



**HAL**  
open science

# Helicobacter pylori : migrations humaines et cancer gastrique

Sébastien Breurec

► **To cite this version:**

Sébastien Breurec. Helicobacter pylori : migrations humaines et cancer gastrique. Sciences agricoles. Université Paris Sud - Paris XI, 2011. Français. NNT : 2011PA114830 . tel-00649643v2

**HAL Id: tel-00649643**

**<https://theses.hal.science/tel-00649643v2>**

Submitted on 10 Jan 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# UNIVERSITÉ PARIS-SUD 11

**ECOLE DOCTORALE :**

INNOVATION THÉRAPEUTIQUE : DU FONDAMENTAL A L'APPLIQUÉ

*PÔLE : MICROBIOLOGIE / THERAPEUTIQUES ANTIINFECTIEUSES*

**DISCIPLINE :**  
MICROBIOLOGIE

ANNÉE 2011 – 2012

SÉRIE DOCTORAT N° 1142

**THÈSE DE DOCTORAT**

Soutenu le 17/11/2011

par

**Sébastien BREUREC**

*Helicobacter pylori* : migrations humaines et cancer gastrique

<b>Président :</b>	Anne COLLIGNON	Professeur, Université Paris Sud 11
<b>Directeur de thèse :</b>	Josette RAYMOND	Maître de conférences, Université Paris Descartes
<b>Rapporteurs :</b>	Christophe BURUCOA Catherine DAUGA	Professeur, Université de Poitiers Chargée de recherche, Institut Pasteur, Paris
<b>Examineurs :</b>	Sylvain BRISSE Hilde DE REUSE Jean-Louis SARTHOU	Chargé de recherche, Institut Pasteur, Paris Chef d'Unité, Institut Pasteur, Paris Directeur, Institut Pasteur, Dakar

## Remerciements

Je souhaiterais, tout d'abord, exprimer ma profonde reconnaissance à Josette Raymond qui a dirigé cette thèse. Sa rigueur scientifique, son enthousiasme et sa chaleur humaine resteront toujours pour moi des modèles.

Je tiens à remercier Benoit Garin, Michel Huerre et Agnès Labigne. Tous les trois sont à l'origine de ce travail. Je les remercie pour la confiance qu'ils m'ont accordée et pour tout ce qu'ils m'ont apporté tant professionnellement que humainement. Le travail présenté dans ce manuscrit est aussi le leur.

Une thèse ne se fait jamais seule et je remercie tous ceux qui m'auront épaulé d'une façon ou d'une autre. J'ai une pensée toute particulière pour Jean-Michel Thiberge (Institut Pasteur de Paris), Yoshan Moodley, Mark Achtman et Bodo Linz (Institut Max Planck à Berlin), Sopheak Hem et Bertrand Guillard (Institut Pasteur du Cambodge), et mes anciens et actuels collègues de l'Institut Pasteur de Dakar, plus particulièrement, Fatou Bintou Dieye, Cheikh Fall, Amadou Sall, Abdoulaye Seck et Jean-Marie Sire.

Je remercie également la Direction Internationale de l'Institut Pasteur de Paris, plus particulièrement Marc Jouan, pour m'avoir permis de concilier mes activités de recherche et de services.

Je remercie également tous mes proches, plus particulièrement Hildegarde, Martine et Aminata, qui ont traqué les fautes d'orthographe, et Hélène pour l'impression du manuscrit.

Enfin, je tiens à remercier très sincèrement les membres du Jury, qui ont accepté de juger ce travail en dépit de leurs contraintes respectives: Anne Collignon, pour avoir présidé ce jury, Christophe Burucoa et Catherine Dauga, pour avoir rapporté cette thèse, Sylvain Brisse, Hilde de Reuse et Jean-Louis Sarthou pour avoir examiné mon manuscrit.

Un immense Merci à Tous...

## Résumé

*Helicobacter pylori* est associée à des pathologies gastro-duodénales sévères mais est également un marqueur bactérien de migrations humaines. Nous avons montré que des populations génétiques distinctes de *H. pylori* ont accompagné au moins quatre migrations en Asie du sud-est et en Océanie : i) une expansion des ancêtres des austronésiens il y a 5000 ans à partir de Taiwan en Océanie, ii) une migration d'Inde en Asie du sud-est depuis 2000 ans, iii) une migration des ancêtres des locuteurs des langues austro-asiatiques au Vietnam et au Cambodge il y a 4000 ans, i) une migration des ancêtres des Thaïs du sud de la Chine vers l'actuelle Thaïlande au début du second millénaire. Ces données confirment la résolution plus élevée de la diversité génétique de *H. pylori* pour retracer les anciennes migrations humaines par comparaison aux marqueurs génétiques humains traditionnels. Nous avons ensuite investigué les facteurs de virulence de souches isolées de patients présentant des symptômes gastriques au Sénégal et au Cambodge. Au Sénégal, une association significative a été observée entre le cancer gastrique et le gène *cagA*, deux motifs EPIYA-C et l'allèle *vacA* s1. De multiples segments EPIYA-C étaient observés moins fréquemment que dans les autres régions du monde, contribuant probablement à la faible incidence du cancer gastrique. Au Cambodge, une introgression fréquente d'allèles *cagA* et *vacA* européens dans des souches d'Asie de l'est a été observée. CagA et VacA ayant des effets antagonistes, cette expansion pourrait entraîner la rupture de l'équilibre entre les effets biologiques de ces deux protéines et être responsable de conséquences graves sur l'évolution de la maladie.

## **Title**

*Helicobacter pylori*, human migrations and cancer gastric

## **Summary**

*Helicobacter pylori* is associated with severe gastroduodenal disorders but is also a bacterial genetic marker of human migrations. First, we provide evidence that distinct *H. pylori* genetic populations accompanied at least four ancient human migrations into Oceania and Southeast Asia: i) an expansion of Austronesian speaking people about 5000 years ago from Taiwan into Oceania, ii) a migration from India into Southeast Asia within the last 2000 years, iii) a migration of Austro-Asiatic speaking people into Vietnam and Cambodia about 4000 years ago, and iv) a migration of the ancestors of the Thai people from Southern China into Thailand during the early second millennium AD. These findings demonstrate that *H. pylori* genetic diversity has more discriminatory power than traditional human genetic markers for tracing old human migrations. Second, we investigated virulence factors of *H. pylori* strains isolated from patients with gastric symptoms in Senegal and Cambodia. In Senegal, a significant association was observed between gastric cancer and the *cagA* gene, two EPIYA-C segments, and the s1 *vacA* allele in Senegal. Multiple EPIYA-C segments were less frequent than reported in other countries, possibly contributing to the low incidence of gastric cancer. In Cambodia, the frequent introgression of European *vacA* and *cagA* alleles into East Asian *H. pylori* strains was observed. As VacA and CagA have opposing effects, this expansion may disrupt the balancing activities on epithelial cells which might result in severe consequences for individual disease outcome.

## **Mots clefs**

*Helicobacter pylori*, population génétique, migrations humaines, Taiwan, Océanie, Inde, Asie du sud-est, cancer gastrique, *cagA*, EPIYA, *vacA*, introgression, Sénégal, Cambodge

## **Key words**

*Helicobacter pylori*, genetic population, human migrations, Taiwan, Oceania, India, Southeast Asia, gastric cancer, *cagA*, EPIYA, *vacA*, introgression, Senegal, Cambodia

**Laboratoire de rattachement**

Institut Pasteur, Unité de Pathogenèse de *Helicobacter*, 28 rue du Docteur Roux, 75724 Paris  
Cedex 15, France

**Laboratoire associé**

Institut Pasteur, Unité de Bactériologie Médicale et Environnementale, 36 avenue Pasteur,  
Dakar, Sénégal

## TABLE DES MATIERES

LISTE DES FIGURES.....	9
LISTE DES TABLEAUX.....	10
LISTE DES ABREVIATIONS.....	11
Partie I. Généralités.....	13
1. Historique et taxonomie.....	14
2. Pathologies associées aux infections à <i>H. pylori</i> .....	16
2.1. Pathologies associées aux infections à <i>H. pylori</i> .....	16
2.2. Indications de la recherche et du traitement de l'infection à <i>H. pylori</i> .....	19
2.2.1. Ulcère gastrique et duodéal.....	19
2.2.2. Lymphome du MALT.....	19
2.2.3. Cancer gastrique et lésions pré-néoplasiques.....	19
3. Epidémiologie de l'infection.....	21
3.1. Prévalence, incidence et réinfection.....	21
3.2. Réservoir de virus.....	22
3.2.1. Arguments en faveur de l'eau.....	23
3.2.2. Arguments en faveur des aliments.....	24
3.2.3. Arguments en faveur du tractus digestif des animaux.....	24
3.3. Transmission de l'infection par <i>H. pylori</i> .....	25
3.3.1. Voies de transmission de l'infection.....	25
3.3.1.1. Arguments en faveur de la transmission gastro-orale.....	25
3.3.1.2. Arguments en faveur de la transmission oro-orale.....	25
3.3.1.3. Arguments en faveur de la transmission féco-orale.....	26
3.3.2. Mode de transmission de l'infection.....	27
3.3.3. Facteurs influençant la transmission de <i>H. pylori</i> .....	28
4. Traitement et épidémiologie des résistances aux antibiotiques.....	29
5. Les déterminants des pathologies associées à l'infection par <i>H. pylori</i> .....	32
5.1. Les polymorphismes génétiques de l'hôte.....	32
5.2. Facteurs environnementaux.....	34
5.3. Facteurs de virulence bactériens.....	34
5.3.1. Inflammation et lésions tissulaires.....	36
5.3.1.1. La cytotoxine vacuolisante VacA.....	36
5.3.1.1.1. Génétique.....	38
5.3.1.1.2. Structure.....	39
5.3.1.1.3. Effets physiopathologiques.....	40
5.3.1.1.3.1. Vacuolisation des cellules eucaryotes.....	40
5.3.1.1.3.2. Induction de l'apoptose.....	41
5.3.1.1.3.3. Action immunosuppressive.....	42
5.3.1.1.3.4. Autres effets de la protéine VacA.....	42
5.3.1.2. L'îlot de pathogénicité <i>cag</i> et la protéine CagA.....	43
5.3.1.2.1. Organisation génétique de l'îlot.....	43
5.3.1.2.2. Activités biologiques.....	44
5.3.1.2.2.1. Effets de CagA sous forme phosphorylée.....	46
5.3.1.2.2.2. Effets de CagA sous forme non phosphorylée.....	49
5.3.1.2.2.3. Effets dépendants de l'appareil de sécrétion de type IV mais indépendants de CagA.....	50
5.3.1.3. L'inflammation.....	50
5.3.2. Adaptation à l'environnement gastrique.....	51
5.3.2.1. Les flagelles et le chimiotactisme.....	52

5.3.2.2. L'uréase .....	53
5.3.2.3. La réponse à l'acidité .....	55
5.3.2.4. Le lipopolysaccharide .....	56
5.3.2.5. Les adhésines.....	57
5.3.3. Instabilité chromosomique et altérations épigénétiques.....	58
6. Structure et dynamique du génome de <i>H. pylori</i> .....	59
6.1. Espèce procaryote .....	59
6.1.1. Définition .....	59
6.1.2. Hybridation ADN-ADN.....	60
6.1.3. Séquençage des ARNr ou séquençage des gènes codant pour les ARNr.....	60
6.1.4. Méthodes alternatives.....	61
6.1.5. Concept d'espèces écologiques.....	61
6.2. Structure génétique de <i>H. pylori</i> .....	64
6.2.1. Génome .....	64
6.2.2. Microdiversité .....	68
6.2.2.1. Populations génétiques modernes .....	69
6.2.2.2. Populations génétiques ancestrales .....	72
6.2.2.3. Coévolution entre l'homme et la bactérie .....	73
6.2.3. Macrodiversité.....	74
6.2.3.1. Organisation génomique : sythénie .....	74
6.2.3.2. "Core" génomique et pan-génome .....	76
6.2.3.3. Evolution de la macrodiversité.....	77
6.2.3.4. Contenu génétique et pathologie gastro-duodénale .....	78
6.2.4. Mécanisme générant la diversité génétique .....	79
6.2.4.1. Mutations ponctuelles .....	81
6.2.4.2. Variation de phase par glissement des brins d'ADN.....	81
6.2.4.3. Recombinaison.....	83
6.2.4.4. Systèmes de réparation de l'ADN.....	84
6.2.4.4.1. Réparation des mésappariements .....	85
6.2.4.4.2. Réparation par excision de nucléotides.....	86
6.2.4.4.3. Réparation par excision de base.....	86
6.2.4.4.4. Réparation par recombinaison.....	87
6.3. Changements génétiques pendant l'infection chronique <i>in vivo</i> .....	88
6.3.1. Au niveau microdiversité .....	88
6.3.2. Au niveau macrodiversité.....	89
6.3.3. Rôle de la variabilité génétique dans la colonisation de l'estomac.....	89
Partie II. Travail expérimental .....	92
1. Présentation des travaux de recherche .....	93
Article n°1: The peopling of the Pacific from a bacterial perspective.....	98
Article n°2: Evolutionary history of <i>Helicobacter pylori</i> sequences reflects past human migrations in Southeast Asia.....	103
Article n°3: Expansion of European <i>vacA</i> and <i>cagA</i> alleles to East-Asian <i>Helicobacter pylori</i> strains in Cambodia .....	108
Article n°4: Clinical relevance of <i>cagA</i> and <i>vacA</i> gene polymorphisms in <i>Helicobacter pylori</i> isolates from Senegalese patients .....	111
2. Discussion .....	115
2.1. Migrations humaines à travers un marqueur bactérien : <i>H. pylori</i> .....	115
2.1.1. <i>H. pylori</i> , le fidèle compagnon de l'homme dans ses migrations .....	115
2.1.2. <i>H. pylori</i> et les migrations humaines en Asie du sud-est et en Océanie .....	120
2.1.2.1. Océanie.....	120



2.1.2.2. Asie du sud-est .....	125
2.1.2.2.1. Anciennes migrations venue d'Inde .....	126
2.1.2.2.2. Anciennes migrations venue de Chine .....	132
2.1.3. Problèmes à résoudre .....	133
2.1.3.1. Dynamique de diffusion différente selon la population génétique .....	133
2.1.3.2. Accélération récente des flux migratoires .....	134
2.1.3.3. Diminution de la prévalence de l'infection .....	135
2.1.3.4. Difficultés pour obtenir de l'ADN de <i>H. pylori</i> .....	135
2.2. <i>H. pylori</i> et le cancer gastrique au Sénégal et au Cambodge .....	135
2.2.1. Le gène <i>cagA</i> .....	138
2.2.1.1. Le statut <i>cagA</i> et le cancer gastrique .....	138
2.2.1.2. L'extrémité 3' du gène <i>cagA</i> , un facteur clef .....	139
2.2.2. La cytotoxine vacuolisante VacA et le cancer gastrique .....	141
2.2.3. Diffusion d'allèles <i>cagA</i> et <i>vacA</i> des souches européennes vers les souches d'Asie de l'est .....	142
2.3. Conclusion .....	143
3. Perspectives .....	146
3.1. Etude des migrations humaines .....	146
3.1.1. Madagascar .....	146
3.1.2. Expansion Bantou .....	146
3.2. Structure génétique des gènes <i>hspA</i> et <i>glmM</i> .....	147
3.3. Relation entre la diversité génétique de l'hôte et de la bactérie infectante .....	147
3.4. Etude de la sensibilité des souches de <i>H. pylori</i> aux antibiotiques .....	148
Références bibliographiques .....	150
Annexes .....	174
Annexe I: Article n°5: CagA and VacA polymorphisms are associated with distinct pathological features in <i>Helicobacter pylori</i> -infected adults with peptic and non-peptic ulcer disease .....	174
Annexe II : Logiciel STRUCTURE (Falush <i>et al.</i> 2003) .....	175
Annexe III : Valorisation scientifique de la thèse .....	177
Index .....	178

## LISTE DES FIGURES

Figure 1. Morphologie de <i>H. pylori</i> en microscopie électronique (x 30 000).....	14
Figure 2. Pathologies associées aux infections à <i>H. pylori</i> .....	18
Figure 3. Prévalence de l'infection à <i>H. pylori</i> dans le monde .....	22
Figure 4. Colonisation de la muqueuse gastrique et genèse des lésions par <i>H. pylori</i> .....	35
Figure 5. Effets pathologiques de VacA sur les cellules de l'hôte.....	37
Figure 6. Organisation génétique du gène <i>vacA</i> et structure tridimensionnelle de la protéine VacA.....	40
Figure 7. Modèle du fonctionnement de l'appareil de sécrétion de type IV de <i>H. pylori</i> .....	45
Figure 8. Diversité génétique de l'extrémité C-terminale de la protéine CagA.....	46
Figure 9. Structure tridimensionnelle de l'uréase de <i>H. pylori</i> .....	54
Figure 10. Structure du lipopolysaccharide de <i>H. pylori</i> .....	57
Figure 11. Modèles théoriques de structure de la population bactérienne .....	62
Figure 12. Sous-populations et populations génétiques de <i>H. pylori</i> .....	70
Figure 13. La diversité génétique de <i>H. pylori</i> reflète les anciennes grandes migrations humaines.....	71
Figure 14. Proportion de nucléotides appartenant à chacune des cinq populations ancestrales dans chaque population génétique moderne de <i>H. pylori</i> déterminée par le logiciel STRUCTURE, modèle « linkage ».....	73
Figure 15. Synthénie entre le génome de la souche de <i>H. pylori</i> B38 et les souches de <i>H. pylori</i> 26695, J99, HPAG1, Shi470, P12, G27 et de <i>H. hepaticus</i> et de <i>H. acinonychis</i> .....	75
Figure 16. Arbres phylogénétiques construits à partir des données de la MLSA et des puces à ADN .....	77
Figure 17. Mécanismes générant la diversité génétique chez <i>H. pylori</i> .....	80
Figure 18. Glissement des brins d'ADN au niveau d'une séquence répétée durant la réplication .....	82
Figure 19. Systèmes de réparation de l'ADN au sein des procaryotes. ....	85
Figure 20. Rôle de la variabilité génétique dans la colonisation de l'estomac .....	91
Figure 21. DISTRUCT plot des proportions de nucléotides issues des populations ancestrales composant des souches de <i>H. pylori</i> isolées dans le nord de l'Inde (Ladakh), chez des musulmans et des bouddhistes (logiciel STRUCTURE V2.0, modèle linkage) .....	119
Figure 22. Migration des Austronésiens à partir de Taiwan en se basant sur la généalogie de 400 langues appartenant aux sous-groupes ou groupes des langues austronésiennes.....	121
Figure 23. <i>hspMaori</i> , marqueur de la migration des Austronésiens en Océanie à partir de Taiwan.....	124
Figure 24. Populations génétiques de <i>H. pylori</i> en Asie continentale, au Japon et à Taiwan	126
Figure 25. La population génétique <i>hpEurope</i> , marqueur d'une migration venue d'Inde.....	129
Figure 26. Topographie du Vietnam, Laos et Cambodge .....	131
Figure 27. Populations génétiques de <i>H. pylori</i> et incidence du cancer gastrique en Asie de l'est et du sud-est.....	137
Figure 28. Migrations humaines décrites à partir du polymorphisme génétique de <i>H. pylori</i>	144

## LISTE DES TABLEAUX

Tableau 1. Impact de la diversité génétique de l'extrémité C-terminale de la protéine CagA	48
Tableau 2. Caractéristiques des génomes de <i>Helicobacter</i>	66
Tableau 3. Fréquence du type de protéine CagA en Asie de l'est et du sud-est et incidence du cancer gastrique	140
Tableau 4. Incidence du cancer gastrique et génotype <i>vacA</i> m1 et m2 dans les pays d'Asie où les souches hspEAsia sont endémiques	142
Tableau 5. Résistance des souches de <i>H. pylori</i> aux antibiotiques testés	148

## LISTE DES ABREVIATIONS

ADNmt : ADN mitochondrial  
AE1: ancestral Europe1  
AE2: ancestral Europe2  
ANI (*average nucleotide identity*)  
ARN: acide ribonucléique  
BabA (*blood group antigen-binding adhesin*)  
BER (*base excision repair*): système de réparation par excision de base  
*C. jejuni* : *Campylobacter jejuni*  
CagA (*cytotoxin associated antigen A*)  
CagA-P: CagA sous sa forme phosphorylée  
*cag* PAI: îlot de pathogénicité *cag*  
CAPE (*caffeic acid phenethyl ester*)  
CDS (coding DNA sequence) : séquence codante  
CG : cancer gastrique  
CM: motif multimérique  
CMI : concentration minimale inhibitrice  
COG (*clusters of orthologous groups*)  
*E. coli*: *Escherichia coli*  
EGF (*epidermal growth factor*) : facteur de croissance épidermique  
ELISA (*enzyme-linked immunosorbent assay*)  
EPIYA: Glu-Pro-Ile-Tyr-Ala  
*H. acinonychis*: *Helicobacter acinonychis*  
*H. bovis*: *Helicobacter bovis*  
*H. canis*: *Helicobacter canis*  
*H. hepaticus*: *Helicobacter hepaticus*  
*H. suis*: *Helicobacter suis*  
HAA: hybridation ADN-ADN  
*H. pylori* : *Helicobacter pylori*  
HGF (*hepatocyte growth factor*) : facteur de croissance hépatocyte  
HVSI : région hypervariable 1 de l'ADN mitochondrial  
HVSII : région hypervariable 2 de l'ADN mitochondrial  
IC : intervalle de confiance  
IFN- $\gamma$  : interferon-gamma  
Ig : immunoglobuline  
IL: interleukine  
IL1  $\beta$ : interleukine-1 bêta  
IL1Ra : antagoniste du récepteur de l'IL1  
ISR (*interspersed sequences of the recipient*)  
JAM (*junctional adhesion molecule*)  
LPS: lipopolysaccharide  
MALT (*mucosal associated lymphoid tissue*)  
MARK (*microtubules affinity regulating kinases*)  
MLSA (*MultiLocus Sequence Analysis*): analyse de séquences multilocus  
*M. tuberculosis* : *Mycobacterium tuberculosis*  
NER (*nucleotide excision repair*): système de réparation par excision de nucléotides  
NF-kappa B (*nuclear factor-kappa B*): facteur de transcription nucléaire kappa B  
NFAT (*nuclear factor of activated T-cells*): facteur nucléaire des lymphocytes T activés  
OMP (*outer membrane protein*): protéine de la membrane externe

OR: odds ratio  
PAR1 : *partitioning-defective 1*  
pb : paires de base  
PCR (*polymerase chain reaction*): amplification en chaîne par polymérase  
PCR-RFLP (*PCR-based restriction fragment length polymorphism*)  
PG: peptidoglycane  
PLP: protéine liant la pénicilline  
PZ (*plasticity zone*): zone de plasticité  
RAPD (*random amplified polymorphism DNA*)  
SabA (*sialic acid-binding adhesin*)  
SH2 (*Src homology 2*)  
ST (*sequence type*)  
SRM (*mismatch repair*): système de réparation des mésappariements  
R-M: restriction-modification  
T4SS (*type four secretion system*) : appareil de sécrétion de type IV  
TAK1 (*transforming growth factor  $\beta$  activated kinase 1*)  
TGF- $\beta$  (*transforming growth factor beta*) : facteur de croissance transformant bêta  
Th-1: T helper de type 1  
TNF-alpha (*tumor necrosis factor-alpha*): facteur de nécrose tumorale alpha  
TRAF6 (*tumour necrosis factor factor 6-mediated receptor associated*)  
Tregs: lymphocytes T régulateurs  
*W. succinogenes* : *Wolinella succinogenes*  
ZO (*zonula occludens*)

**PARTIE I. GENERALITES**

## 1. HISTORIQUE ET TAXONOMIE

Des travaux récents indiquent que *Helicobacter pylori* (*H. pylori*) était déjà présent dans l'estomac des *Homo sapiens* il y a environ 58 000 ans, avant les grandes migrations de l'homme moderne partant de l'Afrique de l'est vers l'Asie et l'Europe (Linz *et al.* 2007). Les premières descriptions de la présence d'une bactérie hélicoïdale dans des estomacs humains datent de 1906 (Krienitz 1906). Diverses publications mentionnent ensuite la présence de bactéries dans l'estomac humain à l'examen histologique mais ce n'est qu'en 1982 que deux chercheurs australiens, R. Warren (pathologiste) et B. Marshall (gastroentérologue), réussirent à la cultiver. Assimilant la morphologie hélicoïdale de la bactérie à celle des bactéries du genre *Campylobacter*, ils pensèrent à réaliser les cultures de biopsies gastriques sur gélose chocolat en atmosphère microaérophile. Un temps d'incubation trop court (2-3 jours) ne leur permit tout d'abord pas d'obtenir de culture. En avril 1982, après 34 essais infructueux et à la faveur d'un week-end de Pâques prolongé pendant lequel les milieux de culture avaient été oubliés dans l'étuve, la culture s'avéra positive au bout de 5 jours, indiquant la nécessité d'un temps de culture plus long pour isoler la bactérie. Elle fut alors assimilée au genre *Campylobacter* sur la base de sa morphologie spiralée et de la présence de flagelles engainés. Dans leur publication originelle, R. Warren et B. Marshall postulèrent que la plupart des ulcères gastriques étaient causés par cette bactérie, et non par le stress ou la nourriture épicée, comme on le pensait auparavant. Cette découverte leur valut le prix Nobel 2005 de physiologie et de médecine.



