicidade para células B e concomitantemente boa afinidade para moléculas MHC, a ausência destas

características em simultâneo, conduziu-nos à preferência de uma região menos conservada, mas mais imunogénica.

Fragmento	Conservação				MHC-II		MHC-I	
	100%	90%	≥80%	<80%	menor valor	maior valor		
NapA DKVTVTYADDQLAKLQKSIWML Dimensão:22aa Localização:118-139	79,3% (23/29)*				12,74	54,84	A2	
							A3	
							A24	
							B8	
							B39	
HpaA LSELDIQEKFLKTTSSHS Dimensão:19aa Localização:184-202	72% (18/25)*				1,64	53,94	A3	
							A24	
							B44	
							B62	
VacA TLYANSGAQRDLLQTLIDSHNAGYARTM Dimensão:30aa Localização:896-925	3,57% (2/56)	80,36% (45/56)	98,21% (55/56)	1,79% (1/56) 23,30%	6,66	50,21	A1	
							A26	
							B8	
							B39	
							B62	
ANQHEFDFEAQGALGSDQSSLNFKSALLRD Dimensão:30aa Localização:1094-1123	10,71% (6/56)	98,21% (55/56)	98,21% (55/56)	1,79% (1/56) 20%	1,94	51,77	A1	
							A24	
							B39	
							B58	
	LNQSYNYLAYSAAATRASGYDFAFFR Dimensão: 26aa Localização:1124-1149	12,50% (7/56)	85,71% (48/56)	85,71% (48/56)	1,79% (1/56) 23,08%	2,16	70,73	A1
								A3
								A24
								A26
								B27
HomB WQNFTPQTAEFTNLMLNMIA Tamanho:21aa Localização:216-236	75% (90/120)	100%			5,42	34,99	A24	
							A26	
							B44	
							B62	
YANLGSQTNYNLNASLNTQDLANSMLST Tamanho:28aa Localização:455-482	12,50% (15/120)	72,50% (87/120)	99,17% (119/120)	0,83% (1/120) 67,86%	0,62	11,01	A26	
							B39	
							B62	

Tabela IV.1. Características dos fragmentos seleccionados para cada alvo antigénico. A localização dos fragmentos indicada é referente à posição que ocupam na sequência consenso. O grau de conservação foi obtido através da ferramenta “Epitope Conservancy Analysis”. Os fragmentos seleccionados para as proteínas NapA e HpaA são 100% conservados. O baixo grau de conservação aparente registado deve-se ao facto de apenas 23 e 18 das sequências é que têm o comprimento total da proteína, respectivamente. A coluna designada por “MHC-II” representa os valores mínimos e máximos relativos ao método “Consensus Percentile Rank” que estima a afinidade de ligação dos epítomos às moléculas MHC-II – ou seja, de todos os epítomos que o fragmento da proteína HpaA pode gerar, estima-se que o mais antigénico apresente um valor de 1,64 e o menos antigénico 53,9. A coluna designada por “MHC-I” representa os supertipos para os quais os fragmentos são reconhecidos.

A.



B.