**Figure 1. Morphologie de *H. pylori* en microscopie électronique (x 30 000)**

Photo fournie par P. Gounon de la Station de microscopie électronique de l'Institut Pasteur

La communauté médicale tarda à reconnaître la responsabilité de cette bactérie dans la genèse des ulcères gastro-duodénaux, pensant qu'aucune bactérie ne pouvait survivre dans l'environnement acide de l'estomac. Il a fallu que B. Marshall absorbe une culture de *H. pylori* et développe une gastrite qu'il traita avec des antibiotiques, satisfaisant ainsi à 3 des 4 postulats de Koch, pour convaincre la communauté médicale. Coghlan en 1987 (Coghlan *et al.* 1987) puis Marshall en 1988 (Marshall *et al.* 1988) confirmèrent l'importance de l'éradication de la bactérie pour prévenir les récurrences d'ulcère. Dans cette dernière étude, chez 100 patients présentant une infection à *H. pylori* et un ulcère duodénal, le taux de récurrence était de 84 % chez les patients pour lesquels la bactérie n'avait pas été éradiquée et de 21 % dans le cas contraire. En 1994, le « National Institutes of Health » affirmait que la plupart des ulcères gastriques récurrents étaient causés par *H. pylori* et recommandait l'adjonction d'antibiotiques. En effet, avant que ne soit reconnu le rôle de *H. pylori*, les ulcères gastriques étaient traités par des antiacides qui n'évitaient pas les rechutes. Le subsalicylate de bismuth était largement utilisé mais bien que vraisemblablement bactéricide, il fut abandonné en raison de sa toxicité neurologique.

En 1994, *H. pylori* a été classée parmi les agents carcinogènes de classe I par l'Organisation Mondiale de la Santé. Il s'agit de la première bactérie carcinogène reconnue. Elle fut initialement nommée *Campylobacter pyloridis*, puis *Campylobacter pylori* (après correction grammaticale latine). Sur la base de la séquence de l'acide ribonucléique (ARN) 16S, de la composition en acide gras et de la morphologie des flagelles, cette bactérie a été renommée en 1989 *Helicobacter pylori* (Goodwin *et al.* 1989), représentant ainsi le chef de file d'un nouveau genre bactérien. Le genre *Helicobacter* appartient à la famille des *Helicobacteriaceae* (Campylobactérial, division epsilon de la classe des *Proteobacteria*), comprenant aussi les genres *Flexispira*, *Sulfuricurvum*, *Sulfurimonas*, *Thiovulum* et *Wolinella*<sup>1</sup>. Le genre *Helicobacter* regroupe actuellement une cinquantaine d'espèces reconnues et plus de 160 souches en cours de classification. Ces espèces sont capables d'infecter de nombreuses espèces de mammifères mais également des oiseaux et des poissons. Un haut degré de spécificité entre les bactéries du genre *Helicobacter* et leur hôte respectif est retrouvé : *H. pylori* est spécifique de l'homme, *H. suis* du cochon, *H. bovis* du bovin, *H. felis* du chat et *H. canis* du chien (Solnick and Schauer 2001).

---

<sup>1</sup>[http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=210&lvl=3&lin=f&keep=1&src\\_hmode=1&unlock](http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=210&lvl=3&lin=f&keep=1&src_hmode=1&unlock)



## 2. PATHOLOGIES ASSOCIEES AUX INFECTIONS A *H. PYLORI*

### 2.1. Pathologies associées aux infections à *H. pylori*

L'infection par *H. pylori*, acquise le plus souvent dans l'enfance, persiste toute la vie. Certaines pathologies qui lui sont reliées n'apparaissent qu'après de nombreuses années, une quarantaine d'années pour le cancer gastrique (CG). L'interaction entre les facteurs de virulence de la bactérie, la susceptibilité individuelle de l'hôte et l'environnement est responsable de l'évolution de l'infection. Toute infection de l'estomac par *H. pylori* entraîne une inflammation de la muqueuse se traduisant par une gastrite histologique. Il n'existe donc pas de porteurs véritablement sains de *H. pylori*.

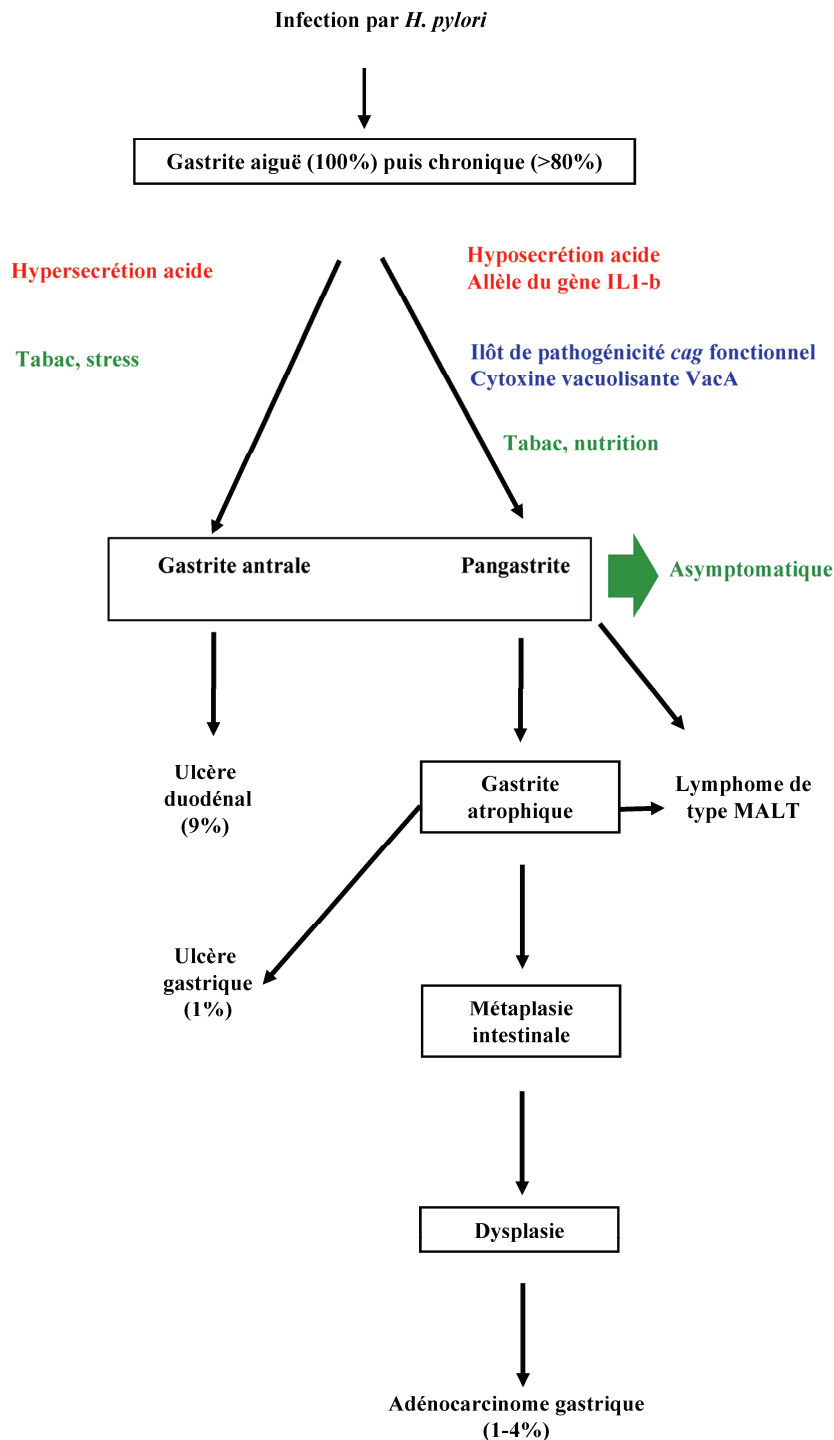
Une gastrite localisée principalement au niveau antral est associée à une hypergastrinémie et à une hyperchlorhydrie, pouvant entraîner une métaplasie gastrique au niveau duodénal puis un ulcère duodénal. Dans ce cas, un effet protecteur vis-à-vis du développement d'un CG a été mis en évidence. Par opposition, une pangastrite (infection de la partie antrale et fundique de l'estomac) est accompagnée d'une hypochlorhydrie liée à la diminution de la sécrétion acide. L'évolution de l'infection vers le cancer suit une progression séquentielle appelée « cascade de Correa », allant de la pangastrite vers la gastrite atrophique, la métaplasie intestinale précurseur de la dysplasie, puis la transformation maligne (Konturek *et al.* 2003) (Figure 2).

Approximativement 90 à 95 % des ulcères duodénaux et 70 à 75 % des ulcères gastriques sont associés à une infection à *H. pylori* (Covacci *et al.* 1999). Les ulcères chez des individus ne présentant pas d'infection à *H. pylori* sont dus à la prise d'anti-inflammatoires non stéroïdiens. Il a été estimé que le risque pour un patient infecté par *H. pylori* de développer un ulcère peptique est de 10 % (Parsonnet 1995). L'infection à *H. pylori* est associée à deux types de néoplasies : les lymphomes de type MALT (Mucosal Associated Lymphoïde Tissue) et les adénocarcinomes. Une infection à *H. pylori* est observée chez 90 % des patients présentant un lymphome du MALT (Konturek *et al.* 2000). Le risque de développer un adénocarcinome au cours de sa vie pour un patient infecté par *H. pylori* est de 1 à 4 % (Kuipers 1999). La relation entre l'infection à *H. pylori* et l'adénocarcinome gastrique est suggérée avant tout par des données épidémiologiques. Une étude cas-témoin, incluant 109 malades atteints de CG et 109 sujets témoins appariés a montré que 84 % des malades étaient infectés par *H. pylori* contre 61 % des témoins et que l'Odds Ratio (OR) (intervalle de confiance à 95 %) de CG chez les patients infectés était de 3.6 (1.8-7.3) (Parsonnet *et al.*

1991). En 2001, Uemura *et al.* ont montré sur une série de 1526 patients suivis sur une période de 8 ans, que seuls les patients présentant une infection à *H. pylori* développaient un CG (2.9 % vs 0 %) (Uemura *et al.* 2001).

Sur le plan histologique, les adénocarcinomes peuvent être de type intestinal bien différencié ou de type diffus indifférencié. Une faible proportion des adénocarcinomes de type diffus est de type héréditaire lié à des mutations germinales du gène suppresseur de tumeur CDH1 codant la molécule d'adhésion intercellulaire E-cadherine (Guilford *et al.* 1998). Des études épidémiologiques ont montré que *H. pylori* joue un rôle clef dans le développement des deux types d'adénocarcinomes (Parsonnet *et al.* 1991; Uemura *et al.* 2001; Nomura *et al.* 2005). Une étude cas-témoin a montré une association significative entre le cancer du cardia et une gastrite atrophique à *H. pylori* (OR = 3.9 ; Intervalle de confiance (IC) 95 % : 1.8 – 8.7), ainsi qu'avec les manifestations cliniques liées au reflux gastro-œsophagien (OR = 10.1 ; IC 95 %: 2.3 - 44.4) (Derakhshan *et al.* 2008). Les adénocarcinomes associés à une atrophie présentaient un type histologique intestinal et diffus tandis que les adénocarcinomes sans atrophie gastrique et associés aux manifestations cliniques de reflux gastro-œsophagien présentaient uniquement une histologie de type intestinal. Une étude rétrospective suggère également que les cancers de la partie haute de l'estomac (après exclusion soigneuse des cancers situés au dessus du cardia) sont majoritairement liés à *H. pylori* (Bornschein *et al.* 2010).

Actuellement, le CG arrive en quatrième position parmi les cancers les plus diagnostiqués dans le monde, avec approximativement un million de nouveaux cas chaque année et est la deuxième cause de décès parmi les cancers, avec approximativement 800 000 morts par an. Son pronostic est mauvais avec une survie de 10 à 15 % à 5 ans (Parkin *et al.* 2005). Alors que l'incidence globale et la mortalité par CG sont en diminution dans les pays développés, son incidence parmi les lésions malignes s'accroît dans les pays en voie de développement. Les incidences de CG les plus élevées sont retrouvées dans l'est de l'Asie (Chine, Corée, Mongolie et Japon) et en Amérique du Sud (Chili, Colombie, Equateur et Pérou) (Globocan 2010). L'identification et un meilleur contrôle des facteurs de risque semblent les mesures les plus efficaces en matière de prévention.



**Figure 2. Pathologies associées aux infections à *H. pylori***

Les principaux facteurs connus de l'hôte, de l'environnement et de la bactérie pouvant influencer l'évolution des pathologies sont indiqués respectivement en rouge, vert et bleu. Les pourcentages indiquent le risque de développement d'une pathologie donnée chez les personnes infectées.

## **2.2. Indications de la recherche et du traitement de l'infection à *H. pylori***

### **2.2.1. Ulcère gastrique et duodéal**

Le rôle de l'éradication de *H. pylori* sur la cicatrisation des ulcères est différent selon qu'il s'agit d'un ulcère duodéal ou d'un ulcère gastrique. Une méta-analyse incluant 53 études a montré que l'éradication permettait un taux de cicatrisation supérieur par rapport à un traitement anti-sécrétoire seul en cas d'ulcère duodéal, mais qu'elle n'avait aucun bénéfice en cas d'ulcère gastrique. En revanche, l'éradication avait un effet bénéfique majeur dans la prévention de la récurrence des ulcères qu'ils soient duodéaux ou gastriques (Ford *et al.* 2004). Selon une méta-analyse, les taux de récurrence à 1 an des ulcères duodéaux et gastriques étaient respectivement de 1 et 3 % après éradication, versus 35 % et 40 % sans éradication de la bactérie (Leodolter *et al.* 2001).

### **2.2.2. Lymphome du MALT**

La très forte prévalence de l'infection à *H. pylori* chez les malades atteints de lymphome du MALT (90 %) est un très fort argument (Konturek *et al.* 2000). Le traitement de l'infection à *H. pylori* permet la régression du lymphome du MALT de bas grade chez 60 à 90 % des patients (Montalban and Norman 2006) avec maintien d'une rémission à 5 ans chez 87 à 97 % des patients initialement répondeurs selon plusieurs études prospectives (Chen *et al.* 2005; Nakamura *et al.* 2005; Wundisch *et al.* 2005). Une méta-analyse de 34 études ayant inclus 1271 patients a montré une régression du lymphome du MALT chez 973 (78.5 %) des 1250 patients traités avec succès pour leur infection par *H. pylori* (Zullo *et al.* 2009). L'éradication de *H. pylori* fait donc partie de la prise en charge initiale de tous les lymphomes du MALT gastriques.

### **2.2.3. Cancer gastrique et lésions pré-néoplasiques**

Des antécédents de CG chez un apparenté du premier degré accroissent le risque de cancer en cas d'infection par *H. pylori*. Une étude de Leung *et al.* a montré que la présence d'une métaplasie intestinale chez de tels patients était significativement associée à l'infection par *H. pylori*, au sexe masculin, à l'âge et à la survenue d'un CG dans la fratrie (Leung *et al.* 2005). Une méta-analyse de Rokkas incluant 11 études récentes confirme l'augmentation des lésions pré-néoplasiques (atrophie et/ou métaplasie intestinale) chez de tels patients infectés par *H. pylori* par rapport au groupe des patients témoins (Rokkas *et al.* 2010).

A ce jour, aucune étude n'a démontré formellement que l'éradication de *H. pylori* permettait de réduire le risque de CG (Malfertheiner *et al.* 2005). Ce type d'étude est difficile à mettre en place puisqu'elle nécessite un suivi prolongé de cohortes importantes car la carcinogenèse gastrique est un phénomène qui s'étend sur plusieurs années et ne concerne qu'un faible pourcentage des patients infectés (Suerbaum and Michetti 2002). De plus, la bactérie n'est en général plus détectée dans l'estomac à partir du stade de métaplasie intestinale, ce qui rend difficile la mise en évidence de ses caractéristiques et de son implication directe dans la maladie. Les conséquences de l'éradication de *H. pylori* sur la survenue des cancers ont été évaluées par une étude comparative incluant 1630 sujets chinois asymptomatiques randomisés pour recevoir un traitement d'éradication ou un placebo (Wong *et al.* 2004). A l'issue du suivi de 7,5 ans, une tendance significative à la réduction de l'incidence des CG était trouvée chez les sujets dont l'infection à *H. pylori* avait été traitée avec succès. L'analyse par sous groupe révélait sept cas (6 %) dans le groupe des patients ayant initialement des lésions pré-néoplasiques et aucun cas de cancer dans le groupe des patients sans lésions pré-néoplasiques. Ces résultats suggèrent que l'éradication de *H. pylori* pourrait réduire le risque de cancer. Dans une étude incluant 80255 patients taïwanais présentant un ulcère gastroduodéal et traités pour une infection à *H. pylori* entre 1997 et 2004, Wu *et al.* montraient que le risque de CG entre le groupe ayant eu un traitement d'éradication précoce (avant 1 an suivant la découverte de l'ulcère) et la population générale n'était pas modifié (OR=1.05 ; IC 95 % : 0.96-1.14). En revanche, une éradication tardive de l'infection à *H. pylori* (après 1 an suivant la découverte de l'ulcère) était associée à un risque accru de CG (OR= 1.36 ; IC 95 % : 1.24-1.49) (Wu *et al.* 2009).

En revanche, il existe de nombreuses études concluant à un effet bénéfique de l'éradication de *H. pylori* sur l'évolution des lésions pré-néoplasiques (Correa *et al.* 2000; Zhou *et al.* 2003; Leung *et al.* 2004; Ley *et al.* 2004; Wong *et al.* 2004). Une étude réalisée en Colombie chez 795 adultes avec des lésions pré-néoplasiques a montré que l'éradication de *H. pylori* s'accompagnait d'une régression des lésions proportionnelle au carré du délai écoulé depuis le traitement (Mera *et al.* 2005). Dans une méta-analyse de 6 études comparatives sur l'effet de l'éradication sur l'évolution des lésions pré-néoplasiques de la muqueuse gastrique, une réduction du risque de progression était observée dans le groupe traité. Quatre de ces études avaient été réalisées en Asie, dans des régions avec de fortes prévalences du CG (Fuccio *et al.* 2009).

En conclusion, l'infection par *H. pylori* doit être recherchée et traitée si besoin chez les patients ayant:

- des antécédents familiaux de CG au premier degré,
- des lésions pré-néoplasiques de la muqueuse gastrique (atrophie avec ou sans métaplasie intestinale),
- eu une résection localisée d'un CG.

### 3. EPIDEMIOLOGIE DE L'INFECTION

Bien que depuis 1982, de nombreuses études aient permis de mieux connaître l'épidémiologie de *H. pylori*, de nombreuses questions restent en suspens, notamment le mode de transmission de cette bactérie qui n'est toujours pas formellement démontré.

#### 3.1. Prévalence, incidence et réinfection

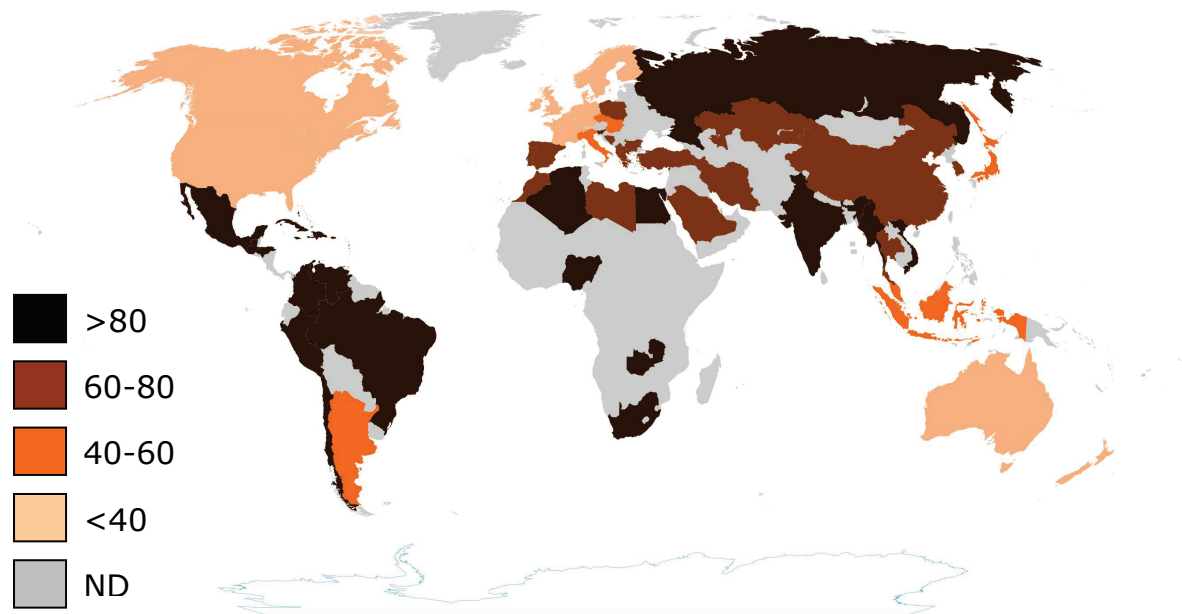
Lorsque l'on étudie la prévalence d'une infection, deux paramètres doivent être pris en compte : i) la population étudiée qui doit être représentative de la population cible, ii) les méthodes de détection choisies qui doivent être suffisamment sensibles et spécifiques.