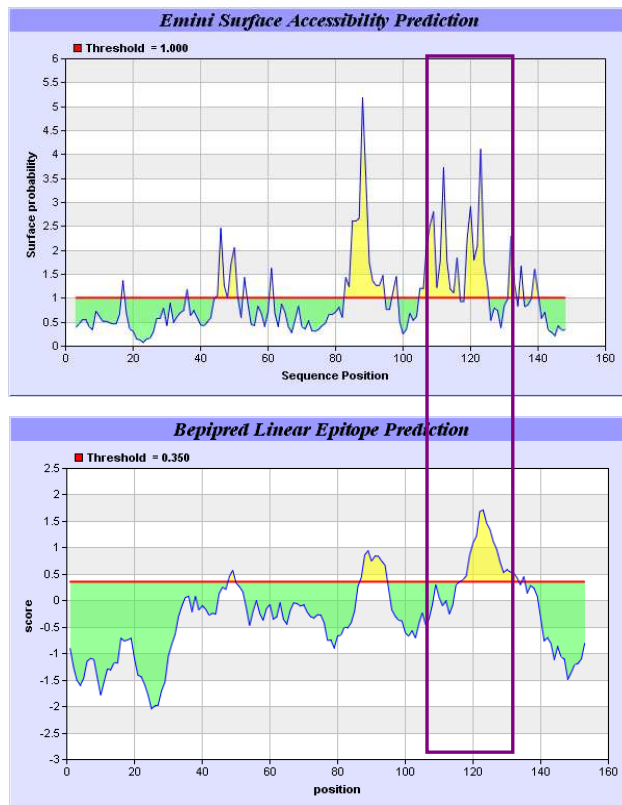


Figura IV.3. Alinhamento das proteínas e análise de antigenicidade de epítopos B para a proteína NapA.

A. Alinhamento de 29 seqüências proteicas de NapA, com indicação da seqüência consenso e da sua conservação entre as estirpes consideradas. O rectângulo assinala a região do fragmento escolhido.

B. Estimativa de epítopos B pelo método de Emini e de Bepipred. A região entre os 120 e os 140 aa assinalada com um rectângulo é das regiões mais antigénicas para epítopos B. No entanto, foi seleccionada a região entre 125-146aa que é a mais conservada e menos antigénica para B em relação ao intervalo referido anteriormente, dado também apresentar mais epítopos Th (figura IV.4).

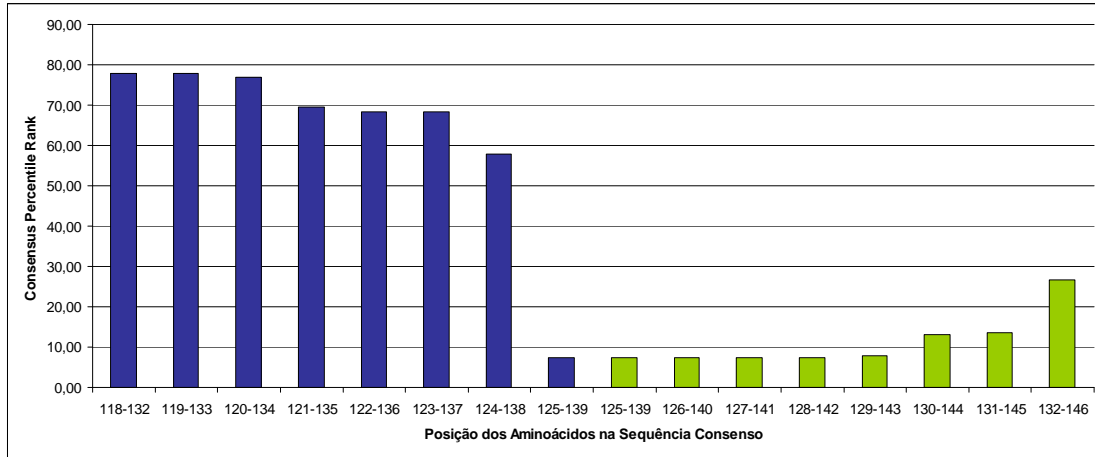


Figura IV.4. Comparação entre duas regiões da proteína NapA quanto à afinidade de ligação a moléculas MHC-II. Nas colunas azuis apresentam-se fragmentos de 15 aa localizados entre o aminoácido 118 e o 139 da sequência consenso e nas colunas verdes fragmentos de 15 aa da posição 125-146 que correspondem ao fragmento DKVTVTYADDQLAKLQKSIWML seleccionado para construção da vacina. Tal como se pode comparar pela figura IV.2.A, estes fragmentos de 15 aa têm o potencial de gerar epítomos de 9 aa que são avaliados quanto à sua antigenicidade. Como se pode observar a região entre os aminoácidos 118-125 aa, muito rica em epítomos B, é pouco rica em epítomos Th (quanto menor o valor, maior a antigenicidade). Em contrapartida, a região seleccionada é menos rica em epítomos B é muito mais rica em epítomos Th.

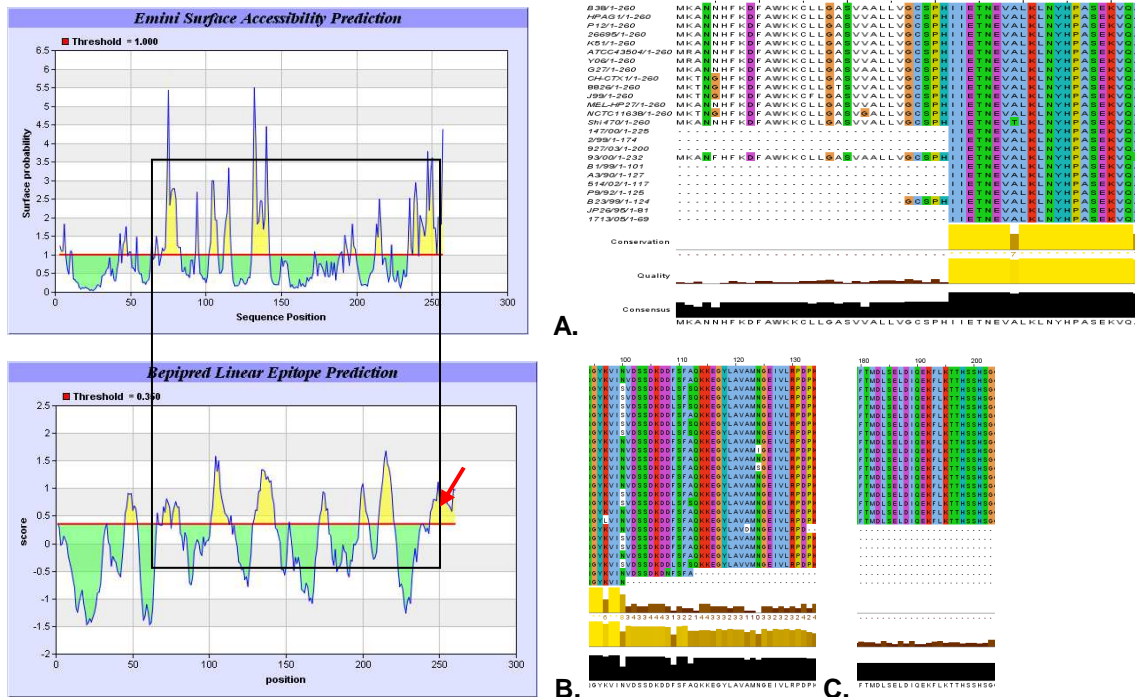


Figura IV.5. Análises efectuadas para a proteína Hpa em relação à conservação entre estipes e estimativa de epítomos B. A região mais rica em epítomos B encontra-se assinalada pelo rectângulo nos gráficos da análise de Emini (gráfico superior) e Bepired (Gráfico inferior). Entre a posição 1-50aa a antigenicidade é bastante baixa mas a região é bastante conservada. Observa-se também que relativamente à presença dos resíduos à superfície da proteína existe uma zona no início que embora não seja antigénica para B de acordo com o método Bepired, encontra-se acessível ao reconhecimento pelo sistema imunitário. Pode-se constatar que esta zona corresponde à região entre os aa 5-25 do alinhamento, que apresentam dois locais variáveis. A região entre os 100-120aa é das mais antigénicas mas pouco conservada. O fragmento final escolhido para ser integrado na construção da vacina localiza-se entre 185-202aa (região assinalada pela seta no gráfico Bepired) e é 100% conservado tal como determinado pelo “Epitope Conservancy Analysis” (para o número de sequências que temos)

A afinidade de ligação a moléculas MHC-I foi considerada como secundária. Estas moléculas ligam-se a proteínas que se localizam no interior da célula humana, implicando a apresentação dos epítomos a células Tc que levariam à destruição da célula humana infectada, típicas por exemplo em infecções virais. As moléculas MHC-II apresentam normalmente antígenos exteriores à célula e que são incorporados, processados e apresentados às células Th (T *helper*). Este tipo de resposta irá estimular e amplificar a resposta humoral por células B e irá conduzir à formação de células memória.