Le test respiratoire et la recherche d'antigènes de *H. pylori* dans les selles sont considérés comme les méthodes les plus fiables mais la sérologie est la méthode la plus utilisée car la plus économique. La plupart des tests sérologiques commercialisés se basent sur la détection des immunoglobulines (Ig) G anti-*H. pylori*, ce qui comporte deux inconvénients majeurs :

- ces tests ne différencient pas une infection passée d'une infection active,
- les IgG se négativent après la guérison de l'infection.

L'infection à *H. pylori* atteint plus de 50 % de la population dans le monde. La prévalence de l'infection est beaucoup plus importante dans les pays en voie de développement (80-95 %) que dans les pays industrialisés (15-30 %) (Figure 3). Cependant, cette séparation devient de moins en moins claire avec l'augmentation rapide du niveau socio-économique de certaines sous-populations des pays en voie de développement et l'utilisation de plus en plus répandue des antibiotiques. Ces dernières années, la prévalence de l'infection a diminué de façon importante à la fois dans les pays développés et les pays en voie de développement. Cette baisse devrait continuer dans le futur, conséquence de l'amélioration des conditions de vie et d'hygiène dans le monde (Frenck and Clemens 2003; Azevedo *et al.* 2007; Lehours and Yilmaz 2007; Bruce and Maaroos 2008; Azevedo *et al.* 2009; Zhang *et al.*

2009). Cette diminution progressive de l'incidence dans les pays développés au cours du siècle dernier a été accompagnée d'une diminution de l'incidence du CG (Parsonnet 1995).



**Figure 3. Prévalence de l'infection à *H. pylori* dans le monde**

D'après Azevedo *et al.* (Azevedo *et al.* 2007)

L'infection par *H. pylori* est acquise essentiellement durant l'enfance (Glynn *et al.* 2002). L'incidence chez les enfants est de 3 à 10 % par an dans les pays en voie de développement et de 1 % par an dans les pays industrialisés (Ernst and Gold 2000). La récurrence (réinfection ou infection mal traitée) d'une infection à *H. pylori* après éradication de la bactérie au niveau de l'estomac semble rare dans les pays développés contrairement aux pays en voie de développement (Gisbert 2005).

### 3.2. Réservoir de virus

Le seul réservoir de virus connu est l'estomac humain. Malgré de nombreuses recherches, un réservoir environnemental n'a pas encore été mis en évidence de façon formelle. Cependant, il pourrait jouer un rôle majeur, particulièrement dans les pays en voie de développement (Azevedo *et al.* 2007).

Plusieurs hypothèses ont été explorées : l'eau, les aliments, les animaux.

### 3.2.1. Arguments en faveur de l'eau

De nombreuses études épidémiologiques ont identifié l'eau potable comme facteur de risque d'infection à *H. pylori* (Klein *et al.* 1991; Goodman *et al.* 1996). A l'aide de techniques d'amplification en chaîne par polymérase (PCR) et d'hybridation *in situ* en fluorescence, l'ADN de *H. pylori* a été retrouvée dans l'eau ou au sein de biofilms présents au niveau de puits, rivières et des réseaux de distribution (Fujimura *et al.* 2004; Piqueres *et al.* 2006; Bockelmann *et al.* 2009). Cependant, la détection d'ADN de *H. pylori* par PCR ne prouve en rien que ces bactéries sont viables et donc transmissibles. Les essais de culture de *H. pylori* à partir de l'eau n'ont jamais été couronnés de succès à de rares exceptions près (Lu *et al.* 2002; Cellini *et al.* 2008).

Des expériences de laboratoire ont montré que des formes coccoïdes de *H. pylori* étaient présentes jusqu'à 1 semaine dans de l'eau contaminée artificiellement et conservée à + 4°C (Fan *et al.* 1998). Les opinions divergent concernant ces formes coccoïdes qui sont incapables de se diviser dans les conditions de cultures usuelles mais maintiennent un métabolisme basal : la concentration d'adénosine triphosphate intracellulaire a été évaluée comme 1000 fois plus faible que dans les formes bacillaires (Sorberg *et al.* 1996). Certains auteurs pensent qu'il s'agit de formes de dégénérescence. L'absence de virulence chez le porcelet gnotobiotique infecté expérimentalement (Eaton *et al.* 1995), les modifications structurales et métaboliques observées lors de la conversion en formes coccoïdes (Sorberg *et al.* 1996 ; Kusters *et al.* 1997) seraient en faveur de cette hypothèse. De plus, des modifications morphologiques similaires surviennent lors de la lyse de formes bacillaires de *H. pylori* médiée par la protéine E, conséquence d'une infection par le bactériophage PhiX174 (Panthel *et al.* 2003). D'autres auteurs suggèrent que les formes coccoïdes sont viables, que ce sont des formes de résistance quand le milieu de culture est appauvri, qu'elles pourraient représenter une forme de transmission de l'infection mais également une forme de résistance aux traitements antibiotiques. Des études portant sur différentes activités cellulaires semblent supporter cette hypothèse (Costa *et al.* 1999 ; Mizoguchi *et al.* 1999 ; Monstein and Jonasson 2001 ; Nilsson *et al.* 2002). De façon intéressante, il semble que la transformation en forme coccoïde est un phénomène actif car prévenue lors de la survenue d'une mutation dans le gène *cdrA*, gène impliqué dans la division cellulaire (Takeuchi *et al.* 1998). Ces formes coccoïdes viables et non cultivables pourraient être une source de contamination, peut-être par l'intermédiaire du biofilm présent dans tous les systèmes de stockage de l'eau (Bunn *et al.* 2002).



### 3.2.2. Arguments en faveur des aliments

Au moins deux études ont mis en évidence une relation entre la consommation de crudités et la transmission de *H. pylori* (Hopkins *et al.* 1993; Goodman *et al.* 1996). La contamination des crudités se ferait par l'intermédiaire d'eau contaminée utilisée pour l'irrigation ou le lavage, ce qui implique que *H. pylori* soit capable de survivre dans l'eau. A notre connaissance, aucune étude n'a essayé de détecter la bactérie par culture ou PCR directement à partir de ces produits. Le lait a été également impliqué comme possible source de contamination par des études épidémiologiques. Au Mexique, une forte consommation de lait était corrélée à une prévalence plus élevée de l'infection, cependant l'inverse était observé en Italie (Azevedo *et al.* 2007). Cette différence pourrait s'expliquer par une qualité microbiologique variable du lait. Dore *et al.* ont mis en évidence une prévalence significativement plus élevée dans une communauté d'éleveurs de brebis (98 %) et les membres de leur famille (73 %), par rapport à un groupe témoin sans contact avec des brebis (43 %) (Dore *et al.* 1999). Cette équipe a même réussi à cultiver *H. pylori* à partir de lait cru de brebis (Dore *et al.* 1999; Dore *et al.* 2001), ce qu'un large screening de 400 échantillons en Turquie n'a pas réussi à confirmer (Turutoglu and Mudul 2002). *H. pylori* a également été cultivée à partir d'un échantillon de lait de vache cru et dans cette même étude, la bactérie était détectée par PCR (gène *ureA*) dans 72 % des échantillons de lait cru et dans 55 % des échantillons de lait pasteurisé testés (Fujimura *et al.* 2002). La viabilité de *H. pylori* a été étudiée à différents intervalles de temps par culture bactériologique de différents produits alimentaires conservés à +4°C après une contamination artificielle. La bactérie a pu être cultivée jusqu'à 5 jours à partir du lait ou du tofu et jusqu'à 2 jours à partir de la laitue ou du poulet cru. Cependant, elle n'a pu être isolée à partir de yaourt (Poms and Tatini 2001). Ces différences de viabilité pourraient être expliquées par une survie prolongée dans un environnement acide et fortement humide (Jiang and Doyle 1998).

### 3.2.3. Arguments en faveur du tractus digestif des animaux

La possibilité que *H. pylori* soit une zoonose a été évoquée après la publication de deux études séro-épidémiologiques concernant les employés d'un abattoir et prenant comme cas-témoin des personnes sans aucune relation avec les animaux. Ces résultats ont été controversés en raison de possibles réactions croisées entre *H. pylori* et *Campylobacter jejuni*, ce dernier étant fréquent dans le tractus gastro-intestinal des animaux (Fox 1995). Une revue de la littérature indique qu'aucune étude épidémiologique n'a montré de façon claire l'existence d'un quelconque réservoir chez l'animal. Récemment, la présence de *Helicobacter*

spp. a été recherchée dans les fèces et la salive du chat (Ghil *et al.* 2009). Bien que 77 % des échantillons aient été positifs par PCR, les amorces spécifiques de l'espèce ne montraient pas d'amplification du gène cible. De plus larges études sont nécessaires pour clarifier le rôle de ce réservoir.

### **3.3. Transmission de l'infection par *H. pylori***

#### **3.3.1. Voies de transmission de l'infection**

Bien qu'identifiée comme bactérie strictement humaine, la voie de transmission reste cependant mal comprise ; certains arguments sont en faveur d'une contamination par voie gastro-orale, d'autres en faveur d'une contamination par voie féco-orale ou oro-orale. Une revue de la littérature publiée sur ce sujet en 2008 identifie un nombre important de travaux sans qu'un consensus sur la voie de transmission apparaisse (Azevedo *et al.* 2009).

##### **3.3.1.1. Arguments en faveur de la transmission gastro-orale**

La prévalence plus élevée de l'infection chez les gastro-entérologues évoquent une contamination par des microgouttelettes de jus gastrique lors des endoscopies (Lin *et al.* 1994; Hildebrand *et al.* 2000). Ce mode de transmission a principalement été évoqué pour les jeunes enfants chez qui le reflux gastro-intestinal et les vomissements sont fréquents (Mitchell *et al.* 1992; Axon 1995; Leung *et al.* 1999). Un argument indirect a été apporté par une récente étude épidémiologique qui a montré que la présence d'un individu infecté présentant une gastroentérite avec vomissements était associée à un risque d'infection 6 fois plus élevé pour la communauté habitant sous le même toit (Perry *et al.* 2006). Dans une étude réalisée par Parsonnet *et al.*, tous les échantillons de vomissements de personnes infectées étaient positifs, ainsi que 37 % des échantillons d'air prélevés lors de l'épisode de vomissement. Deux autres études ont également montré la présence de *H. pylori* par culture à partir des vomissements (Leung *et al.* 1999; Young *et al.* 2000). Récemment, il a été montré par biologie moléculaire que la bactérie était 600 fois plus abondante dans les vomissements que dans les selles (Janzon *et al.* 2009).

##### **3.3.1.2. Arguments en faveur de la transmission oro-orale**

Au Bangladesh, les mères hindoues, contrairement aux mères musulmanes, appliquent régulièrement de la salive sur leur mamelon avant de donner le sein. Une étude a montré que la prévalence de l'infection était significativement plus élevée chez les bébés hindous que

chez les bébés musulmans et pourrait être due à une transmission orale (Clemens *et al.* 1996). La prémastication d'aliments par la maman a été également avancée comme source de contamination de leurs enfants dans différentes populations (Azevedo *et al.* 2009). L'utilisation de baguettes et le partage d'un même plat au sein de communautés chinoises ont été identifiés comme facteurs de risque de transmission (Chow *et al.* 1995), bien que *H. pylori* n'ait été que rarement retrouvée sur des baguettes (Leung *et al.* 1999). Un contre-argument est l'absence d'une corrélation claire entre la prévalence de l'infection et le métier de dentiste (Lin *et al.* 1998; Honda *et al.* 2001; Loster *et al.* 2009). Après recherche de *H. pylori* dans la cavité buccale par PCR et dans l'estomac par culture chez 94 personnes (Burgers *et al.* 2008), 17 % des individus infectés par *H. pylori* présentait également la bactérie au niveau de la cavité buccale. Cependant, les résultats de la détection de *H. pylori* dans la cavité buccale par PCR restent contradictoires. La culture de la bactérie à partir de la cavité buccale a été décrite dans peu d'études et pourrait être une conséquence de régurgitations transitoires à partir de l'estomac (Luman *et al.* 2002; Dowsett and Kowolik 2003). Des études dans un milieu expérimental mimant la salive a montré que la culture commençait à être inhibée après 1 heure à 37°C mais cette inhibition diminuait lors de l'ajout de salive humaine (Haukioja *et al.* 2004). Par ailleurs, la présence d'une bactérie au niveau d'une localisation particulière ne démontre pas forcément l'implication de ce site dans la transmission.

### 3.3.1.3. Arguments en faveur de la transmission féco-orale

Bien qu'à l'évidence *H. pylori* transite par l'intestin, cette bactérie n'est pas adaptée à ce milieu. Elle est sensible à la bile et il est peu vraisemblable qu'elle survive au transit intestinal (Azevedo *et al.* 2007). Une étude épidémiologique récente a montré que le risque d'infection pour une communauté habitant sous le même toit n'était pas significativement plus élevé lorsqu'un des membres infecté présentait une diarrhée (Perry *et al.* 2006). Des études ont exploré la transmission fécale-orale de *H. pylori* en utilisant le modèle de l'hépatite A, virus signant la contamination fécale, mais les résultats sont peu convaincants [pour une revue (BinSaeed 2010)]. La culture bactériologique à partir des selles a été rarement couronnée de succès sauf dans de rares études où la plupart des patients présentaient un transit intestinal accéléré (Luman *et al.* 2002; Dowsett and Kowolik 2003). Des essais de détection de *H. pylori* par PCR ou par méthode immuno-enzymatique (ELISA ou « enzyme-linked immunosorbent assay ») ont identifié la présence de la bactérie dans les selles de populations infectées (Russo *et al.* 1999; Ndip *et al.* 2004; Queralt *et al.* 2005), mais il est essentiel de se

rappeler que le fait de trouver de l'ADN ou un antigène de *H. pylori* n'implique pas forcément que les bactéries soient vivantes.

### 3.3.2. Mode de transmission de l'infection

Plusieurs études de typage génétique ont montré que l'acquisition de l'infection peut survenir à la fois de l'intérieur et de l'extérieur de la famille mais que la transmission intrafamiliale était prédominante (Vincent *et al.* 1994; Kivi *et al.* 2003; Owen and Xerry 2003). Cependant, ces études comportaient un ou plusieurs biais majeurs : i) l'étude du mode de transmission de l'infection est compliquée par la possibilité d'une infection mixte chez un même individu, paramètre qui n'a en général pas été pris en compte, un seul isolat dans une région précise de l'estomac a été le plus souvent investigué, ii) les méthodes utilisées manquaient bien souvent de pouvoir discriminant pour identifier des différences génétiques entre isolats (ribotypage, PCR-based restriction fragment length polymorphism (PCR-RFLP) analysis, random amplified polymorphism DNA (RAPD) analysis, séquençage d'un nombre limité de gènes de ménage). De plus, ces études ont été réalisées dans des familles vivant dans des pays développés et aucune n'avait été réalisée dans des pays en voie de développement.

Plusieurs études récentes ont essayé de prendre en compte ces paramètres. La diversité génétique de *H. pylori* a été étudiée de façon large au sein de l'estomac d'une famille dont les parents sont nés en Algérie et les quatre enfants en France (Raymond *et al.* 2004; Raymond *et al.* 2008). Pour chaque membre de la famille, 9 à 10 isolats présents au niveau de l'antrum et du fundus ont été caractérisés par séquençage de deux gènes de ménage (*glmM* et *hspA*). Plusieurs clones ont été retrouvés dans l'estomac de chacun des individus. L'étude de la diversité génétique globale montrait une circulation intrafamiliale des clones, entre parents et enfants mais aussi entre frères et sœurs. De façon intéressante, un enfant possédait sa propre souche non retrouvée chez les autres membres de la famille. Une autre étude a analysé la diversité génétique d'isolats de *H. pylori* dans deux familles de 52 individus vivant en milieu rural en Afrique du Sud et dans 11 familles de 43 individus vivant en milieu urbain au Royaume Uni, aux Etats-Unis, en Corée du Sud et en Colombie (Schwarz *et al.* 2008) par analyse de la diversité de sept gènes de ménage. La circulation intrafamiliale de souches était beaucoup plus fréquente en milieu urbain, suggérant une transmission entre parents et enfants ou entre frères et sœurs. Dans les familles sud-africaines, la diversité génétique était beaucoup plus importante et aucune similitude n'a été retrouvée entre les souches hébergées par les parents et les enfants, ni entre les souches présentes chez les individus vivant sous le même toit. Cependant, un seul isolat de l'antrum et du fundus ont été investigués par séquençage chez

la presque totalité des membres des familles, impliquant une probable sous-estimation du nombre d'infections mixtes par rapport à l'étude précédente. Le mode de transmission de *H. pylori* dans des zones de forte prévalence de l'infection tel que dans un milieu rural en Afrique du Sud, semblait être différent de celui rencontré dans des zones de faible ou moyenne prévalence de l'infection. La transmission par l'eau et les aliments, la promiscuité pourraient jouer un rôle plus important dans les pays en voie de développement, conséquence probable de différences dans le mode de vie et de culture. Récemment, la diversité génétique de dix isolats de *H. pylori* de l'antra et du fundus collectés dans trois familles, deux françaises et une algérienne, a été analysée par séquençage de deux gènes de ménage et par étude du contenu génétique à l'aide de membranes d'hybridation à haute densité (Raymond *et al.* 2008). Les infections mixtes étaient rares dans les familles françaises puisque tous les individus n'hébergeaient qu'une souche dans leur estomac. Dans la famille algérienne, plusieurs enfants présentaient des infections mixtes. Deux à cinq souches circulaient à l'intérieur d'une même famille et une même souche était retrouvée au moins chez deux membres d'une même famille, arguments en faveur d'un modèle de transmission intrafamiliale (notamment entre enfants et parents et entre frères et sœurs). De façon surprenante, la mère n'était pas impliquée dans la transmission dans les deux familles françaises.

Ces trois études ont apporté un éclairage nouveau sur le mode de transmission de l'infection : i) les infections mixtes semblent fréquentes dans les pays en voie de développement, ii) le mode de transmission intrafamiliale semble prépondérant dans les pays développés, notamment entre frères et sœurs, et iii) le mode de transmission horizontale semble prépondérant dans les pays en voie de développement.

### **3.3.3. Facteurs influençant la transmission de *H. pylori***

De nombreuses études à travers le monde ont montré qu'un niveau socio-économique bas est associé à une plus grande prévalence de l'infection à *H. pylori* et que le niveau socio-économique dans l'enfance est considéré comme déterminant dans le risque d'acquisition de l'infection [pour une revue (Frenck and Clemens 2003; Bruce and Maaros 2008; Azevedo *et al.* 2009; Cherian *et al.* 2009; Zhang *et al.* 2009; Ford and Axon 2010; Jafri *et al.* 2010)]. Un argument fort est une prévalence de l'infection plus forte chez le jumeau homozygote vivant dans le milieu le plus pauvre (Malaty *et al.* 1998).

Les principaux facteurs de risque sont la promiscuité, le manque d'hygiène, de l'eau de boisson souillée, un partage du lit pendant l'enfance. D'autres facteurs ont été également

identifiés comme un niveau d'éducation bas des parents (en particulier celui de la mère), le fait de ne pas allaiter au sein, des habitudes culturelles telles que la prémastication ou la partage de plats, le nombre de personnes dormant dans la même chambre, le nombre de frères et sœurs, un âge plus élevé.

#### **4. TRAITEMENT ET EPIDEMIOLOGIE DES RESISTANCES AUX ANTIBIOTIQUES**

Une des particularités de *H. pylori* est sa capacité à acquérir des résistances essentiellement par mutation. La transmission de la résistance est essentiellement verticale et les taux de résistance augmentent progressivement du fait de la pression de sélection. Tous les antibiotiques sont concernés bien que la résistance à l'amoxicilline soit exceptionnelle (Megraud 2007; Megraud and Lehours 2007; Boyanova and Mitov 2010; De Francesco *et al.* 2010).

L'activité bactéricide de l'amoxicilline est la conséquence de l'inhibition de la synthèse du peptidoglycane par liaison aux protéines liant la pénicilline (PLP). En 1996, une souche (souche Hardenberg) avec une résistance stable à l'amoxicilline (concentration minimale inhibitrice (CMI) = 8 mg/L) a été isolée chez un patient hollandais ayant reçu au préalable des injections d'amoxicilline par voie parentérale. Le remplacement du gène codant pour la PLP-1A d'une souche sensible par celui de la souche Hardenberg s'est traduit par une CMI de l'amoxicilline 100 fois plus élevée. Plusieurs substitutions d'acides aminés ont été décrites dans ce gène, mais seule la substitution Ser 414 Arg semble impliquée dans la résistance. La culture d'une souche sensible en présence de concentrations progressivement croissantes d'amoxicilline entraîne l'augmentation de la CMI de 0,03-0,06 mg/l à 4-8 mg/l. Cette résistance liée à une diminution de l'affinité de la PLP-1 pour l'amoxicilline est stable (Gerrits *et al.* 2002). Schiller *et al.* suggèrent que le haut niveau de résistance à l'amoxicilline (CMI > 0.5mg/L) pourrait être dû à une combinaison de substitutions d'acides aminés à la fois au niveau de la PLP-1A et au niveau de porines (Gerrits *et al.* 2006). Récemment, il a été montré que de multiples substitutions d'acides aminés au niveau de la PLP-2 et PLP-3, en plus de celles au niveau de la PLP-1, conféraient une résistance plus élevée à l'amoxicilline (Rimbara *et al.* 2008). En conclusion, bien que le mécanisme de résistance ne soit pas encore totalement élucidé, il est maintenant clair qu'il concerne principalement les PLP.

L'augmentation de la prescription de macrolides, notamment dans les pays développés pour le traitement des infections respiratoires, a eu pour conséquence une augmentation de la

résistance de *H. pylori* à la clarithromycine. La cible des macrolides est la boucle peptidyl transférase du domaine V de l'ARN ribosomique 23S avec un arrêt de l'élongation de la chaîne peptidique. Des mutations dans la boucle V de l'ARN ribosomique 23S sont responsables d'une diminution de la fixation des macrolides et donc de l'apparition d'une résistance (Versalovic *et al.* 1996). La plupart des souches résistantes présentent les mutations A2142G, A2143G et A2142C. D'autres mutations plus rares ont été identifiées comme les mutations T2717C et T2182C, mais leur impact sur la résistance est discuté (Fontana *et al.* 2002; Hao *et al.* 2004; Khan *et al.* 2004; Toracchio *et al.* 2004).

Une étude multicentrique européenne réalisée en 1998 dans 17 pays, avait montré un taux moyen de résistance primaire à la clarithromycine de 9,9 % (IC 95 % 0 - 10,8 %), plus élevé dans les pays d'Europe de l'est et surtout dans les pays du sud de l'Europe (Megraud *et al.* 1999). Ces résultats ont été ultérieurement confirmés (Megraud 2007; Megraud and Lehours 2007; Boyanova and Mitov 2010; De Francesco *et al.* 2010). Récemment, des taux de résistance primaire supérieurs à 15 % ont été observés dans plusieurs pays européens: Bulgarie (18.4 %), Italie (21.3 % et 16.9 %), Hongrie (17.3 %) et Pologne (15.0 %) (Boyanova and Mitov 2010), ce qui remet en question l'utilisation de cet antibiotique en traitement de première intention si l'on sait que la résistance à la clarithromycine diminue le taux d'éradication d'environ 70 % (Megraud 2004). Une étude récente sur 455 souches collectées en France entre 2004 et 2007 a montré un taux de résistance primaire de 19.1 %. Pendant cette période d'inclusion, aucune augmentation de la résistance à cet antibiotique n'a été observée, ce qui peut être expliqué par une prescription principalement dans les infections respiratoires hautes (Raymond *et al.* 2010). En dehors de l'Europe, la prévalence était inférieure à 4 % au Canada et de 10 à 15 % aux Etats-Unis (Fallone 2000; Meyer *et al.* 2002). En ce qui concerne l'Asie, des prévalences de 20 % ou plus ont été décrites au Japon et en Corée du sud (Lee *et al.* 2005; Kobayashi *et al.* 2007). Par ailleurs, il a été décrit des taux de résistance aussi élevés chez l'enfant que chez l'adulte (Boyanova and Mitov 2010).

L'étude de la résistance au métronidazole est peu fiable : la plupart des études ont montré une absence de reproductibilité des résultats inter et intralaboratoires (Glupczynski *et al.* 2002). Le travail réalisé par Goodwin *et al.* a permis de mieux comprendre les mécanismes impliqués dans la résistance aux 5-nitroimidazolés : la présence de mutations dans le gène d'une nitroréductase insensible à l'oxygène (*rdxA*) n'est pas suffisante pour conférer à une souche une résistance au métronidazole (Goodwin *et al.* 1998). Plus récemment, il a été montré que le mécanisme responsable de la résistance au métronidazole impliquerait d'autres réductases (FrxA, FdxB, Ribf, MdaB...) (Jeong *et al.* 2000; Kwon *et al.* 2000; Tankovic *et al.*

2000; Megraud 2004; Kaakoush *et al.* 2009). Ces données expliquent la difficulté à évaluer la résistance au métronidazole.

Néanmoins, la prévalence de la résistance au métronidazole catégorisée en basse (<30 %), moyenne (30 % - 50 %) ou haute (>50 %) semble fiable. Dans l'étude multicentrique européenne réalisée en 1998, la prévalence globale était de 33,1 % (IC 95 % 7,5 – 58,9), sans différence significative entre le Nord et le Sud, mais avec une prévalence significativement plus basse en Europe Centrale et Orientale (Megraud *et al.* 1999). Depuis, d'autres études individuelles réalisées en Europe ont mis en évidence une prévalence comprise entre 20 et 40 % (Cabrita *et al.* 2000; Toracchio and Marzio 2003; Elviss *et al.* 2004; Megraud 2007; Megraud and Lehours 2007). Des données récentes en France indiquent une augmentation de la résistance primaire avec un taux proche de 60 % (Raymond *et al.* 2010). Dans les pays en voie de développement où le métronidazole est très utilisé pour traiter certaines parasitoses, la prévalence de la résistance est très élevée (50-80 %) (Prazeres Magalhaes *et al.* 2002; Nahar *et al.* 2004; Datta *et al.* 2005; Seck *et al.* 2009).

Peu d'études ont porté sur la prévalence de la résistance aux fluoroquinolones. Les fluoroquinolones agissent au niveau de la sous-unité A de l'ADN gyrase qui est un tétramère composé de 2 sous-unités A et B, codées respectivement par les gènes *gyrA* et *gyrB*. De nombreuses mutations au niveau du gène *gyrA* ont été incriminées dans la résistance aux fluoroquinolones, elles concernaient presque toutes les 2 acides aminés suivants : Asn-87 et/ou Asp-91 (Moore *et al.* 1995; Wang *et al.* 2001; Kim *et al.* 2005; Miyachi *et al.* 2006; Kist 2007). La résistance est de l'ordre de 5 % dans la plupart des pays, des taux proches de 20 % ont été trouvés en Italie, au Portugal, en Corée, au Japon et en Belgique (Debets-Ossenkopp *et al.* 1999; Boyanova *et al.* 2000; Cabrita *et al.* 2000; Tankovic *et al.* 2003; Kim *et al.* 2004; Bogaerts *et al.* 2006; Miyachi *et al.* 2006; Zullo *et al.* 2007). En France, une augmentation significative du taux de résistance primaire a été observée entre la période 2004-2005 (7.3 %) et 2006-2007 (14.1 %) (Raymond *et al.* 2010). En France comme dans d'autres pays, cette augmentation pourrait refléter l'utilisation accrue de fluoroquinolones pour traiter de nombreux types d'infections communautaires (Malferttheiner *et al.* 2002).

La combinaison d'un inhibiteur de la pompe à protons à deux antibiotiques est proposée dans toutes les conférences de consensus. Le traitement de référence inclus la clarithromycine (1 g en 2 prises) et l'amoxicilline (2 g en 2 prises) associés à un inhibiteur de la pompe à protons à double dose pendant au moins 7 jours. Le problème majeur de ce traitement est la sélection de bactéries résistantes à la clarithromycine. Dans la conférence de consensus sur *Helicobacter pylori* Maastricht III, il est recommandé de ne pas utiliser cet



antibiotique et traitement probabiliste lorsque le taux de résistance est supérieur à 15-20 % (Malfertheiner *et al.* 2007). Etant donné la difficulté d'étude de la sensibilité aux antibiotiques par culture, le développement de nombreuses méthodes moléculaires permettant la détection de la résistance à la clarithromycine, mais également aux fluoroquinolones, est une réponse à ce défi, d'autant plus qu'elles peuvent être réalisées directement sur le prélèvement (Romano *et al.* 2003; Cambau *et al.* 2009). La complexité du mécanisme moléculaire responsable de la résistance au métronidazole n'a pas permis pour l'instant le développement de méthodes moléculaires de détection. En cas de résistance à la clarithromycine, l'alternative est le métronidazole ou les fluoroquinolones (lévofloxacine, moxifloxacine). D'autres antibiotiques sont utilisables comme la tétracycline et la rifabutine dont le taux de résistance est encore faible à ce jour. L'étude de la sensibilité aux antibiotiques et des mécanismes de résistance est donc plus que jamais à l'ordre du jour.

## **5. LES DETERMINANTS DES PATHOLOGIES ASSOCIEES A L'INFECTION PAR *H. PYLORI***

Dans le cadre des CG de type sporadique (qui constitue environ 98 % de tous les CG), il a été démontré que les sujets infectés par *H. pylori* ont un risque accru de développer un CG. La réponse inflammatoire de l'hôte serait un élément clef du processus menant vers le CG et serait influencée par des facteurs de prédisposition génétique de l'hôte, des facteurs environnementaux et la virulence de la bactérie.

### **5.1. Les polymorphismes génétiques de l'hôte**

La prédisposition génétique individuelle pourrait être le point central dans une variété de processus relatifs à la carcinogenèse gastrique: i) la réponse inflammatoire, qui régit le maintien, la sévérité et l'issue de l'infection par *H. pylori*, ii) la protection de la muqueuse gastrique face à l'infection par *H. pylori* et à d'autres carcinogènes, iii) la protection contre l'agression oxydative et d'autres inducteurs d'altération de l'ADN, iv) la capacité de désintoxication carcinogène et de protection antioxydante et v) la capacité de prolifération cellulaire (par l'activation des oncogènes ou l'inactivation des gènes tumoraux suppresseurs) (Gonzalez *et al.* 2002).

La cytokine interleukine (IL)-1 bêta (IL1  $\beta$ ) est l'un des inhibiteurs physiologiques les plus puissants de la sécrétion gastrique acide. Elle est ainsi la cytokine clé de la réponse inflammatoire de l'hôte à *H. pylori* et donc un candidat de premier plan dans les variations

génétiques associées à la prédisposition au CG. Certains polymorphismes du gène codant pour l'IL1  $\beta$  (Hwang *et al.* 2002) et du gène de l'antagoniste du récepteur de l'IL1 (IL1RA) qui fonctionne comme régulateur des taux d'IL1  $\beta$  (Machado *et al.* 2001), prédisposent à l'augmentation de l'expression de l'IL1  $\beta$ . Ils sont associés à un risque accru de développer une hypochlorhydrie et une gastrite atrophique en réponse à l'infection par *H. pylori*, conditions préalables au développement d'un CG par *H. pylori*. Il a été démontré que les effets combinés des génotypes pro-inflammatoires de l'IL1  $\beta$  et des facteurs de virulence bactériens de *H. pylori* (présence du gène *cagA*, génotype *vacA* s1/m1) conduisent à un risque nettement accru de développer un CG, illustrant le rôle important de l'interaction entre l'hôte et la bactérie (Figueiredo *et al.* 2002).

L'allèle TNFA-308\*A, qui est présumé accroître la production de facteur de nécrose tumorale alpha (TNF-alpha), a également été identifié comme facteur de risque accru de développement d'un CG (Duarte *et al.* 2005). L'effet combiné du polymorphisme génétique des gènes codant pour l'IL1  $\beta$ , l'IL1RA et le TNF-alpha sur la survenue d'un CG a été également étudié. Les individus porteurs de trois génotypes à haut risque sont exposés à un risque accru de CG avec un OR d'environ 10 (El-Omar *et al.* 2003; Machado *et al.* 2003). Bien que les cellules T helper de type 1 (Th-1) soient surtout recrutées par des pathogènes intracellulaires, les cellules T spécifiques de la muqueuse gastrique recrutées par *H. pylori* sont généralement du phénotype Th1. L'IL10 est une cytokine anti-inflammatoire répressive de l'expression des cytokines qui sont les premières concernées dans des vecteurs inflammatoires Th1 tels que l'IL1  $\beta$  et le TNF-alpha. Il a été rapporté que le génotype induisant une expression réduite de l'IL10 augmente significativement le risque de survenue d'un CG (El-Omar *et al.* 2003).

Le point commun des études portant sur la susceptibilité génétique de l'hôte est le manque de reproductibilité des résultats parmi les différentes populations, dû le plus souvent à une méthodologie inadéquate. Même en respectant tous les critères, les résultats peuvent diverger. Un des exemples le plus net est le polymorphisme de l'IL8-251\*T/\*A pour lequel une association significative avec le CG a été mise en évidence dans plusieurs études menées chez des populations asiatiques, association qui était absente dans des populations caucasiennes (Shirai *et al.* 2006; Canedo *et al.* 2008). Le terrain génétique différent selon les populations étudiées pourrait expliquer ces résultats discordants.

## 5.2. Facteurs environnementaux

Le terme « facteurs environnementaux » englobe les paramètres généraux inhérents au mode de vie [pour une revue (Gonzalez and Lopez-Carrillo 2010)]. Il semble qu'il existe une interaction positive entre le tabac, la consommation de viande rouge et l'infection à *H. pylori* dans la survenue d'un CG. Un effet protecteur potentiel des anti-oxydants tels que la vitamine C et E, le bêta-carotène semble exister chez les patients infectés par *H. pylori*, bien que pas toujours retrouvé. En Asie, une alimentation riche en sel (poissons fumés...) associée à une infection par *H. pylori* entraîne un risque plus élevé de CG. Cependant, il est très difficile d'évaluer l'impact de chacun des facteurs sur la survenue d'un CG chez des patients infectés par *H. pylori*: i) l'impact est mesuré en comparant des sujets infectés et non infectés par *H. pylori*, or une atrophie sévère ou une métaplasie intestinale diffuse peut éradiquer l'infection et entraîner une sous-estimation du nombre de cas associés à *H. pylori*, ii) le faible effectif, particulièrement le nombre de cas de CG non infectés par *H. pylori*, ne permet pas toujours de conclure à l'impact d'un facteur sur la survenue du CG, iii) la supposition que pour certains cas de CG, le processus carcinogène est totalement indépendant d'une infection à *H. pylori* est controversée. De larges études prospectives seront nécessaires pour conclure définitivement.

## 5.3. Facteurs de virulence bactériens

*H. pylori* est particulièrement bien adapté à la colonisation de la muqueuse gastrique grâce à de nombreux facteurs bactériens tels qu'une uréase et des flagelles. D'autres facteurs sont responsables d'une inflammation locale au site de colonisation tels que le lipopolysaccharide (LPS), des adhésines, un appareil de sécrétion de type IV (T4SS) et une cytotoxine VacA. L'intervention de ces différents facteurs bactériens amène à l'affaiblissement de la barrière épithéliale et à des lésions gastriques apparentes lors de l'observation par endoscopie de l'estomac (Figure 4). *H. pylori*, armé de son pouvoir colonisateur et de ses facteurs de virulence, entraîne l'apparition de gastrites qui peuvent être chroniques et asymptomatiques ou évoluer vers des pathologies plus graves.

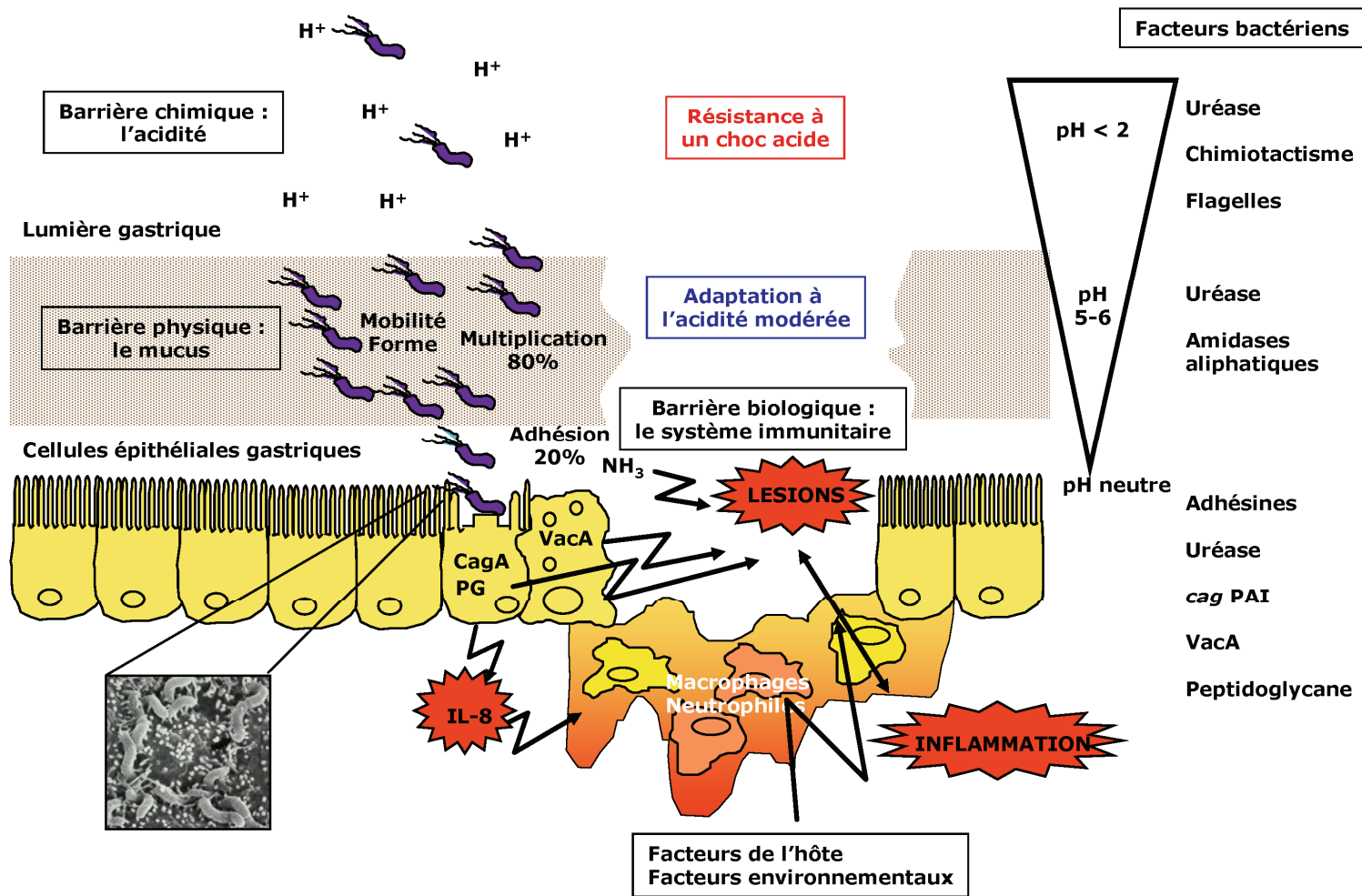


Figure 4. Colonisation de la muqueuse gastrique et genèse des lésions par *H. pylori*

### **5.3.1. Inflammation et lésions tissulaires**

#### **5.3.1.1. La cytotoxine vacuolisante VacA**

VacA est un facteur de virulence majeur chez *H. pylori*. La protéine VacA a la particularité d'être une cytotoxine multifonctionnelle, incluant entre autres une vacuolisation cellulaire, la formation de canaux membranaires, la perturbation des fonctions endosomales/lysosomales, l'apoptose et un effet immunosuppresseur (Figure 5). La capacité de la toxine à former des pores anioniques est une caractéristique importante de son mécanisme d'action [pour une revue (Isomoto *et al.* 2010)].

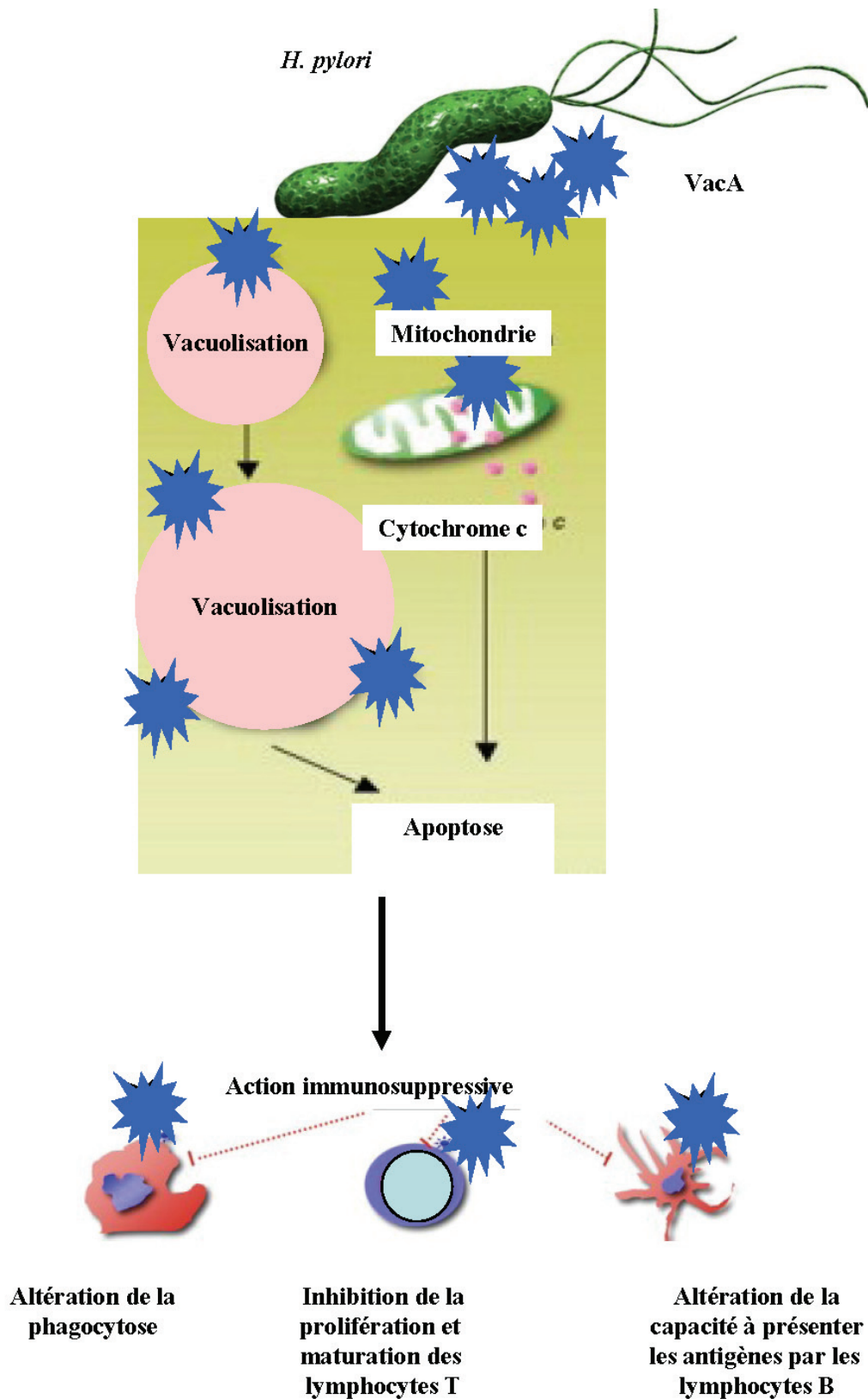


Figure 5. Effets pathologiques de VacA sur les cellules de l'hôte.