Todos os fragmentos seleccionados são específicos das proteínas alvo, excepto HomB, que apresentou homologia com outras OMP da *H. pylori*, nomeadamente HomA. Isto deverá prender-se com a elevada homologia que as OMP apresentam entre si. Os alvos NapA e HpaA apresentaram ainda alguma homologia com as proteínas Nap e Hpa quer de *H. acinonichys*, como de *H. hepaticus*.

Para as proteínas VacA e HomB não foi possível a selecção de fragmentos antigénicos 100% conservados em todas as estirpes de estudo. A eficácia da vacina, entre outros factores, passa também pela capacidade de cobrir toda a variabilidade do patogéneo.

Capítulo V. Conclusões e Perspectivas

Neste trabalho pretendeu-se definir um conjunto de fragmentos conservados e antigénicos que possam ser utilizados no desenho de uma vacina de DNA com base em dados de imunoproteómica e da relevância dos alvos antigénicos no processo de infecção, considerou-se para o desenho de uma vacina de DNA as seguintes cinco proteínas: NapA, HpaA, VacA, HomB e Omp9. Dada a ausência de informação quanto à variabilidade genética dos alvos NapA, HpaA e Omp9, foi necessário proceder à sua sequenciação em diversas estirpes de *H. pylori* associadas a patologias gástricas distintas, previamente amplificadas por PCR e PCR *touchdown*,

Observou-se que uma grande diversidade genética e elevado número de alelos nos genes *napA*, *hapA*, *vacA* e *homB*, e que estes genes se encontram sob uma selecção de novas variantes antigénicas, dado o elevado número de substituições não sinónimas. Visto que existem muitas locais de substituições sinónimas num número relativamente grande de estirpes por gene e dado o processo de selecção em curso, especulou-se que é provável que estes locais dêem origem a substituições não sinónimas. Estes dados são fulcrais para a construções de uma vacina, de forma a cobrir-se a variabilidade antigénica do patogéneo. Nesta perspectiva a vacina deve ser constituída por diversos fragmentos de diversos alvos antigénicos. Cada fragmento deve conter multi-epítomos conservados e antigénicos de modo a que, caso ocorram mutações que alterem o epítopo codificado, a probabilidade de estas ocorrerem em 3 ou 5 epítomos simultaneamente é mais reduzida. Uma vez que as regiões mais antigénicas são geralmente menos conservadas, optou-se pela selecção de fragmentos antigénicos ainda que não fossem tão conservados como seria preferível.

O próximo passo do trabalho consiste sequenciação do gene *omp9*, seguido da construção da vacina de DNA constituída por fragmentos conservados e antigénicos dos alvos NapA, HpaA, VacA, HomB e Omp9.

Capítulo VI. Bibliografía

Achtman M. & Suerbaum S. (2000). Sequence variation in *Helicobacter pylori*. Elsevier Science Ltd. Comment

Alm R.A. et al. (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397: 176-180.

Alm R.A., Bina J., Andrews B.M., Doig P., Hancock R.E.W., Trust T.J. (2000). Comparative Genomics of *Helicobacter pylori*: Analysis of the Outer Membrane Protein Families. *Infection and Immunity* 68(7): 4155–4168.

Baltrus D.A., Amieva M.R., Covacci A., Lowe T.M., Merrell D.S., Ottemann K.M., Stein M., Salama N.R., Guillemin K. (2009). The complete genome sequence of *Helicobacter pylori* strain G27. *J Bacteriol* 191: 447-448.

Bui H.H., Sidney J., Peters B., Sathiamurthy M., Sinichi A., Purton K.A., Mothé B.R., Chisari F.V., Watkins D.I., Sette A. (2005). *Immunogenetics*. 57:304-314.