### 5.3.1.1.1. Génétique

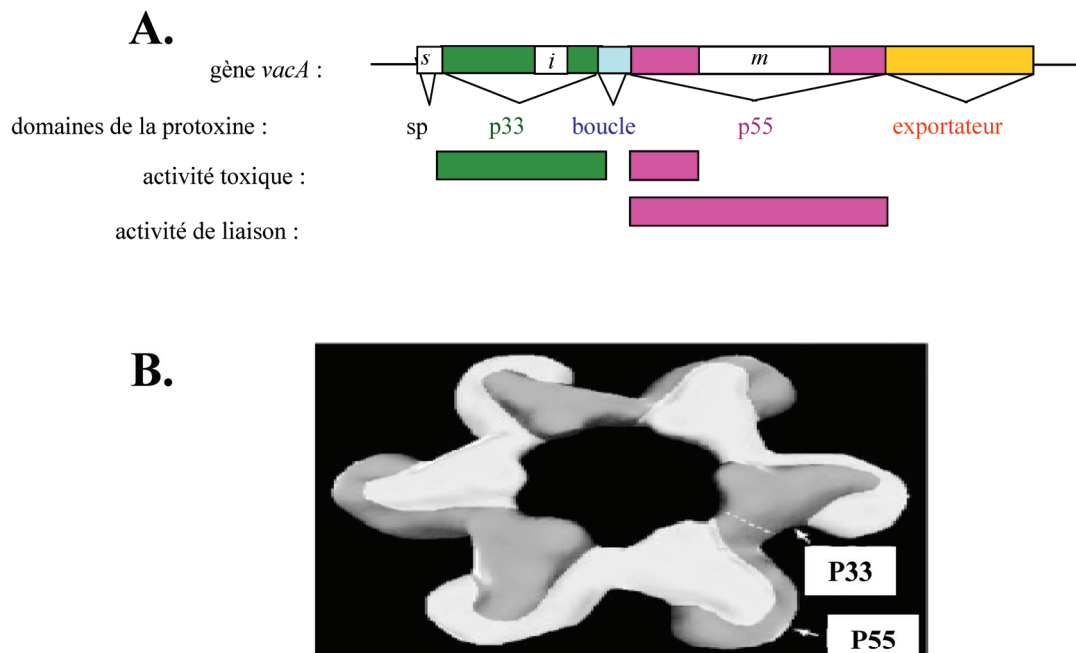
Bien que toutes les souches de *H. pylori* n'aient pas toutes une activité cytotoxique, elles possèdent toutes le gène *vacA* codant pour une cytotoxine vacuolisante VacA. Le gène *vacA* est sujet à un polymorphisme au niveau de : i) la région « s » codant pour la séquence signal de la protéine (allèle s1 fonctionnel, subdivisé en s1a, s1b, s1c et l'allèle non fonctionnel s2), ii) la région « i » (allèles i1 et i2) et iii) la région « m » située au milieu du gène (allèles m1 et m2 subdivisés en m2a et m2b) (Rhead *et al.* 2007) (Figure 6A). Une étude réalisée sur 735 souches provenant de 24 pays par Van Doorn *et al.* a montré en Europe un gradient du sous-type s1. Dans le nord et l'est de l'Europe, le sous-type s1a est majoritaire (89 %) tandis qu'en France et en Italie, le sous-type s1a et s1b est retrouvé de façon à peu près égal. En Espagne et au Portugal, le sous-type majoritaire est le sous-type s1b (89 %). En Asie du sud-est, le sous-type s1c a été retrouvé dans 77 % des souches (Japon, Thaïlande). Pour la région « m », les allèles m1 et m2a ont une distribution géographique proche, excepté dans la péninsule ibérique, l'Amérique Centrale et du Sud où m1 (86.2 %) est plus répandu que m2 (13.8 %). Le sous-type m2b a été trouvé exclusivement chez les souches du sud-est asiatique du sous-type s1c (Van Doorn *et al.* 1999). Les souches s1/m1 sont i1, s2/m2 sont i2 et s1/m2 peuvent être i1 ou i2 (Rhead *et al.* 2007; Basso *et al.* 2008), L'allèle s1/i2/m1 n'a jamais été décrit, suggérant que cette combinaison est défavorable à *H. pylori* en terme de fitness.

Ce mosaïcisme se traduit, *in vitro*, par des niveaux d'activités différents de la toxine. Les souches de type s1/i1/m1-ou-m2 induisent en général un haut niveau de vacuolisation tandis que les souches s1/i2/m2 induisent peu de vacuolisation, et les souches s2/i2-ou-i1/m1-ou-m2 sont dépourvues d'effet cytotoxique (Rhead *et al.* 2007). Les souches de type s1 ont été associées à un risque accru de survenue d'un ulcère peptique (Atherton *et al.* 1995; Rudi *et al.* 1998; Van Doorn *et al.* 1999) et les souches de type s1/m1 sont plus fréquemment associées à un CG par rapport à celles présentant les allèles s1/m2 ou s2 (Kidd *et al.* 1999; Miehle *et al.* 2000). De plus, le type i1 a été associé au CG chez des patients iraniens (Rhead *et al.* 2007) et était prédictif de l'ulcère peptique chez des patients italiens (Basso *et al.* 2008). Cependant, chez des patients vivant en Asie, il n'était pas associé à la survenue de pathologies gastro-duodénales graves (Ogiwara *et al.* 2008).

#### 5.3.1.1.2. Structure

Le gène *vacA* code pour une protéine précurseur de 140 kDa qui est maturée en une toxine d'environ 88 kDa après les clivages suivants : i) lors de sa traversée de la membrane interne via le système de sécrétion de type II (Sec-dépendant), une séquence signal de 33 acides aminés en N-terminal est clivée, ii) lors de sa traversée de la membrane externe, la partie C-terminale de 33 kDa de la protéine reste attachée à la membrane externe permettant l'export et la sécrétion de la toxine mature par un mécanisme de type auto-transport, et iii) la protéine mature peut subir un clivage au niveau d'une boucle sensible à la protéolyse entre le fragment N-terminal de 33 kDa (p33) et le fragment C terminal de 55 kDa (p55), les deux fragments restant physiquement associés (Cover and Blaser 1992; Schmitt and Haas 1994) (Figure 6A). Le domaine p33 présente une activité de formation de pores nécessaire pour la vacuolisation (Busler *et al.* 2006), tandis que le domaine p55 est responsable de la liaison à la cellule cible (Reyrat *et al.* 1999). Récemment, il a été montré que la partie N-terminale de p55 est essentielle pour l'induction de la vacuolisation et de la dépolarisation de la membrane cellulaire de l'hôte (Ivie *et al.* 2008). La toxine est libérée dans le milieu extracellulaire ou peut être retenue à la surface bactérienne (Telford *et al.* 1994). Les monomères de VacA peuvent s'agencer en une structure de plus de 900 kDa dite "en fleur", de 6 à 7 pétales constitués chacun par l'association des deux polypeptides de 33 et 55 kDa (monomère de 88 kDa) (Lupetti *et al.* 1996; Reyrat *et al.* 1999) (Figure 6B). La forme oligomérique de VacA présente une faible activité sauf lorsqu'elle est exposée à pH acide ou alcalin, pH auxquels la protéine se monomérise (Cover *et al.* 1997). La toxine de *H. pylori* possède donc des propriétés remarquables qui pourraient lui permettre d'optimiser dans le milieu gastrique son action délétère vis-à-vis de l'épithélium.





**Figure 6. Organisation génétique du gène *vacA* et structure tridimensionnelle de la protéine VacA**

A. Représentation schématique de l'organisation du gène *vacA* sujet à un polymorphisme au niveau de : i) la région « s » codant pour la séquence signal de la protéine (allèle s1 fonctionnel et l'allèle non fonctionnel s2), ii) la région « i » (allèles i1 et i2) et iii) la région « m » située au milieu du gène (allèles m1 et m2).

B. Modèle de l'agencement multimérique en forme de « fleur ». D'après Reyrat *et al.* (Reyrat *et al.* 1999).

### 5.3.1.1.3. Effets physiopathologiques

#### 5.3.1.1.3.1. Vacuolisation des cellules eucaryotes

Il a été observé que des surnageants de culture de *H. pylori* étaient capables d'induire *in vitro* la formation de larges vacuoles intracellulaires dans des cellules, activité qui a été attribuée ultérieurement à la protéine VacA (Leunk *et al.* 1988). L'utilisation de la protéine purifiée et activée par l'acidité, ou l'expression du gène *vacA* dans la cellule eucaryote, permet d'induire la vacuolisation en présence de bases faibles, telle que le chlorure d'ammonium (Cover and Blaser 1992). La microinjection de VacA ou la transfection de plasmides contenant le gène *vacA* dans des cellules HeLA entraîne la formation de vacuoles

intracellulaires, montrant que la protéine agit sur une cible intracellulaire (de Bernard *et al.* 1997). VacA est capable d'agir dans la cellule eucaryote sous forme d'oligomères ou de monomères. L'administration intragastrique chez la souris d'une dose massive de la protéine VacA purifiée permet d'observer des dommages de la muqueuse gastrique ainsi qu'une inflammation [pour une revue (Cover and Blanke 2005)]. Cependant, cette dose de VacA ne correspond pas à des conditions physiologiques pouvant être rencontrées chez l'homme.

La toxine VacA forme des pores dans des bi-couches lipidiques et sur des membranes plasmiques : des monomères de VacA, activés à pH acide et insérés dans une bi-couche lipidique, se réassocient au niveau membranaire pour former un pore hexamérique impliqué dans un transport d'anions (Iwamoto *et al.* 1999; Szabo *et al.* 1999). Après endocytose des pores de surface, ils sont traités au niveau des compartiments endosomaux tardifs, qui subissent un gonflement osmotique pour devenir de larges vacuoles (Papini *et al.* 1994). Les vacuoles induites par VacA portent les marqueurs des endosomes tardifs et des lysosomes et ne sont pas létales pour les cellules. Cependant, la protéine VacA engendre des dégâts au niveau du compartiment endosomal tardif (Cover and Blanke 2005):

- dégradation du facteur de croissance épidermique (EGF pour « epidermal growth factor »),
- inhibition de la maturation de la procathepsine D,
- inhibition de la présentation d'antigènes.

#### **5.3.1.1.3.2. Induction de l'apoptose**

La protéine VacA engendre d'autres phénomènes cytotoxiques tels que l'induction de l'apoptose. Lorsque des cellules en culture sont traitées avec la protéine VacA purifiée, il a été observé une réduction du potentiel transmembranaire mitochondrial et la libération de cytochrome c (Kimura *et al.* 1999; Willhite *et al.* 2003; Willhite and Blanke 2004). Ces deux effets sont caractéristiques d'un affaiblissement de la perméabilité membranaire mitochondriale et suggèrent l'initiation d'une mort cellulaire. Des fusions de la partie N-terminale de VacA avec un fluorochrome ont permis de mettre en évidence la localisation mitochondriale de VacA (Galmiche *et al.* 2000).

De récents travaux suggèrent une action de VacA au niveau de certaines protéines appartenant à la famille Bcl-2. Cette famille se subdivise en 2 sous-groupes, selon leur fonction apoptotique ou anti-apoptotique. En réponse à VacA, l'activation de Bak et Bax, protéines pro-apoptotiques, est responsable de la libération de cytochrome c et de l'apoptose cellulaire. En parallèle, l'activation de la caspase 3, étape clef de l'apoptose, et le clivage de la

poly (ADP-ribose) polymérase renforceraient le processus d'apoptose. Les principales protéines anti-apoptotiques sont Bcl-2, Bcl-XL et Mcl-1. VacA semble réprimer l'expression de ces protéines via le facteur de transcription STAT3 [pour une revue (Isomoto *et al.* 2010)].

#### 5.3.1.1.3.3. Action immunosuppressive

Elle est suggérée par : i) une altération dans la maturation du phagosome, mais la survie de *H. pylori* ne semble pas modifiée que le gène *vacA* soit exprimé ou non (Zheng and Jones 2003), ii) une altération de la capacité à présenter les antigènes par les lymphocytes B (Molinari *et al.* 1998), iii) une inhibition de la prolifération et de la maturation des lymphocytes T (Gebert *et al.* 2003; Sundrud *et al.* 2004; Torres *et al.* 2007) et B (Torres *et al.* 2007). L'activité immunosuppressive observée de VacA a conduit certains auteurs à proposer pour cette toxine un rôle dans la persistance de *H. pylori* dans l'estomac.

#### 5.3.1.1.3.4. Autres effets de la protéine VacA

D'autres effets de la protéine VacA ont été observés :

- Tuo *et al.* ont montré que VacA inhibait la sécrétion de bicarbonates au niveau de la muqueuse duodénale induite par la prostaglandine E2 en stimulant le relargage d'histamines, et donc pourrait par ce mécanisme contribuer au développement de l'ulcère duodéal (Tuo *et al.* 2009).
- affaiblissement des jonctions intercellulaires (de Bernard *et al.* 2005) qui permettrait le passage de nutriments pour la bactérie vers la lumière gastrique.
- stimulation de la sécrétion de cytokines pro-inflammatoires, IL1  $\beta$ , TNF-alpha, IL6, IL10, IL13 et cyclo-oxygénase 2 (Supajatura *et al.* 2002; de Bernard *et al.* 2005). Récemment, Hisatsune a montré que VacA augmentait la production et la sécrétion d'IL8, un important médiateur dans le processus pathologique (Hisatsune *et al.* 2008).

Malgré l'abondance d'effets cytotoxiques observés *in vitro*, le rôle de VacA *in vivo* n'est pas clarifié. L'inoculation par voie orogastrique de la toxine VacA purifiée ou de souches de *H. pylori* produisant cette protéine à des souris, provoquait des lésions de l'épithélium gastrique chez ces animaux (Ghiara *et al.* 1995). Des études similaires réalisées dans d'autres modèles animaux (porcelets et gerbilles) n'ont pas permis de confirmer le rôle de VacA dans la genèse de lésions épithéliales. Bien que le gène *vacA* ne soit pas nécessaire pour la colonisation dans un modèle animal, une souche produisant une cytotoxine fonctionnelle

possède un avantage compétitif sur une souche dépourvue de cytotoxine VacA (Salama *et al.* 2001).

### 5.3.1.2. L'îlot de pathogénicité *cag* et la protéine CagA

Les souches de *H. pylori* ont été classées en deux catégories : celles qui possèdent un îlot de pathogénicité *cag* (*cag* PAI) complet et fonctionnel et celles qui en sont totalement ou partiellement dépourvues (Covacci *et al.* 1999). Il semble que le *cag* PAI a été probablement importé dans le génome de *H. pylori* par transfert horizontal d'une autre espèce après la spéciation, tôt dans l'évolution de l'espèce (Gressmann *et al.* 2005; Olbermann *et al.* 2010).

Le *cag* PAI code pour un T4SS. Il est typiquement composé de l'assemblage de 11 protéines (VirB 1 à 11) couplées à une NTPase (VirD4) (Backert and Selbach 2008). Il est accepté que la protéine CagF joue un rôle de chaperonne dans la translocation de la protéine « cytotoxin associated antigen A » (CagA) (Couturier *et al.* 2006; Pattis *et al.* 2007) et que la protéine CagL assure la spécificité du TSS4 pour les cellules hôtes en se liant aux intégrines cellulaires de type  $\alpha 5\beta 1$  (Kwok *et al.* 2007). Les travaux de Wiedemann ont permis de mieux préciser le rôle du T4SS dans l'évolution de la maladie gastro-duodénale en utilisant un modèle animal (gerbille) sur une période longue (2-64 semaines) (Wiedemann *et al.* 2009). Le T4SS est essentiel pour l'induction d'une inflammation précoce et sévère du corpus, caractérisée par une augmentation de l'expression de cytokines pro-inflammatoires ainsi que des changements histologiques avec la survenue d'une atrophie gastrique et d'une métaplasie intestinale. En occident, de nombreuses études montrent une corrélation entre une infection par des bactéries possédant un *cag* PAI fonctionnel et la survenue d'ulcères peptiques ou de CG (Huang *et al.* 2003) ; ces corrélations ne sont toutefois pas retrouvées pour les souches asiatiques qui possèdent quasiment toutes un *cag* PAI, qu'elles soient ou non associées à une pathologie sévère (Yamaoka *et al.* 1999).

#### 5.3.1.2.1. Organisation génétique de l'îlot

Le *cag* PAI est une région génomique de 40 kb contenant une trentaine de gènes. Le *cag* PAI est inséré dans le gène chromosomique *glr* codant la glutamate racémase et présente les caractéristiques d'un îlot de pathogénicité : i) une grande taille (environ 40 kb, 31 gènes putatifs), ii) un contenu en GC de 35 % différent de celui du reste du génome (39 %), iii) la présence de séquences répétées de 31 paires de base (pb) aux extrémités, iv) une présence variable de séquences d'insertion, v) des homologues à certains facteurs de virulence, et vi) une instabilité génétique (Censini *et al.* 1996).

#### **5.3.1.2.2. Activités biologiques**

Le T4SS de *H. pylori*, en contact étroit avec les cellules épithéliales, est impliqué dans plusieurs processus majeurs : (i) un phénomène d'élongation cellulaire également appelé phénotype colibri, de par les extensions cellulaires que l'on peut observer lors d'une infection de cellules épithéliales *in vitro* avec une souche de *H. pylori*, résultat d'une augmentation de la mobilité et de l'élongation cellulaire (ii) une activité pro-inflammatoire, pro-apoptotique et mitogène, et (iii) une perturbation des jonctions serrées et adhérentielles, et de la polarité cellulaire [pour une revue (Backert and Selbach 2008)] (Figure 7).

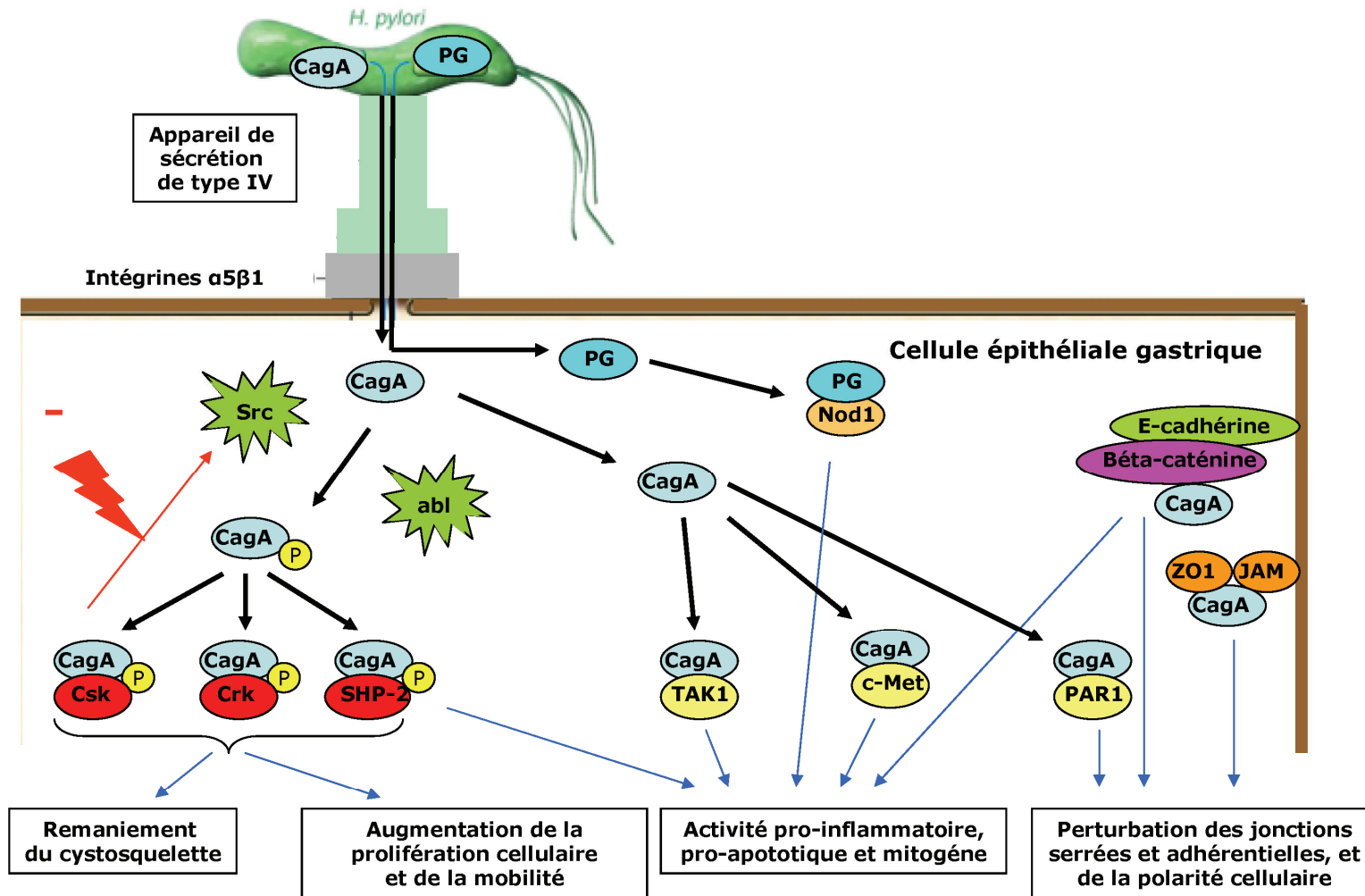
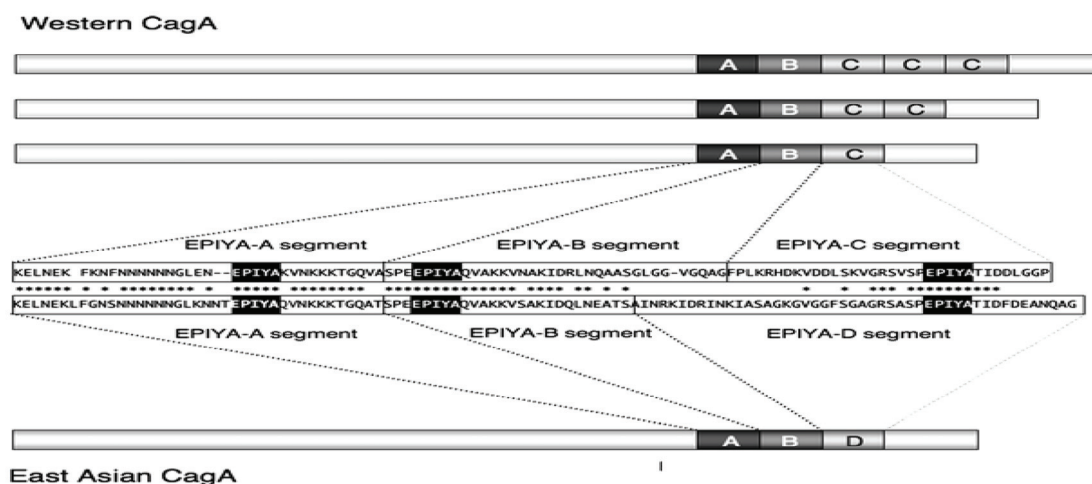


Figure 7. Modèle du fonctionnement de l'appareil de sécrétion de type IV de *H. pylori*

### 5.3.1.2.2.1. Effets de CagA sous forme phosphorylée

La protéine CagA est injectée via le T4SS dans les cellules épithéliales gastriques mais aussi dans les cellules phagocytaires. Après sa translocation, CagA est phosphorylée sur des résidus tyrosines présents au niveau de motifs Glu-Pro-Ile-Tyr-Ala (EPIYA) sur l'extrémité C-terminale de la protéine par des kinases cellulaires de la famille Src et Abl (Higashi *et al.* 2002). De façon intéressante, *H. pylori* contrôlerait leur activité: Src est activée durant la phase initiale de l'infection (0.5-2 heures) avant d'être rapidement inactivée tandis que Abl est activée de façon continue mais de manière plus importante dans les phases plus tardives de l'infection (2-8 heures). Ces éléments sont en faveur d'un modèle de phosphorylations successives de CagA par Src et Abl (Tammer *et al.* 2007). Selon la séquence peptidique entourant le motif EPIYA, quatre types de segments ont été définis, EPIYA-A, EPIYA-B, EPIYA-C et EPIYA-D, chacun d'eux contenant un motif EPIYA. Les segments EPIYA-A et EPIYA-B sont quasiment toujours présents quelque soit l'origine géographique des patients chez qui les souches ont été isolées. Le segment EPIYA-C est caractéristique des souches d'Afrique, d'Amérique et d'Europe (« Western CagA ») tandis que le motif EPIYA-D est présent quasiment uniquement chez les souches isolées de patients vivant dans l'est de l'Asie (« East Asian CagA »). Le motif EPIYA est très conservé et seulement une petite fraction des souches présente un motif modifié tel que EPIYT, ESIYA ou ESIYT (Xia *et al.* 2009). De façon intéressante, le segment EPIYA-C varie en nombre (1 à 3 motifs le plus souvent) tandis que le segment EPIYA-D est en général unique (Figure 8).