Carlsohn E., Nystrom J., Bolin I., Nilsson C.L., Svennerholm A.-M. (2006). HpaA Is Essential for *Helicobacter pylori* Colonization in Mice. *Infection and Immunity* 74(2):920-926.

Carlsohn E., Nystrom J., Karlsson H., Svennerholm A.-M., Nilsson C.L. (2006). Characterization of the Outer Membrane Protein Profile from Disease-Related *Helicobacter pylori* Isolates by Subcellular Fractionation and Nano-LC FT-ICR MS Analysis. *Journal of Proteome Research* 5:3197-3204.

Clamp M., Cuff, J., Searle S.M., Barton, G.J. (2004). The Jalview Java Alignment Editor. *Bioinformatics* 20;426-7.

Del Giudice G., Covacci A., Telford J.L., Montecucco C., Rappuoli R. (2001). The Design of Vaccines against *Helicobacter pylori* and their development. *Annu. Rev. Immunol.* 19:523–63.

Del Giudice G., Malfertheiner P., Rappuoli R. (2009). Development of vaccines against *Helicobacter pylori*. *Expert Rev Vaccines* 8(8):1037-1049.

Dong Q.-J., Wang Q., Xin Y.-N., Li N., Xuan S.-Y. (2009). Comparative genomics of *Helicobacter pylori*. *World J Gastroenterol* 15(32): 3984-3991.

Don R.H., Cox P.T., Wainwright, B.J. Baker K. and Mattick J.S. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research*, 19(14).

Dong Q.-J., Wang Q., Xin Y.-N., Li N., Xuan S.-Y. (2009). Comparative genomics of *Helicobacter pylori*. *World J Gastroenterol* 15(32): 3984-3991.

Dubois A. & Borén T. (2007). *Helicobacter pylori* is invasive and it may be a facultative intracellular organism. *Cell Microbiol.* 9(5): 1108–1116.

Emini E.A., Hughes J.V., Perlow D.S., Boger J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 55:836-839.

Falush D., Kraft C., Taylor N.S., Correa P., Fox J.G., Achtman M., Suerbaum S. (2001). Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: Estimates of clock rates, recombination size, and minimal age. *PNAS* 98(26): 15056–15061.

Falush D. et al. (2003). Traces of Human Migrations in *Helicobacter pylori* Populations. *Science* 299, 1582-1585.

Felsenstein J (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783-791.