**Figure 8. Diversité génétique de l'extrémité C-terminale de la protéine CagA**  
D'après Hatakeyama (Hatakeyama 2009)

CagA sous sa forme phosphorylée (CagA-P) interagit alors avec des facteurs de l'hôte contenant un domaine SH2 (Src homology 2) comme la tyrosine phosphatase SHP-2, la tyrosine kinase Csk et la protéine adaptatrice Crk (Higashi *et al.* 2002; Tsutsumi *et al.* 2003; Suzuki *et al.* 2005). La combinaison de ces trois facteurs est responsable de remaniement du cytosquelette et d'une augmentation de la mobilité cellulaire. D'autres protéines ont été identifiées comme pouvant interagir avec CagA-P par protéomique quantitative, P13k, SHP-1, Ras-Gap1 et Grb7 (Selbach *et al.* 2009) mais les conséquences physiopathologiques de ces interactions sont inconnues. CagA-P a été montré comme inactivant Src directement ou indirectement via le recrutement de Csk, un régulateur négatif de Src (Selbach *et al.* 2003; Tsutsumi *et al.* 2003). Comme Src intervient dans la phosphorylation de CagA, l'inhibition de Src par CagA-P génère un rétrocontrôle négatif qui va réguler la quantité intracellulaire de CagA-P, et donc son activité. La vinculine, la cortactine et l'ezine sont des substrats de Src et son inactivation par CagA-P est responsable de la déphosphorylation de ces protéines, étape essentielle dans le remaniement du cytosquelette. Enfin, l'interaction de CagA-P avec CrkII peut activer Rac1 via Dock180 et entraîner notamment des réarrangements du cytosquelette (Suzuki *et al.* 2005). L'analyse d'une série de mutants EPIYA révèle que SHP-2 se fixe spécifiquement aux segments EPIYA-C ou -D phosphorylés (Higashi *et al.* 2002). La séquence flanquant le site de phosphorylation tyrosine du segment EPIYA-D se lie de façon parfaite à la séquence consensus du domaine SH2 de SHP-2 tandis que celle flanquant le site de phosphorylation tyrosine du segment EPIYA-C diffère de la séquence consensus par un seul acide aminé à la position pY+5. De ce fait, le « East Asian CagA » présente une affinité plus importante pour le domaine SH2 que le « Western CagA » (Higashi *et al.* 2002; Naito *et al.* 2006) et à l'intérieur des souches présentant un « Western CagA », plus le nombre de motifs EPIYA-C est important, plus l'interaction avec SHP-2 est importante (Hatakeyama 2009). Il a été montré que les souches présentant un « East Asian CagA » induisaient une sécrétion plus importante d'IL8 *in vitro* sur cellules AGS que les souches avec un « Western CagA » et que plus le nombre de motifs EPIYA-C chez ces dernières était important, plus la sécrétion d'IL8 était importante (Argent *et al.* 2008) (Tableau 1). Dans un modèle de souris transgénique, le « East Asian CagA » a été décrit comme plus carcinogène que le « Western CagA ». Cependant, des études ultérieures sont nécessaires pour confirmer ce résultat puisque des tumeurs s'étaient également développées dans des organes autres que l'estomac (Miura *et al.* 2009). Récemment, une méta-analyse a montré que les souches avec 2 segments EPIYA-C étaient significativement associées à la survenue d'un CG (Xia *et al.* 2009).



De façon physiologique, les protéines de la famille Gab et la protéine phosphatase SHP-2 jouent un rôle clef dans l'activation de la voie des Ras/MAP kinases. La stimulation de récepteurs à activité tyrosine kinase tels que le récepteur du facteur de croissance hépatocyte (HGF pour « hepatocyte growth factor ») et le récepteur à l'EGF, mène à l'activation des protéines de la famille Gab après phosphorylation sur des résidus tyrosine qui vont interagir avec le domaine SH2 de SHP-2, laquelle mène à l'activation de la cascade des Ras/MAP kinases. Cette voie est impliquée dans le contrôle de nombreux processus fondamentaux tels que la prolifération cellulaire, la différenciation et la mobilité cellulaire. De façon surprenante, l'HGF est capable d'induire des modifications morphologiques des cellules épithéliales ressemblant à l'allongement en "bec de colibri" induit par CagA (Segal *et al.* 1999). CagA mime la fonction des protéines de la famille Gab bien qu'elle ne possède aucune similarité d'un point de vue séquence et active de façon aberrante SHP-2 (Hatakeyama 2003). La liaison de CagA-P au domaine SH2 entraîne un changement de conformation de SHP-2 qui empêche l'inhibition du domaine catalytique portant l'activité phosphotyrosine phosphatase par le domaine SH2 à l'extrémité N-terminale (Higashi *et al.* 2002) d'où un emballement de la voie des Ras/MAP kinases. L'inhibition de la phosphorylation de CagA, la perturbation de l'interaction CagA-SHP-2, l'inactivation de l'expression de SHP-2 abolissent l'induction de l'allongement en "bec de colibri" des cellules épithéliales par la protéine CagA (Higashi *et al.* 2002; Higashi *et al.* 2002; Higashi *et al.* 2005). La dérégulation de SHP-2 est potentiellement d'une grande importance dans le processus de transformation cellulaire puisque des mutations du gène *PTPN11* codant pour SHP-2 ont été identifiées dans de nombreuses formes de tumeurs, principalement les leucémies (Tartaglia *et al.* 2003).

**Tableau 1. Impact de la diversité génétique de l'extrémité C-terminale de la protéine CagA**

CagA	Nombre de segments EPIYA				Sécrétion IL8	Remaniement du cytosquelette	Association avec le CG
	-A	-B	-C	-D			
Western	Absent	Absent	Absent	Absent	-	+	+-
	1	1	1	Absent	++	++	++
	1	1	2	Absent	+++	+++	+++
	1	1	3	Absent	++++	++++	++++
East Asian	1	1	Absent	1	+++++	+++++	+++++

#### **5.3.1.2.2.2. Effets de CagA sous forme non phosphorylée**

Sous forme non phosphorylée, CagA peut aussi induire des réponses cellulaires de l'hôte telles qu'une perturbation des jonctions serrées et adhérentielles, une perte de polarité cellulaire et une activité mitogène et pro-inflammatoire [pour une revue (Backert and Selbach 2008)]. CagA s'associe avec ZO (zonula occludens)-1 et JAM (junctional adhesion molecule) et altère l'organisation et la fonction des jonctions serrées (Amieva *et al.* 2003). De plus, CagA inhibe l'activité kinase de la partitioning-defective 1 (PAR1), membre de la famille des « microtubules affinity regulating kinases » (MARK), entraînant une perturbation des jonctions serrées et de la polarité cellulaire des cellules épithéliales (Saadat *et al.* 2007; Lu *et al.* 2009; Umeda *et al.* 2009). La formation du complexe CagA/PAR1 augmente la stabilité de la protéine CagA dans la cellule épithéliale (Ishikawa *et al.* 2009). Cette interaction CagA-PAR1 est médiée par le motif multimérique (CM) présent au niveau de l'extrémité C-terminale de CagA. Il a été montré que le motif CM des souches possédant un « East-Asian CagA » se lie plus fortement à PAR1 que ne le fait celui des souches possédant un « Western CagA », et que dans ces dernières la capacité de liaison à PAR1 est proportionnelle au nombre de motifs CM (Lu *et al.* 2008). CagA interagit également avec la protéine d'adhésion intercellulaire E-cadhérine de façon dépendante du motif CM, détériorant le complexe E-cadhérine-béta-caténine et activant de façon aberrante la bêta-caténine (Murata-Kamiya *et al.* 2007; Kurashima *et al.* 2008). La conséquence de l'ensemble de ces processus est une déstabilisation de l'épithélium gastrique qui potentiellement contribue au processus carcinogène de *H. pylori in vivo*. Récemment, Lamb *et al.* ont démontré que CagA s'associait physiquement à la protéine kinase TAK1 (« transforming growth factor  $\beta$  activated kinase 1 »), entraînant l'activation de TAK1 par ubiquitination en Lys63 médiée par la protéine TRAF6 (« tumour necrosis factor factor 6-mediated receptor associated »). La conséquence est l'activation du complexe IKK permettant la phosphorylation d'I $\kappa$ B qui va se décrocher du facteur de transcription nucléaire kappa B (NF-kappa B). Ce dernier peut alors activer la transcription des gènes codant pour l'IL8 et le TNF-alpha (Lamb *et al.* 2009). De façon intéressante, l'inoculation intragastrique de CAPE (« caffeic acid phenetyl ester »), un inhibiteur du NF-kappa B, à des gerbilles infectées par *H. pylori* diminuait significativement l'infiltration inflammatoire de la muqueuse gastrique et le relargage de médiateurs de l'inflammation (Toyoda *et al.* 2009).

### **5.3.1.2.2.3. Effets dépendants de l'appareil de sécrétion de type IV mais indépendants de CagA**

Certains gènes codant pour des composants du T4SS sont requis pour l'induction du signal pro-inflammatoire (Tummuru *et al.* 1995; Censini *et al.* 1996), incluant l'activation du facteur nucléaire NF-kappa B, la protéine activatrice 1 et enfin l'expression des proto-oncogènes c-fos et c-jun (Sharma *et al.* 1998; Meyer-ter-Vehn *et al.* 2000). L'IL8 est une cytokine majeure pour le recrutement de cellules immunitaires au niveau de la muqueuse gastrique. L'activation de NF-kappa B engendrant la production d'IL8 est dépendante de la protéine Nod1 (Viala, Chaput *et al.* 2004). Cette protéine Nod1 est responsable de la détection intracellulaire de petits fragments de peptidoglycane (PG) générés par la bactérie dans les cellules eucaryotes. Le PG de *H. pylori* est transloqué dans les cellules épithéliales de façon dépendante de la présence d'un T4SS fonctionnel.

### **5.3.1.3. L'inflammation**

*H. pylori* induit une infection chronique de la muqueuse gastrique qui conduit à un état inflammatoire chronique. Le *cag* PAI induit la synthèse par les cellules épithéliales gastriques d'IL8, une puissante cytokine pro-inflammatoire (Partie I. Généralités : 5.3.1.2. *L'Ilot de pathogénicité cag et la protéine CagA*). Les polymorphismes génétiques de l'hôte influencent directement les variations inter-individuelles au niveau des taux de production de cytokines anti- et pro-inflammatoires et contribuent de façon évidente à l'issue de l'évolution clinique individuelle (cf *Partie I. Généralités : 5.1. Les polymorphismes génétiques de l'hôte*). *H. pylori* induit chez l'homme une forte réponse immunitaire cellulaire et humorale. La réponse humorale semble inefficace pour entraîner une guérison de l'infection (Ferrero *et al.* 1998). Elle ne semble pas non plus impliquée dans la protection contre l'infection dans un modèle murin (Ermak *et al.* 1998). En revanche, la réponse cellulaire joue un rôle majeur dans l'élimination de *H. pylori* et dans la pathogenèse chez l'animal. La réponse cellulaire prédominante chez l'homme induite par *H. pylori* est une réponse de type Th1 et est associée à la libération de cytokines pro-inflammatoires (Interferon-gamma (IFN- $\gamma$ ), IL12, IL18, TNF-alpha) ainsi qu'à une activation des macrophages (D'Elisio *et al.* 1997; Ernst and Gold 2000), contribuant aux dommages tissulaires. D'autres cytokines jouent également un rôle important, l'IL17, l'IL21 et l'IL23 dans le fin équilibre entre protection et lésions cellulaires liées à *H. pylori* (Luzza *et al.* 2000; Amedei *et al.* 2006; Cooper 2009). Des souris infectées par *H. pylori* avec une forte réponse Th1 développent une intense gastrite mais ont une faible charge bactérienne au niveau de la muqueuse gastrique, tandis que l'inverse est observé en présence

d'une réponse de type Th2 (Smythies *et al.* 2000). L'impact important de la réponse cellulaire acquise dans la pathogenèse est mis en évidence par l'observation d'une atrophie gastrique induite par l'INF- $\gamma$  (cytokine de type Th1) chez la souris, même en l'absence d'une infection par *H. pylori*, bien que celle-ci potentialise son effet (Atherton and Blaser 2009).

Pour persister, la bactérie se doit d'échapper à la réponse immunitaire de l'hôte. Sa persistance est essentielle pour sa pathogénicité puisque le CG survient des décennies après l'infection. *H. pylori* produit les protéines CagA et VacA, qui ont une activité immunosuppressive (cf *Partie I. Généralités : 5.3.1.1. La cytotoxine vacuolisante VacA et 5.3.1.2. L'Ilot de pathogénicité cag et la protéine CagA*). Sa localisation majoritairement au niveau du mucus gastrique rend difficile une action efficace du système immunitaire. Les lymphocytes T régulateurs (Tregs) jouent un rôle essentiel dans le contrôle de la réponse immune cellulaire. Ils sont capables de supprimer la réponse cellulaire de type T par contact cellulaire et par la production de cytokines immunosuppressives de type IL10 et le facteur de croissance transformant bêta (TGF- $\beta$  pour « transforming growth factor beta »). Les Tregs ont été mis en évidence au niveau de la muqueuse gastrique de patients infectés par *H. pylori*, avec des concentrations élevées d'IL10 et de TGF- $\beta$  (Lundgren *et al.* 2005), suggérant un effet immunosuppresseur. De plus, les Tregs suppriment la sécrétion d'IL8 par les cellules épithéliales ainsi que la réponse cellulaire mémoire dirigée vers *H. pylori* (Lundgren *et al.* 2003). Une souris génétiquement modifiée pour ne pas sécréter de l'IL10 développera une gastrite de forte intensité suite à une infection par *H. pylori*, puis éliminera le pathogène (Atherton and Blaser 2009). Des patients avec un ulcère peptique ont moins de Tregs au niveau de la muqueuse gastrique que les patients sans ulcère peptique et infectés par *H. pylori* (Robinson *et al.* 2008). Pris collectivement, ces résultats montrent l'importance des Tregs dans la persistance de l'infection et le contrôle de l'inflammation et qu'une régulation du système immunitaire inappropriée peut contribuer au développement de la maladie.

### **5.3.2. Adaptation à l'environnement gastrique**

Les cellules épithéliales gastriques secrètent le mucus qui forme une couche épaisse de protection (100  $\mu\text{m}$ ) pour la muqueuse gastrique vis-à-vis de l'acidité gastrique. Il est produit par les cellules cardiales, pyloriques et fundiques (cellules du collet). Lors de son entrée dans l'estomac, *H. pylori* est confronté à l'acidité extrême de la lumière gastrique qu'il fuit rapidement à l'aide de puissants flagelles, vers le mucus. L'analyse du transcriptome à pH acide montre une augmentation de l'expression de certains gènes impliqués dans la mobilité et le chimiotactisme (Bury-Mone *et al.* 2004). Le pH à la surface de la barrière est de 1.5 alors

qu'il est de 7 en zone profonde au contact des cellules. Ce gradient de pH permettrait à la bactérie de s'orienter dans la couche de mucus et d'atteindre son site de colonisation (Schreiber *et al.* 2004). La répartition de *H. pylori* dans l'habitat gastrique est la suivante : une infime proportion des bactéries se situe dans la lumière, 20 % à la surface de l'épithélium et environ 80 % dans le mucus dans une zone d'acidité modérée (pH 4.5-6) (Schreiber *et al.* 1999; Schreiber *et al.* 2000).

### 5.3.2.1. Les flagelles et le chimiotactisme

La mobilité de *H. pylori* est un facteur indispensable à la colonisation de la muqueuse gastrique par la bactérie. Des mutants aflagellés de *H. pylori* sont incapables de persister dans la muqueuse gastrique de porcelets mais génèrent une réponse immunitaire témoignant de leur survie transitoire dans l'estomac (Eaton *et al.* 1992). *H. pylori* possède 5 à 7 flagelles polaires engainés de 3 µm de long et une morphologie spiralée qui lui permettent, lors de son ingestion, d'abrèger son séjour dans le suc gastrique, de pénétrer dans la couche de mucus et de s'y mouvoir. La mobilité de *H. pylori* dans le mucus gastrique est assurée par la machinerie flagellaire dont la fonctionnalité requiert l'expression d'une quarantaine de gènes identifiés dans le génome de *H. pylori* (Tomb *et al.* 1997).

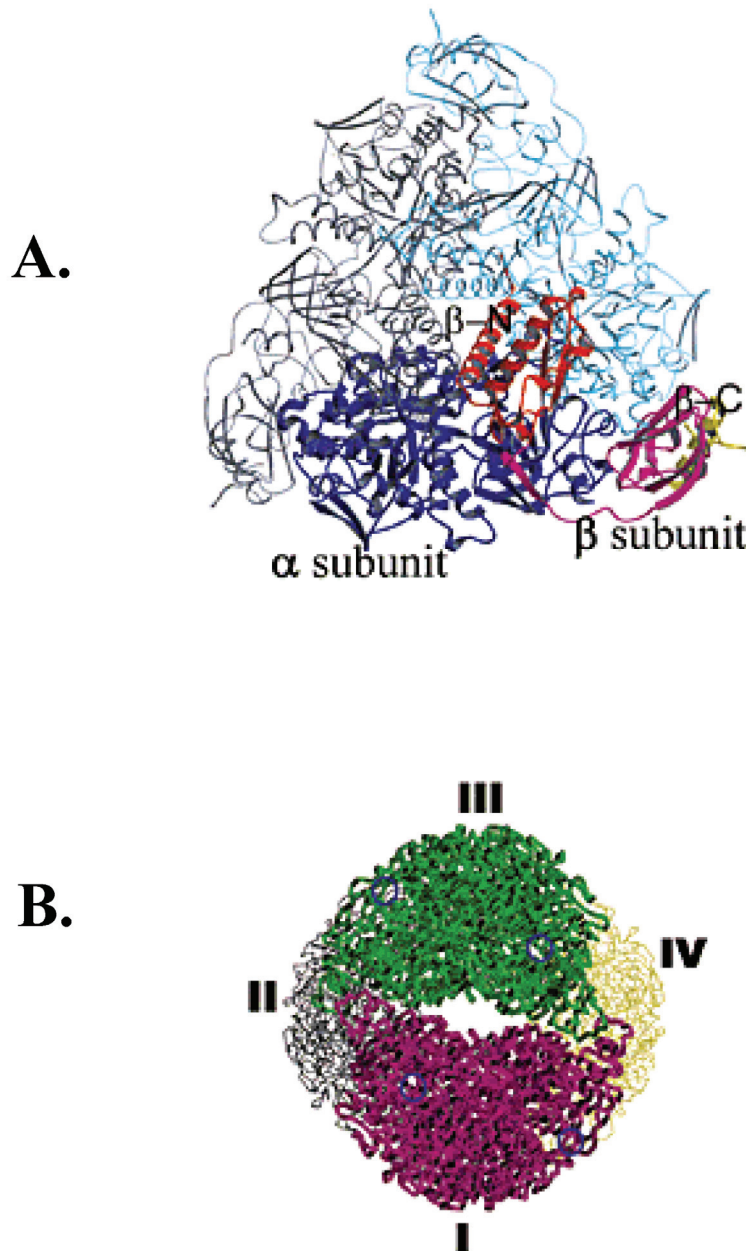
L'appareil flagellaire se décompose en trois éléments : i) un corps basal qui sert d'ancrage à la membrane bactérienne et porte les constituants du moteur flagellaire et certains éléments de chimiotactisme, ii) le crochet (constitué par la polymérisation de la protéine FlgE) qui relie le corps basal aux filaments flagellaires, et iii) les filaments flagellaires. Les filaments flagellaires de *H. pylori* sont constitués par la polymérisation de deux sous-unités d'environ 53 kDa : la flagelline majeure FlaA (codée par *flaA*) et la flagelline mineure FlaB (codée par *flaB*), retrouvées uniquement à la base du filament (Kostrzynska *et al.* 1991). Des expériences *in vivo* et *in vitro* d'inactivation des gènes *flaA* ou *flaB* montrent qu'ils sont tous les deux nécessaires à la mobilité de la bactérie (Josenhans *et al.* 1995; Eaton *et al.* 1996). Le gène *fliD* qui code pour une protéine homologue HAP-2 (famille des protéines associées au crochet pour « hook-associated proteins ») intervient dans la morphogénèse des filaments flagellaires et plus spécifiquement dans l'élongation. Une souche dont le gène *fliD* a été inactivé est incapable de coloniser la muqueuse gastrique de la souris (Kim *et al.* 1999). L'inoculation de souches de *H. pylori* dont le gène *flgk* (codant pour une protéine de la famille des protéines associées au crochet) a été inactivé ou l'inoculation d'une souche sauvage après l'immunisation de la souris par la protéine recombinante Flgk diminue la capacité de colonisation de la muqueuse gastrique par cette bactérie, suggérant que le flagelle et ses

constituants sont des cibles potentielles pour le développement de vaccins (Wu *et al.* 2006). Une souche de *H. pylori* inactivée dans le gène *motB* dont le produit fait partie du moteur flagellaire a été contruite. Elle possède des flagelles mais ne présente aucune mobilité et colonise la muqueuse gastrique murine avec une efficacité moindre que la souche parentale et présente un désavantage sélectif lors d'une compétition avec cette souche (Ottemann and Lowenthal 2002). L'analyse du transcriptome à pH acide montre une augmentation de l'expression de certains gènes impliqués dans la mobilité et le chimiotactisme (Merrell *et al.* 2003; Wen *et al.* 2003; Bury-Mone *et al.* 2004), suggérant la nécessité pour la bactérie de fuir rapidement l'acidité extrême pour un environnement plus favorable.