- Fu Y.-X. & Li W.-H. (1993). Statistical tests of neutrality of mutations. *Genetics* 133: 693-709.
- Ghiara P., Rossi M., Marchetti M., Tommaso A., Vindigni C., Ciampolini F., Covacci A., Telford J.L., Magistris M.T., Pizza M., Rappuoli R., Del Giudice G. (1997). Therapeutic Intra-gastric Vaccination against *Helicobacter pylori* in Mice Eradicates an Otherwise Chronic Infection and Confers Protection against Reinfection. *Infection and Immunity* 65(12):4996-5002.
- Go M.F., Kapur V., Graham D.Y., Musser J.M. (1996). Population Genetic Analysis of *Helicobacter pylori* by Multilocus Enzyme Electrophoresis: Extensive Allelic Diversity and Recombinational Population Structure. *J. Bacteriol.* 178(13), 3934–3938.
- Golubchik T., Wise M.J., Eastal S., Jermin L.S. (2007). Mind the Gaps: Evidence of Bias in Estimates of Multiple Sequence Alignments. *Mol. Biol. Evol.* 24(11):2433–2442.
- Haas G. et al. (2002). Immunoproteomics of *Helicobacter pylori* infection and relation to gastric disease *Proteomics* 2, 313–324.
- Hudson, R. R. (1987). Estimating the recombination parameter of a finite population model without selection. *Genet. Res.* 50: 245-250.
- Jungblut P. R., Bumann D., Haas G., Zimny-Arndt U., Holland P., Lamer S., Siejak F., Aebischer A., Meyer T.F. (2000). Comparative proteome analysis of *Helicobacter pylori* *Molecular Microbiology* 36(3), 710-725.
- Kansau I., Raymond J., Bingen E., Courcoux P., Kalach N., Bergeret M., Braimi N., Dupont C., Labigne A. (1996). Genotyping of *Helicobacter pylori* isolates by sequencing of PCR products and comparison with the RAPD technique. *Res. Microbiol.* 147, 661-669.
- Khoder G., Yamaoka Y., Fauchère J.-L., Burucoa C., Atanassov C. (2009). Proteomic *Helicobacter pylori* biomarkers discriminating between duodenal ulcer and gastric cancer. *Journal of Chromatography B* 877:1193–1199.
- Kimmel B., Bosserhoff A., Frank R., Gross R., Goebel W., Beier D. (2000). Identification of Immunodominant Antigens from *Helicobacter pylori* and Evaluation of Their Reactivities with Sera from Patients with Different Gastrointestinal Pathologies. *Infection and Immunity* 68(2):915–920.
- Kuipers E.J., Israel D.A., Kusters J.G., Gerrits M.M., Weel J., van der Ende A., van der Hulst R.W.M., Wirth H.P., Niskanen J.H., Thompson S.A., Blaser M.J. (2000). Quasispecies Development of *Helicobacter pylori* Observed in Paired Isolates Obtained Years Apart from the Same Host. *The Journal of Infectious Diseases* 181:273–82.
- Kusters J.G., J.G. van Vliet J.G., Kuipers E.J. (2006). Pathogenesis of *Helicobacter pylori* Infection. *Clinical Microbiology Reviews* 19 (3): 449-490.
- Larsen M.V., Lundegaard C., Kasper Lamberth, Buus S., Brunak S., Lund O., Nielsen M. (2005). An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC-I binding, TAP transport efficiency, and proteasomal cleavage predictions. *European Journal of Immunology* 35(8):2295-303.
- Larsen J.E., Lund O., Nielsen M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome Res* 2:2.
- Larsen M.V., Lundegaard C., Lamberth K., Buus S., Lund O., and Nielsen M. (2007). Large-Scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 8:424.
- Librado P., & Rozas J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.

Lock R.A., Cordwell S.J., Coombs G.W., Walsh B.J., Forbes G.M. (2001). Proteome Analysis of *Helicobacter pylori*: major proteins of type strain NCTC 11637. *Pathology* 33:365–374.

Malfertheiner P., Megraud F., O`Morain C., Bazzoli F., El-Omar E., Graham D., Hunt R., Rokkas T., Vakil N., Kuipers E. (2007). Current concepts in the management of *Helicobacter pylori* infection – The Maastricht III Consensus Report. *Gut* doi:10.1136/gut.2006.101634.

Malfertheiner P., Schultze V., Rosenkranz B., Kufmann S. H.E., Ulrichs T., Novicki D., Norelli F., Contorni M., Peppoloni S., Berti D., Tornese D., Ganju J., Palla E., Rappuoli R., Scharschmidt B.F., Del Giudice G. (2008). Safety and Immunogenicity of an Intramuscular *Helicobacter pylori* Vaccine in Noninfected Volunteers: A Phase I Study. *Gastroenterology* 135(3):787-795.

Marshall B.J., Armstrong J.A., McGeachie D.B., Glancy R.J. (1985). Attempt to fulfil Koch's postulates for pyloric campylobacter. *The Medical Journal of Australia* 142:436-439.

Marshall, B. J., Warren J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* i:1311–1315.

McAtee C.P., Lim M.Y., Fung K., Velligan M., Fry K., Chow T., Berg D.E. (1998). Identification of Potential Diagnostic and Vaccine Candidates of *Helicobacter pylori* by Two-Dimensional Gel Electrophoresis, Sequence Analysis, and Serum Profiling. *Clinical and Diagnostic Laboratory Immunology* 5(4): 537–542.

Nei M. (1987). *Molecular Evolutionary Genetics*. Columbia Univ. Press, New York.