### 5.3.2.2. L'uréase

*H. pylori* exprime une métalloenzyme à nickel de localisation cytoplasmique, l'uréase, produite en quantité abondante (6 à 10 % des protéines totales). Elle hydrolyse l'urée présente dans l'environnement gastrique en permettant la production d'ammoniac et de dioxyde de carbone. L'ammoniac produit par l'uréase capte un proton pour former l'ammonium, ce qui permet à la bactérie de maintenir son pH intracellulaire proche de la neutralité. La production d'ammonium participerait également à l'endommagement de la barrière épithéliale gastrique afin de récupérer des nutriments et des ions essentiels (Smoot *et al.* 1990).

L'uréase est codée par un opéron bicistronique comprenant les gènes de structures de l'apoenzyme (*ureA* et *ureB*) et cinq gènes (*ureIEFGH*) dont les produits participent à l'activation de l'apoenzyme par incorporation des ions nickel. Des données de cristallographie montrent que l'uréase de *H. pylori* forme un complexe dodécamérique  $[(UreAUreB)^3]^4$  contenant 12 sous-unités catalytiques (UreB) fixant chacune 2 ions de nickel (24 ions/molécule) et formant un complexe sphérique extrêmement compact, conférant à l'enzyme une résistance à l'acidité (Ha *et al.* 2001) (Figure 9).



**Figure 9. Structure tridimensionnelle de l'uréase de *H. pylori***

D'après Ha *et al.* (Ha *et al.* 2001)

A. Unité trimérique de dimères des protéines UreA et UreB - Les dimères (UreA-UreB) ont des couleurs différentes. L'un des dimères est représenté par une sous-unité UreB bleue et une sous-unité UreA dont l'extrémité N terminale est rouge, l'extrémité C terminale est magenta et l'extension responsable de la formation des tétramères est jaune.

B. Structure supramoléculaire  $[(\text{UreA-UreB})^3]^4$  - Les quatre trimères  $(\text{UreA-UreB})^3$  sont représentés par des couleurs différentes.

Une des caractéristiques uniques de l'opéron uréasique de *H. pylori* est liée à la présence du gène *ureI* codant une protéine membranaire qui forme au niveau de la membrane cytoplasmique un pore à urée qui s'ouvre à pH acide permettant le transport efficace de l'urée et donc une adaptation rapide à un environnement acide (Weeks *et al.* 2000). L'uréase et la protéine *ureI* sont essentiels pour la colonisation de la muqueuse gastrique de différents modèles animaux (Ferrero *et al.* 1992; Eaton and Krakowka 1994; Skouloubris *et al.* 1998; Bury-Mone *et al.* 2001; Mollenhauer-Rektorschek *et al.* 2002).

### 5.3.2.3. La réponse à l'acidité

L'influence du pH acide sur la transcription des gènes, reflet de la dynamique de la réponse adaptative de la bactérie à son environnement, a été étudiée en détail par des analyses de transcriptome (Merrell *et al.* 2003; Wen *et al.* 2003; Pflock *et al.* 2006). Cependant, la complexité de la réponse *in vivo* à des variations de pH est difficile à transposer *in vitro*, expliquant probablement les différences qualitatives et quantitatives retrouvées entre les différentes études: environ 100 à 280 gènes selon les études ont une expression modifiée.

La réponse à l'acidité est caractérisée par la mise en œuvre de mécanismes de protection [pour une revue (de Reuse and Bereswill 2007)]: i) augmentation de l'expression de certains gènes impliqués dans la mobilité et le chimiotactisme (cf *Partie I. Généralités : 5.3.2.1. Les flagelles et le chimiotactisme*), ii) maintien de son pH intracellulaire par l'augmentation de la transcription de gènes codant pour des enzymes générant l'ammoniac, l'uréase (UreA) et les amidases aliphatiques (AmiE et AmiF), iii) diminution de la transcription de gènes codant pour des protéines transmembranaires telles que des transporteurs, perméases ou des protéines de la membrane externe (OMP pour « outer membrane protein ») entraînant une diminution des influx transmembranaires (protons, ions sodiques), iv) protection contre l'accumulation d'ions métalliques, principalement le nickel et le fer (à pH acide, leur solubilité est augmentée. L'expression des protéines impliquées dans le stockage est augmentée et l'expression des protéines impliquées dans l'import de nickel, NixA, et du fer, FecA, est diminuée), et v) augmentation de l'expression de la protéine VacA et de l'adhésine SabA à pH neutre, ce qui optimiserait leurs actions à proximité de la cellule gastrique.

Par rapport à d'autres bactéries, *H. pylori* possède un nombre restreint de régulateurs transcriptionnels (Tomb *et al.* 1997; Alm *et al.* 1999). Au moins, trois régulateurs transcriptionnels ont été identifiés comme impliqués dans la réponse adaptative à l'acidité : NikR, Fur et ArsRS [pour une revue (de Reuse and Bereswill 2007)]. Ces protéines présentent

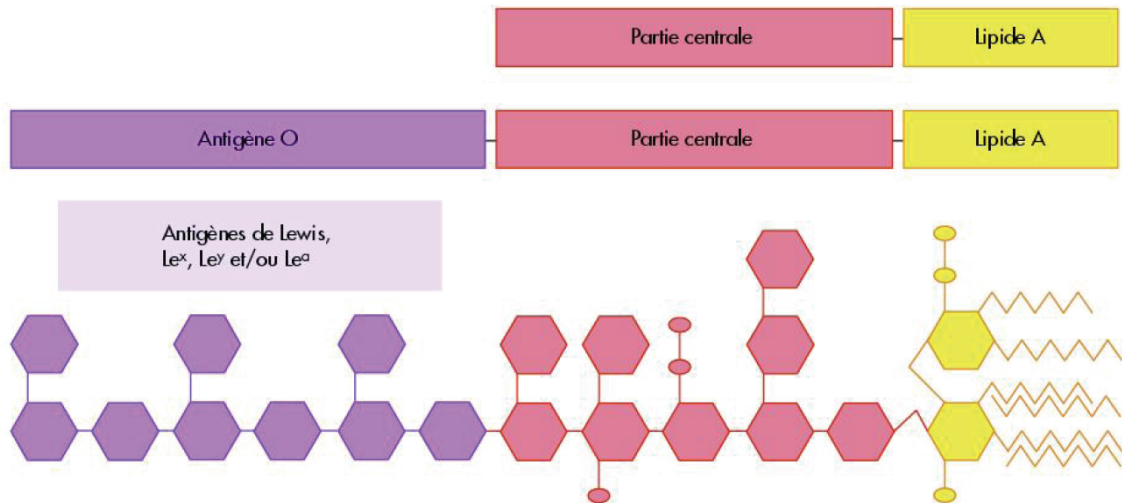


la particularité d'être des régulateurs pleiotropiques en agissant sur des catégories fonctionnelles de gènes très diverses. Les protéines NikR et Fur sont deux métallo-régulateurs sensibles respectivement à la concentration intracellulaire en nickel et en fer. Une analyse de transcriptome d'un double mutant  $\Delta nikR-\Delta fur$  indique une modification de l'expression de 36 gènes à pH acide contre 105 pour la souche sauvage, montrant le rôle central joué par ces deux régulateurs dans la réponse à l'acidité. Fur et NikR sont tous les deux importants pour une colonisation optimale de la muqueuse gastrique de la souris. Le système à deux composants ArsRS chez *H. pylori* est composé des protéines ArsS et ArsR. Un pH acide induirait un signal responsable de l'autophosphorylation de la protéine kinase ArsS. Le groupement phosphate est ensuite transféré à la protéine ArsR, laquelle régule l'expression de gènes cibles. Des études de mutagenèse ont montré qu'ArsS est indispensable à la colonisation de la souris et qu'ArsR est essentielle à la croissance *in vitro* de *H. pylori*. Cependant, une souche exprimant une protéine ArsR dont le site de phosphorylation a été muté est viable, suggérant que ArsR sous sa forme non phosphorylée joue également un rôle important. Les études de transcriptome mettent en évidence l'existence de réseaux de régulation, d'autant plus complexes qu'il semble exister une hiérarchie entre ces régulateurs et un chevauchement entre les régulons qu'ils définissent. A pH acide, l'expression de NikR augmente, celles de Fur et ArsR diminuent. NikR contrôle négativement l'expression de Fur et vice-versa. ArsR ne régule pas l'expression de NikR et Fur mais Fur réprime l'expression de ArsR en réponse au fer.

#### 5.3.2.4. Le lipopolysaccharide

Le LPS est un constituant de la paroi qui est ancré dans la membrane externe par sa partie lipidique, appelée lipide A. A ce lipide A, est greffée une partie centrale courte composée de sucres. Cette partie centrale fait elle-même le lien entre le lipide A et de longues chaînes glucosidiques formant l'antigène O. Il a été observé que jusqu'à 85 % des souches de *H. pylori* sont porteuses d'antigènes Lewis qui correspondent à l'antigène O (Simoons-Smit *et al.* 1996; Wirth *et al.* 1996). Une particularité unique est la présence au niveau de l'antigène O de motifs antigéniques de type Lewis,  $Le^x$ ,  $Le^y$  dans 80 à 90 % des souches ou plus rarement  $Le^a$ ,  $Le^b$ ,  $Le^c$  ou  $Le^x$ -sialylé (Figure 10). Les antigènes Lewis sont normalement présents au niveau de la surface des cellules eucaryotes telles que les érythrocytes et les cellules épithéliales de la muqueuse gastrique. La synthèse des antigènes Lewis est sous la dépendance de trois fucosyl-transférases, FutA, FutB et FutC, leur expression peut-être modulé par un mécanisme de variation de phase, créant une population microbienne avec des

profils de glycosylation du LPS très divers. De fait, il a été proposé que le LPS de *H. pylori* puisse jouer un rôle dans l'échappement de la bactérie au système immunitaire pour (Moran and Prendergast 2001).



**Figure 10. Structure du lipopolysaccharide de *H. pylori***

D'après Chaput et Bonecca (Chaput and Bonecca 2006)

Le LPS confère certaines activités biologiques et/ou immunologiques aux bactéries à Gram négatif telles que les entérobactéries, dont le LPS a été bien étudié. Il a été constaté que le LPS de *H. pylori* est peu immunogène, car les patients infectés par la bactérie ne présentent pas ou très peu d'anticorps dirigés contre ce constituant. De plus, le lipide A de *H. pylori* montre une composition en acide gras inhabituelle (Moran *et al.* 1997). Cette particularité est d'ailleurs associée à une diminution de ses activités pyrogènes, mitogènes et endotoxiques : il est 1000 fois moins pyrogène et mitogène que le lipide A de *Salmonella typhimurium*. De plus, la dose létale du LPS de *H. pylori* chez la souris est supérieure à celle du LPS de *S. typhimurium*, d'un facteur 500 (Muotiala *et al.* 1992).

### 5.3.2.5. Les adhésines

*H. pylori* a de nombreux moyens d'adhésion à la muqueuse gastrique, ce qui la protège du péristaltisme et de la desquamation de la muqueuse gastrique. Au regard des sept génomes séquencés (Thiberge *et al.* 2010), 46 OMPs ont été prédites. Cette catégorie de OMPs regroupe plusieurs adhésines dont un petit nombre a été caractérisé : AlpA, AlpB, BabA, BabB, HopZ, HopQ, SabA. Les adhésines BabA (« blood group antigen-binding adhesin ») et

SabA (« sialic acid-binding adhesin ») lie respectivement, les antigènes Lewis du type b (Le<sup>b</sup>) et les antigènes Lewis portant un acide sialique (sialyl-Le<sup>x</sup>). Il est d'ailleurs intéressant de noter que l'infection par *H. pylori* de souris transgéniques porteuses de Le<sup>b</sup> humain conduit à une inflammation plus sévère que celle de souris non transgéniques (Covacci *et al.* 1999). Ceci illustre l'importance de l'adhésion bactérienne dans le processus d'induction de la réponse immunitaire. De façon intéressante, une augmentation de sialyl-Le<sup>x</sup> était corrélée à une persistance de la colonisation de l'estomac de jeunes souris infectées de façon précoce (Yang *et al.* 2008). Le locus *sabA* a une expression soumise à une variation de phase. Ce système permet à la bactérie d'échapper à la réponse immunitaire de l'hôte, expliquant sa persistance et son adaptation à l'hôte (Goodwin *et al.* 2008). Des modifications de l'expression de BabA ont été également mises en évidence (Styer *et al.* 2010). Deux autres adhésines sont connues, AlpA et AlpB, mais leurs ligands sont inconnus. Un double mutant alpA/alpB n'a plus la capacité d'adhérer à la muqueuse gastrique (Odenbreit *et al.* 1999). De simples mutants alpA ou alpB sont affectés dans leur capacité à coloniser l'estomac du cobaye (de Jonge *et al.* 2004). Pour les protéines HopZ et OipA, leurs ligands sont inconnus. Cependant, un simple mutant est désavantagé lors d'une compétition avec la souche parentale pour la colonisation de l'estomac de cobaye (de Jonge *et al.* 2004). Des mutations de *hopQ* ont augmenté l'adhérence de *H. pylori* aux cellules épithéliales gastriques, la translocation de CagA dans la cellule hôte et des altérations cellulaires, démontrant l'importance de HopQ pour l'adhérence bactérienne. L'allèle *hopQ* type I était associé à une atrophie plus importante que celle générée par le type II et a été proposé comme marqueur de survenue d'une pathologie gastro-duodénale sévère (Ohno *et al.* 2009).

L'expression de la protéine OipA est associée à une production d'IL8 accrue par les cellules épithéliales, à une inflammation plus sévère et à une infiltration neutrophile plus importante. Elle est également associée à la survenue du CG et de l'ulcère duodéal (Yamaoka *et al.* 2000; Yamaoka *et al.* 2006).

### 5.3.3. Instabilité chromosomique et altérations épigénétiques

L'instabilité chromosomique du génome de l'hôte après infection par *H. pylori* a été investiguée par Machado *et al.* (Machado *et al.* 2010). Une diminution de l'activité et de l'expression des systèmes de réparation par excision de base et de réparation des mésappariements a été mise en évidence *in vivo* chez la souris et *in vitro* sur cellule AGS, contribuant ainsi au processus de carcinogénèse.

L'épigénétique est définie comme « l'étude des modifications de l'expression des gènes qui sont transmissibles lors de la mitose et/ou la méiose, mais qui ne découlent pas de modifications dans la séquence de l'ADN » (Wolffe and Matzke 1999). Les dinucléotides CpG sont répartis de façon non uniforme dans le génome, avec une sous-représentation générale, mais un enrichissement dans de courtes régions, appelées îlots CpG, positionnées au niveau du promoteur et/ou du premier exon de plus de 60 % des gènes humains (Bird 2002). Dans les cellules cancéreuses (où l'hyperméthylation de certains gènes s'inscrit de façon paradoxale dans un contexte d'hypométhylation globale du génome de ces cellules), l'hyperméthylation des CpG entraîne une extinction transcriptionnelle par modification épigénétique de gènes suppresseurs de tumeur impliqués dans certaines fonctions essentielles pour la cellule, comme le contrôle du cycle cellulaire, la réparation des dommages de l'ADN et l'apoptose, à l'exception notable de la méthylation de l'ADN (Deltour *et al.* 2005). L'hyperméthylation des CpG est considérée comme un facteur important de carcinogénèse et a été décrite chez les patients infectés par *H. pylori* (Chan *et al.* 2003; Maekita *et al.* 2006; Park *et al.* 2009). Il semble que le niveau de méthylation aberrant des CpG augmente lors de la progression séquentielle des lésions histologiques vers le CG chez les patients infectés par *H. pylori* (Park *et al.* 2009). Cependant, ce type d'étude est difficile à mettre en place puisque l'impact est mesuré en comparant des sujets infectés et non infectés par *H. pylori* or une atrophie sévère ou une métaplasie intestinale diffuse peut éradiquer l'infection et entraîner une sous-estimation du nombre de cas associés à *H. pylori*. Récemment, Schneider *et al.* ont montré que l'infection par une souche *cagA* positive et de type *vacA* s1m1 était significativement associée à des niveaux plus élevés de méthylation aberrante de certains gènes suppresseurs de tumeur (Schneider *et al.* 2010).

## **6. STRUCTURE ET DYNAMIQUE DU GENOME DE *H. PYLORI***

### **6.1. Espèce procaryote**

#### **6.1.1. Définition**

Chez les procaryotes, l'espèce est caractérisée par la convergence de propriétés génotypiques et phénotypiques :

- L'espèce englobe l'ensemble des souches ayant des similitudes ADN-ADN qui se traduisent à la fois par des valeurs d'hybridation supérieures ou égales à 70 % et par une valeur  $\Delta T_m$  inférieure ou égale à 5 °C.

- La plupart des souches d'une même espèce doivent posséder un certain nombre de propriétés phénotypiques communes.
- La plupart des espèces non cultivables ne peuvent être assignées à des espèces classiques en raison de l'absence de connaissance de leur phénotype.
- Un taxon non cultivable peut être assigné à la catégorie « Candidatus » i) si les séquences de l'ARN 16S sont suffisamment différentes de celles des espèces classiques, ii) si l'hybridation *in situ* peut être utilisée pour les détecter de façon spécifique, iii) et si une description basique de leurs morphologie et biologie existe.

### **6.1.2. Hybridation ADN-ADN**

Bien qu'introduites depuis plusieurs décennies, les données de l'hybridation ADN-ADN (HAA) sont cohérentes avec celles récentes issues du séquençage complet de génomes et du séquençage de multiples gènes de ménage. Cependant, il s'agit de techniques lourdes et délicates à réaliser, difficiles à standardiser et proposées par très peu de laboratoires. De plus, cette méthode de classification présente des limites : si les valeurs d'hybridation sont supérieures à 70 % entre la souche A et B et entre la souche B et C, elles ne sont pas nécessairement supérieures à 70 % entre la souche A et C, la conséquence en est un assignement inconstant ou ambigu de souches dans une espèce commune.

### **6.1.3. Séquençage des ARNr ou séquençage des gènes codant pour les ARNr**

Les ARNr (ARN ribosomiaux) ont été choisis dans le cadre d'études taxonomiques pour trois raisons:

- Ils sont présents dans toutes les cellules, ce qui permet des comparaisons entre procaryotes et eucaryotes.
- Ils présentent une structure conservée car toute modification importante peut avoir des conséquences sur les synthèses protéiques. Il existe d'ailleurs des portions d'ARNr dont la séquence est identique chez tous les êtres vivants.
- Ils sont abondants dans la cellule et faciles à purifier.

Les ARNr s'associent à des protéines pour former les ribosomes qui, chez les procaryotes, sont constitués d'une sous-unité 30S et d'une sous-unité 50S. La sous-unité 30S d'un ribosome contient de l'ARNr 16S et la sous-unité 50S contient de l'ARNr 5S et de l'ARNr 23S. Les ribosomes sont des organites responsables de la traduction des ARNm en

protéines dont la fonction est essentielle à la vie. En taxonomie, on utilise le plus souvent le séquençage de l'ARNr 16S ou le séquençage des gènes codant pour les ARNr (ADNr 16S).

Cependant, cette méthode manque de pouvoir discriminant par rapport à l'HAA. Des isolats avec une similarité de séquence des gènes codant pour les ARNr de moins de 97 % partagent des valeurs d'HAA inférieures ou égales à 70 % et appartiennent à des espèces différentes. A l'inverse, des isolats présentant une similarité de séquence de plus de 97 % peuvent ou non partager des valeurs d'HAA de plus de 70 % et ainsi appartenir ou non à la même espèce. Des bactéries dont l'homologie de séquence de l'ARNr 16S est inférieure ou égale à 98.7 % appartiennent toujours à une espèce différente. Cependant, le contraire n'est pas forcément vrai.

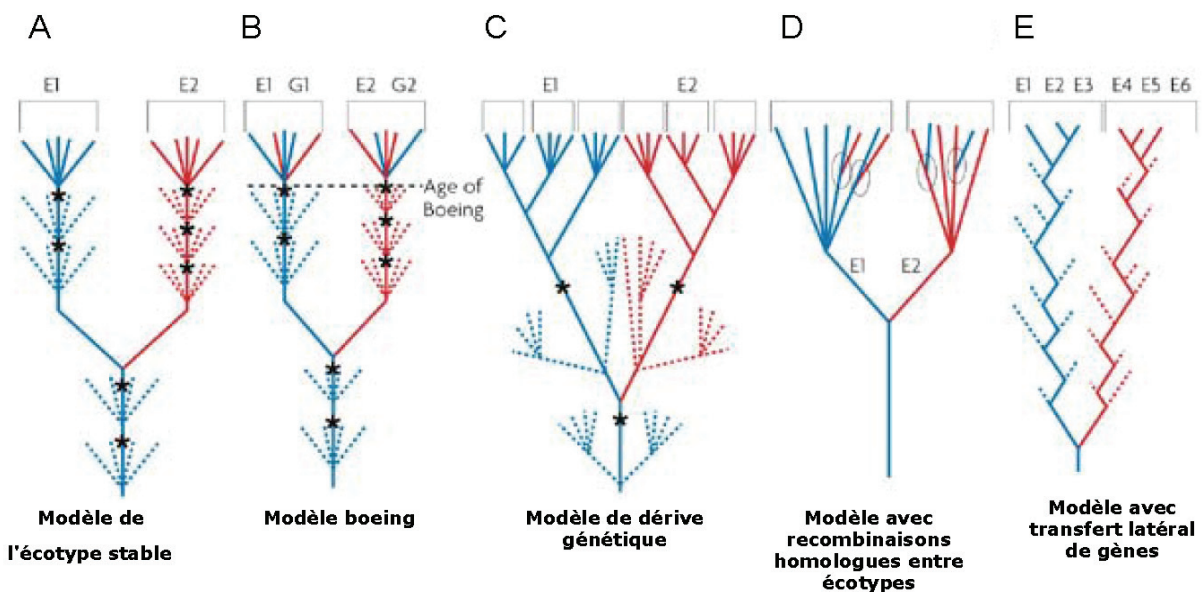
#### **6.1.4. Méthodes alternatives**

Pour être validées, les nouvelles méthodes de définition d'une espèce microbienne doivent donner des résultats cohérents avec l'HAA (Achtman and Wagner 2008). Le séquençage d'un « gène de ménage » (« housekeeping gene », gène qui assure des fonctions indispensables au métabolisme de base) ou de plusieurs "gènes de ménage" (MLSA ou « MultiLocus Sequence Analysis ») s'est avéré être une très bonne méthode pour classer des espèces au sein de genres bactériens. De nombreuses espèces bactériennes semblent composer des populations génétiques différentes, suggérant que la MLSA pourrait permettre de délimiter l'étendue d'une espèce bactérienne. Une autre alternative pourrait être basée sur l'étude de la distance génétique entre deux génomes en se basant sur l'ANI (« average nucleotide identity ») des gènes orthologues communs aux deux génomes. Une étude portant sur 28 espèces montre que des valeurs d'hybridation ADN-ADN supérieur à 70 % correspondent à une ANI de 95 %, suggérant que cette valeur pourrait être un cut-off pour définir une espèce bactérienne (Achtman and Wagner 2008).