Nielsen M., Lundegaard C., Lund O. (2007). Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics*. 8:238.

Nyström J. & Svennerholm A.M. (2007). Oral immunization with HpaA affords therapeutic protective immunity against *H. pylori* that is reflected by specific mucosal immune responses. *Vaccine*. 25(14):2591-8.

Oh J.D., Kling-Backhed H., Giannakis M., Xu J., Fulton R.S., Fulton L.A., Cordum H.S., Wang C., Elliott G., Edwards J., Mardis E.R., Engstrand L.G., Gordon J.I. (2006). The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. *Proc Natl Acad Sci USA* 103: 9999-10004.

Oleastro M., Monteiro L., Lehours P., Mégraud F., Ménard A. (2006). Identification of Markers for *Helicobacter pylori* Strains Isolated from Children with Peptic Ulcer Disease by Suppressive Subtractive Hybridization. *Infection and Immunity* 74(7): 4064–4074.

Oleastro M., Cordeiro R., Ferrand J., Nunes B., Lehours P., Carvalho-Oliveira I., Mendes A.I., Penque D., Monteiro L., Mégraud F., Ménard A. (2008). Evaluation of the Clinical Significance of homB, a Novel Candidate Marker of *Helicobacter pylori* Strains Associated with Peptic Ulcer Disease. *JID* 198.

Oleastro M., Cordeiro R., Yamaoka Y., Queiroz D., Mégraud F., Monteiro L., Ménard A. (2009). Disease association with two *Helicobacter pylori* duplicate outer membrane protein genes, homB and homA. *Gut Pathogens* 2009, 1:12.

Pan Z.-J., Van Der Hulst R.W.M., Feller M., Xiao S.-D., Tytgat G.N.J., Dankert J., Van Der Ende A. (1997). Equally High Prevalences of Infection with cagA-Positive *Helicobacter pylori* in Chinese Patients with Peptic Ulcer Disease and Those with Chronic Gastritis-Associated Dyspepsia. *Journal of Clinical Microbiology* 35(6):1344-1347.

Rossi G., Ruggiero P., Peppoloni S., Pancotto L., Fortuna D., Lauretti L., Volpini G., Mancianti S., Corazza M., Taccini E., Di Pisa F., Rappuoli R., Del Giudice G. (2004). Therapeutic Vaccination against *Helicobacter pylori* in the Beagle Dog Experimental Model: Safety, Immunogenicity, and Efficacy. *Infection and Immunity* 72(6):3252-3259.

Sabarth N., Lamer S., Zimny-Arndt U., Jungblut P.R., Meyer T.F., Bumann D. (2002). Identification of Surface Proteins of *Helicobacter pylori* by Selective Biotinylation, Affinity Purification, and Two-dimensional Gel Electrophoresis. *The Journal of Biological Chemistry* 277(31):27896–27902.

Saitou N & Nei M (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.

Sturniolo T., Bono E., Ding J., Radrizzani L., Tuereci O, Sahin U., Braxenthaler M., Gallazzi F., Protti M.P., Sinigaglia F., Hammer J. (1999). Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol.* 17(6):555-561.

Suerbaum S. & Achtman M. (1999). Evolution of *Helicobacter pylori*: the role of recombination. *Trends in Microbiology* 7(5): 182-184.

Suerbaum S., Smith J.M., Bapumia K., Morelli G., Smith N.H., Kunstmann E., Dyrek I., Achtman M. (1998). Free recombination within *Helicobacter pylori* (nucleotide sequencing/horizontal genetic exchange/evolution/linkage equilibrium). *Proc. Natl. Acad. Sci. USA* 95:12619–12624.

Tajima F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Tamura K, Dudley J, Nei M., Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24:1596-1599.

Tamura K, Nei M., Kumar S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. USA* 101:11030-11035.

Tomb J.F. et al. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388: 539-547.

Voland P., Hafsi N., Zeitner M., Laforsch S., Wagner H., Prinz C. (2003). Antigenic Properties of HpaA and Omp18, Two Outer Membrane Proteins of *Helicobacter pylori*. *Infection and Immunity* 71(7):3837–3843.

Wang G., Humayun M. Z., Taylor D.E. (1999). Mutation as an origin of genetic variability in *Helicobacter pylori*. *Trends in Microbiology* 7(12):488-493.

Wang P., Sidney J., Dow C., Mothé B., Sette A., et al. (2008) A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. *PLoS Comput Biol* 4(4): e1000048. doi:10.1371/journal.pcbi.1000048

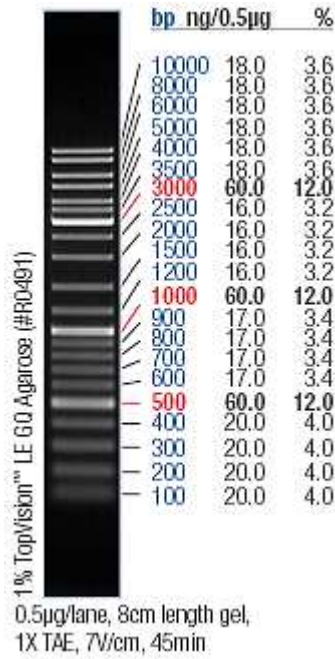
Watterson G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theor. Pop. Biol.* 7: 256-276.

Yamazaki S., Yamakawa A., Okuda T., Ohtani M., Suto H., Ito Y., Yamazaki Y., Keida Y., Higashi H., Hatakeyama M., Azuma T. (2005). Distinct Diversity of *vacA*, *cagA*, and *cagE* Genes of *Helicobacter pylori* Associated with Peptic Ulcer in Japan. *Journal of Clinical Microbiology* 43(8):3906–3916.

Zhang H. et al., Oral immunogenicity and protective efficacy in mice of a carrot-derived vaccine candidate expressing *UreB* subunit against *Helicobacter pylori*, *Protein Expr. Purif.* (2009), doi:10.1016/j.pep.2009.07.016.

Zhou W.-Y., et al. Therapeutic efficacy of a multi-epitope vaccine against *Helicobacter pylori* infection in BALB/c mice model. *Vaccine* (2009), doi:10.1016/j.vaccine.2009.05.009.

ANEXOS



Anexo I. Marcador de pesos moleculares *GeneRuler™ 1 kb Plus DNA Ladder, ready-to-use* (Fermentas, Vilnius, Lituânia).

Table 3: Representative alleles

Supertype	NetCTL	Epijen	MAPP	MHC-pathway	WAPP
A1	HLA-A*0101	HLA-A*0101	HLA-A1	HLA-A*0101	HLA-A*01
A2	HLA-A*0201	HLA-A*0201	HLA-A*0201	HLA-A*0201	HLA-A*0201
A3	HLA-A*0301	HLA-A*0301	HLA-A3	HLA-A*0301	HLA-A*03
A24	HLA-A*2402	HLA-A*24	HLA-A24	HLA-A*2402	N/A
A26	HLA-A*2601	N/A	N/A*	HLA-A*2601	N/A
B7	HLA-B*0702	HLA-B*07	HLA-B7	HLA-B*0702	N/A
B8	HLA-B*0801	N/A	HLA-B8	HLA-B*0801	N/A
B27	HLA-B*2705	HLA-B*27	HLA-B*2705	HLA-B*2705	HLA-B*2705
B39	HLA-B*3901	N/A	HLA-B*3901	N/A	N/A
B44	HLA-B*4001	HLA-B*40	HLA-B40	HLA-B*4002	N/A
B58	HLA-B*5801	N/A	HLA-B*5801	HLA-B*5801	N/A
B62	HLA-B*1501	N/A	HLA-B62	HLA-B*1501	N/A
# epitope-protein pairs	216	188	214	215	131

Anexo II. Tabela dos alelos MHC-I incluídos nas análises de antigenicidade (Larsen *et al.*, 2007).