#### **6.1.5. Concept d'espèces écologiques**

La notion d'écotypes proposée par Cohan *et al.* reprend le concept d'espèces écologiques et le développe (Cohan and Perry 2007). Les niches écologiques spécifiques joueraient dans le monde bactérien le rôle sélectif que l'impossibilité de se reproduire opère dans les règnes animal et végétal. Selon cet auteur, les écotypes sont des groupes d'individus occupant la même niche écologique et susceptibles d'être l'objet de sélection périodique (modèle de « l'écotype stable ») (Figure 11A). Les individus d'un écotype différent ne sont pas sensibles aux mêmes événements de sélection périodique. La sélection périodique a pour

conséquences d'augmenter la divergence entre écotypes et de limiter la diversité intra-écotype. Ce mécanisme, qui purge régulièrement la diversité de l'écotype, est une force de cohésion qui permet à un écotype de conserver son identité génétique. Avec le temps la diversité inter-écotype est plus importante que la diversité intra-écotype et les écotypes forment des entités. A terme, les différents écotypes sont suffisamment divergents pour qu'il s'établisse entre eux une barrière génétique. Ainsi, ce concept d'écotype permet à la fois d'appréhender les causes de la spéciation (les conditions écologiques), les forces de cohésions garantant de l'intégrité de l'espèce (la sélection périodique) et les conséquences de la spéciation bactérienne (l'isolement sexuel). Des scénarios alternatifs ont été proposés. Dans le modèle « boeing », des populations géographiquement isolées peuvent diverger suffisamment pour donner des populations génétiques différentes (ou géotypes). Cependant, la diffusion géographique de ces populations à travers l'accélération des flux migratoires depuis 500 ans a pu brouiller ces limites génétiques, aboutissant à la présence de plusieurs populations génétiques appartenant au même écotype dans une même région du monde (Figure 11B). Dans le modèle de « dérive génétique », la diversité génétique peut persister au lieu d'être éliminée par la sélection périodique, notamment quand la taille de la population est petite (Figure 11C). La recombinaison homologue entre écotypes peut ralentir l'élimination de la diversité génétique (Figure 11D) et le transfert latéral de gènes peut faciliter l'émergence continue de nouveaux écotypes avec une concomitante extinction d'autres écotypes « rivaux » (Figure 11E).



**Figure 11. Modèles théoriques de structure de la population bactérienne**

D'après Achtman et Wagner (Achtman and Wagner 2008)

Ces modèles sont toutefois contestés par Achtman et Wagner qui prônent une définition plus large de l'espèce, indépendante d'une méthode d'analyse et d'un critère universel (Achtman and Wagner 2008). Par exemple, toutes les souches de *Salmonella enterica* subsp. *enterica* serovar Thyphi, agent de la typhoïde, descendent d'un ancêtre commun qui a infecté l'homme entre 10 000 et 50 000 ans avant notre ère. Depuis cette période, de nombreux génotypes sont apparus mais contrairement à ce que le principe de sélection périodique laisserait présager, la plupart des anciens génotypes sont toujours présents avec une diversité génétique formant un continuum et non des « gaps ». De plus, alors que des mutants résistants à l'acide nalidixique ont été sélectionnés par l'usage intensif des quinolones en thérapeutique humaine, ils cohabitent avec des souches sensibles à l'acide nalidixique, ce qui va à l'encontre de la sélection périodique (Roumagnac *et al.* 2006). Ce modèle illustre notre manque de compréhension sur la rapidité de purge de la diversité génétique et le manque d'observations de modèles présentant une purge complète. *Yersinia pestis*, l'agent de la peste, illustre d'autres limites de la notion d'écotype proposée par Cohan. *Y. pestis* est génétiquement très proche de *Y. pseudotuberculosis* et ne formerait pas une espèce différente selon les critères exposés précédemment pour la définir (cf *Partie I. Généralités : 6.1.1. Définition*). Cependant, *Y. pseudotuberculosis* possède une niche écologique et un mécanisme de transmission totalement différent de celui de *Y. pestis*. En analysant la structure génétique de *Y. pestis*, le modèle le plus approprié est le modèle de dérive génétique (Achtman *et al.* 2004). Les limites des écotypes putatifs restent donc à définir : toutes les souches de *Y. pestis* correspondent-elles à un écotipe ou chaque population est-elle un écotipe différent ?

L'espèce *H. pylori* est constituée d'un certain nombre de populations génétiques qui sont retrouvées de façon prédominante dans les différentes régions du monde : Europe, Afrique de l'ouest, du nord-est et du sud, Asie de l'est et du sud-est, Océanie et Sahul (Australie, Tasmanie et Nouvelle-Guinée). Cette distribution géographique rappelle fortement celle des différentes populations humaines. La diversité génétique de *H. pylori* est maximale pour les souches provenant du nord-est de l'Afrique et décroît au fur et à mesure que la population se trouve éloignée de cette région. Cette caractéristique est partagée par la diversité génétique de l'espèce humaine (Falush *et al.* 2003; Linz *et al.* 2007). L'idée qui émerge est qu'au sein d'une humanité originelle vivant en Afrique, la population ancestrale de *H. pylori* a vécu assez longtemps dans son berceau africain pour développer une importante diversité génétique. Puis une partie de la population humaine a migré depuis l'Afrique de l'est



vers d'autres régions du globe, emportant avec elle une partie du patrimoine génétique de la population ancestrale de *H. pylori*. Dans chacune des régions colonisées par l'homme, les bactéries présentes dans l'estomac se sont à nouveau diversifiées mais isolément et donc différemment. Cependant, le transfert horizontal de souches entre groupes ethniques différents et les recombinaisons lors d'infections mixtes favorisées par les migrations humaines récentes ont commencé à brouiller ces frontières génétiques ancestrales.

## 6.2. Structure génétique de *H. pylori*

### 6.2.1. Génome

Deux génomes de *H. pylori* ont été rapidement séquencés: celui de la souche 26695 (Tomb *et al.* 1997) isolée en 1987 en Angleterre chez un patient souffrant d'une gastrite, et celui de la souche J99 (Alm *et al.* 1999) isolée en 1994 aux Etats-Unis chez un patient souffrant d'ulcère duodéal. A la date du 11/10/2011, 32 génomes ont été séquencés et annotés. Les pathologies sévères liées à *H. pylori* et en particulier son rôle dans la genèse du cancer gastrique justifie l'intense effort de recherche dont cette bactérie fait l'objet. De plus, le génome d'autres espèces appartenant au genre *Helicobacter*, telles *H. hepaticus*, bactérie impliquée dans la genèse des cancers hépatiques chez la souris immunodéprimée, *H. mustelae* impliqué dans la survenue de CG chez le furet ou les autres mustélinés, ou *H. acinonychis* impliqué dans la survenue de gastrites chez les grands fauves. La comparaison du génome des souches séquencées a mis en évidence l'exceptionnelle diversité de cette bactérie: une macrodiversité (divergence dans l'organisation des gènes, présence ou absence d'un gène) et une microdiversité (polymorphisme intragénique), qui seront détaillées ultérieurement (cf *Partie I. Généralités : 6.2.2. Microdiversité et 6.2.3. Macrodiversité*).

*H. pylori* possède un génome circulaire (contenu en guanine et cytosine (GC %) proche de 39) relativement petit, compris entre 1.58 Mb et 1.67 Mb, ce qui reflète probablement la forte adaptation de la bactérie à sa niche écologique unique (McClain *et al.* 2009; Thiberge *et al.* 2010). En se basant sur l'annotation automatique par le système MaGe de sept génomes séquencés de *H. pylori*, entre 1528 (souche B38) et 1630 (souche P12) séquences codantes (CDS pour « coding DNA sequence ») étaient présents selon le génome séquencé, dont environ 86 % codaient pour des protéines (taille moyenne proche de 965 pb) (Tableau 2). Les caractéristiques des génomes séquencés de *H. pylori* sont présentées dans le tableau 2. Au total, 73 à 78 % des CDSs étaient assignés à au moins une catégorie fonctionnelle de la classification COG (« clusters of orthologous groups »). Le génome de la

souche B38 possède 1528 CDSs dont 1189 (77 %) se sont vus assigner une fonction: 454 (38 %) appartenait à la catégorie « système de signalisation et de machinerie cellulaire », 342 (28 %) à la catégorie « système de stockage et de transmission de l'information » et 595 (50 %) étaient impliqués dans le « métabolisme ». Il est à noter qu'un même CDS peut être assigné à plusieurs catégories.

**Tableau 2. Caractéristiques des génomes de *Helicobacter***D'après Thiberge *et al.* (Thiberge *et al.* 2010)

Caractéristiques	<i>Hp</i> <sup>a</sup> B38	<i>Hp</i> 26695	<i>Hp</i> J99	<i>Hp</i> HPAG1 <sup>d</sup>	<i>Hp</i> Shi470	<i>Hp</i> G27 <sup>d</sup>	<i>Hp</i> P12 <sup>d</sup>	<i>Ha</i> <sup>b</sup> Sheeba <sup>d</sup>	<i>Hh</i> <sup>c</sup> ATCC 51449
<i>cag</i> PAI	absent	présent	présent	présent	présent	présent	présent	HacGI	HHGI1
Taille (pb)	1,576,758	1,667,867	1,643,831	1,596,366	1,608,547	1,652,982	1,673,813	1,553,927	1,799,146
Contenu en GC (%)	39.2	38.9	39.2	39.1	38.9	38.9	38.8	38.2	35.9
CDSs totaux (nb) <sup>e</sup>	1,528	1,637	1,543	1,539	1,592	1,611	1,639	1,696	1,851
CDSs complets (nb) <sup>e</sup>	1,393	1,501	1,446	1,441	1,473	1,469	1,505	1,397	1,824
Longueur moyenne (bp) <sup>e</sup>	971	964	988	971	955	954	957	933	914
Fraction codante (%) <sup>e</sup>	85.0	86.3	86.6	87.3	87.2	84.6	85.8	83.6	92.3
CDSs partiels (nb) <sup>e</sup>	135	136	97	98	119	142	134	299	27
Gènes tronqués (nb) <sup>e</sup>	2	9	10	7	4	7	7	11	11
Pseudogènes (nb) <sup>e</sup>	133(8.7 %) <sup>f</sup>	127(7.8 %)	87(5.6 %)	91(5.9 %)	115(7.2 %)	135(8.4 %)	127(7.8 %)	288(17 %)	16(0.9 %)
Séquence d'insertion (nombre d'ORFs)	<i>ISHp609</i> (5) <sup>g</sup>	<i>IS606</i> (1) <i>IS605</i> (5) <i>IS200</i> (1)	<i>IS606</i> (1 remnant) <i>ISHp609</i> (1 remnant)	<i>IS606</i> (2 remnants) <i>ISHp609</i> (1)	<i>IS606</i> en 3 fragments	<i>IS605</i> (4)		<i>ISHa1152</i> (2) <i>ISHa1942</i> (1) <i>ISHa1675</i> (1)	<i>ISHp609</i> (1 remnant)

<sup>a</sup> *Hp*, *Helicobacter pylori*; <sup>b</sup> *Ha*, *Helicobacter acinonychis*; <sup>c</sup> *Hh*, *Helicobacter hepaticus*<sup>d</sup> Ces génomes possèdent un plasmide de 9369 bp (HPAG1), 10031 pb (G27), 10225 pb (P12) et 3661 pb (Sheeba).<sup>e</sup> Nombre révisé par le système MaGe et curation manuelle.<sup>f</sup> Rapport nombre de pseudogènes / nombre de CDSs totaux<sup>g</sup> Nombre de copies

Le génome de *H. pylori* se caractérise par (Thiberge *et al.* 2010):

- sa richesse inhabituelle en systèmes de restriction-modification (R-M) (63 CDSs chez la souche B38 dont 33 potentiellement actifs),
- sa richesse en régions homopolymériques ou en répétitions de dinucléotides,
- la présence de séquences répétées sur des régions étendues distribuées régulièrement le long de l'ADN chromosomique,
- un contenu plasmidique variable,
- de multiples copies de séquences d'insertion.

En 2000, le séquençage complet du génome d'une souche de *Campylobacter jejuni* a permis de le comparer avec le génome de *H. pylori*, deux espèces phylogénétiquement proches (Parkhill *et al.* 2000). Le génome de *C. jejuni* a été totalement réannoté et réanalysé en 2007 (Gundogdu *et al.* 2007). Environ 55 % des 1643 gènes prédits de *C. jejuni* ont des orthologues chez *H. pylori* mais ces gènes ne sont pour la plupart pas impliqués dans la persistance de l'infection, la transmission et la pathogénèse, reflétant sans doute la spécificité de ces bactéries. Les gènes constituant l'opéron uréase, les gènes de la cytotoxine VacA et du *cag* PAI de *H. pylori* n'ont pas d'orthologues chez *C. jejuni*.

Si l'on compare le génome de *H. pylori* à celui de *H. hepaticus*, 940 gènes (50 %) possèdent un orthologue au sein du génome de *H. hepaticus* (Suerbaum *et al.* 2003). *H. hepaticus* ne possède pas la majorité des facteurs de virulence connus de *H. pylori* (adhésines, cytotoxine VacA et pratiquement tous les gènes composant le *cag* PAI). Cependant, il possède des gènes codant pour une uréase, deux facteurs de virulence de *C. jejuni*, l'adhésine PEB1 et la cytotoxine CDT (cytolethal distending toxin), ainsi qu'un îlot génomique de 71 kb, potentiellement associé à la pathogénicité (codant des homologues d'un système de sécrétion de type IV). Le génome de *Wollinella succinogenes*, une bactérie commensale du rumen appartenant aux *Helicobacteriaceae*, est disponible depuis 2003 : 655 et 614 gènes du génome de *W. succinogenes* possèdent un orthologue au sein respectivement du génome de *H. pylori* et de *C. jejuni*. Le génome de *W. succinogenes* porte des orthologues des gènes du plasmide de virulence pVir de *C. jejuni* codant pour un système de sécrétion de type IV. Il contient des orthologues de plusieurs facteurs de virulence de *H. pylori* et de *C. jejuni*. La présence de ces gènes est surprenante chez une bactérie définie comme non pathogène (Baar *et al.* 2003). En comparant les séquences génomiques, le nombre de gènes communs aux genres *Helicobacter* (*H. pylori*, *H. hepaticus*), *Campylobacter* (*C. jejuni*) et *Wolinella* (*W. succinogenes*) a été évalué à 700 (Eppinger *et al.* 2004).

*H. acinonychis* est une espèce très proche de *H. pylori* sur la base des séquences de l'ARNr ou de l'analyse du polymorphisme d'autres gènes (Eppinger *et al.* 2006). Cette bactérie colonise l'estomac des grands félins tels le tigre, le lion ou le guépard. L'analyse du génome a montré la présence d'un grand nombre de gènes fragmentés par décalage du cadre de lecture ou par la présence d'un codon stop (16 % des CDSs du génome). Combinée aux analyses phylogénétiques résultant de la comparaison des séquences des génomes de *H. acinonychis* et de *H. pylori* (souches 26695 et J99), cette observation suggère qu'une souche de *Helicobacter* spp. est passée de l'homme aux félins. Ce passage est supposé avoir eu lieu il y a environ 200 000 ans, suivi par la spéciation de *H. acinonychis* (Eppinger *et al.* 2006). La plupart des gènes fragmentés codent pour des protéines de membrane externe, dont l'inactivation pourrait avoir empêché les effets délétères de la réponse immunitaire de l'hôte. De plus, le génome de *H. acinonychis* contient un cluster de gènes impliqués potentiellement dans la sialylation des polysaccharides de surface, ce qui permettrait à la bactérie d'échapper au système immunitaire.

Le génome de *H. mustelae* a été récemment séquencé (O'Toole *et al.* 2010). Il s'agit de la seule bactérie responsable de la survenue d'un ulcère ou d'un cancer gastrique chez son hôte naturel (furet ou les autres mustélinés) en dehors de *H. pylori*. *H. mustelae* ne possède pas la majorité des orthologues des facteurs de virulence majeurs de *H. pylori* tels que CagA, VacA, BabA, SabA et OipA, suggérant que les mécanismes permettant la colonisation de la muqueuse gastrique, la persistance de l'infection et la genèse des lésions sont différents.

### 6.2.2. Microdiversité

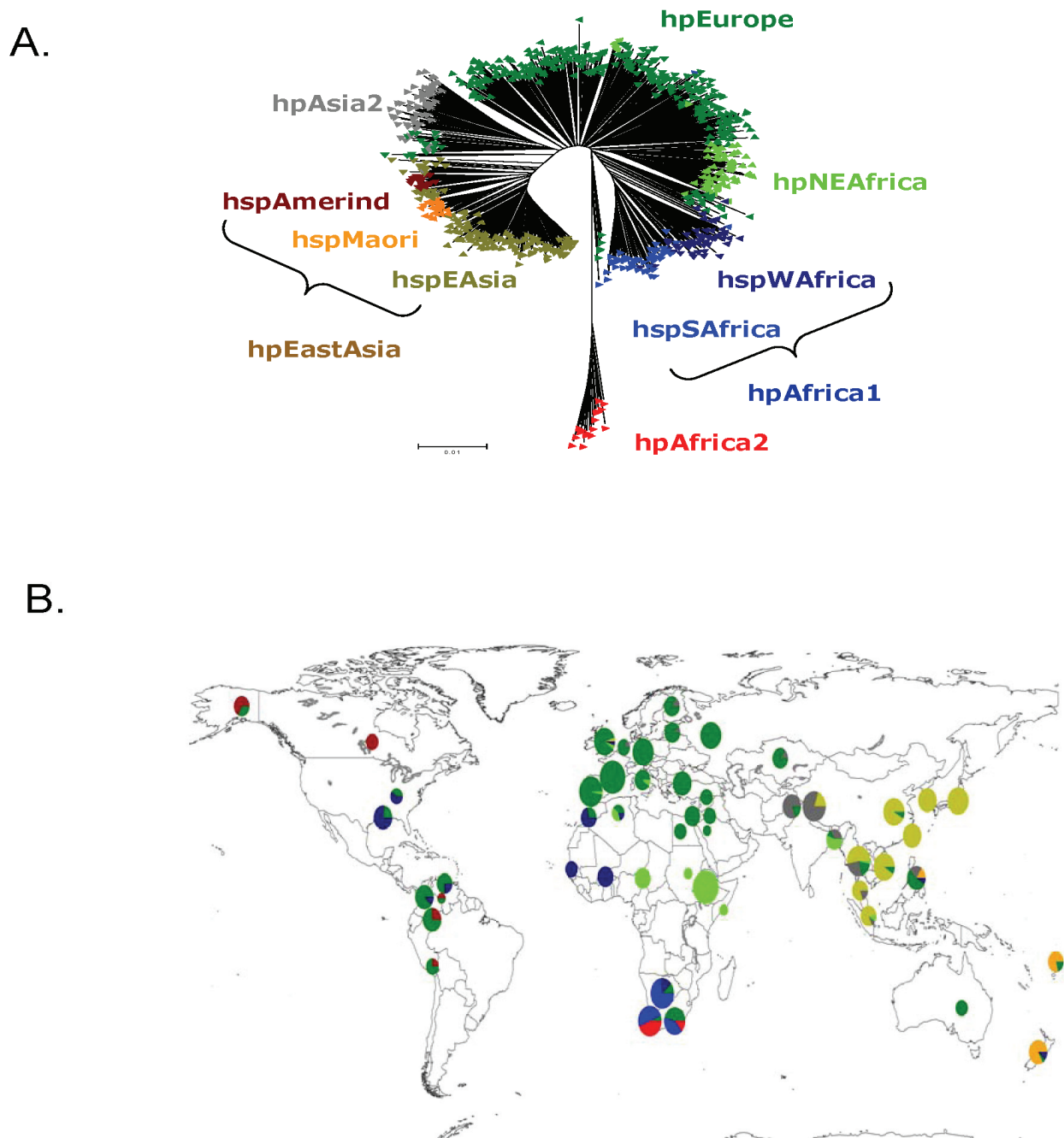
La grande variabilité génétique de *H. pylori* a été mise en évidence très tôt par les techniques d'électrophorèse en champ pulsé, de RAPD, d'empreintes génomiques et par l'établissement des cartes de restrictions génétiques (Taylor *et al.* 1992; Jiang *et al.* 1996; Kansau *et al.* 1996; Akopyants *et al.* 1998). Rapidement, l'analyse du polymorphisme génétique de gènes de ménage est devenue la méthode de référence. Dès 1996, Kansau *et al.* montraient que la séquence nucléotidique d'un fragment de 294 paires de base interne au gène *glmM* était unique à chaque souche étudiée et pouvait être utilisée comme marqueur moléculaire (Kansau *et al.* 1996). L'analyse de 7 gènes de ménage a montré que la recombinaison est si fréquente que pratiquement tous les isolats sont assignés à un nouveau séquence type (ST) et que 50 % du génome peut être remplacé par recombinaison homologue pendant un intervalle de temps compris entre 40 et 2000 ans (Falush *et al.* 2001). Cette exceptionnelle microdiversité a été confirmée par la comparaison des séquences génomiques :

les 1406 gènes communs aux génomes des deux souches 26695 et J99 présentait une identité nucléotidique moyenne de 93 %. Seuls 8 gènes présentait une identité nucléotidique d'au moins 98 % et pour 2 gènes, les séquences nucléotidiques étaient identiques. Les différences nucléotidiques sont le plus souvent des mutations silencieuses ou conservatrices (Alm and Trust 1999).

#### 6.2.2.1. Populations génétiques modernes

Malgré son exceptionnelle diversité génétique, *H. pylori* possède une forte structure phyllogéographique (Falush *et al.* 2003; Linz *et al.* 2007). L'analyse du polymorphisme génétique de 7 gènes de ménage (*atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI*, *yphC*) chez 769 souches de *H. pylori* isolées de patients originaires de 51 groupes géographiques, ethniques et linguistiques différents a permis d'assigner les souches à 6 groupes génétiques différents (logiciel STRUCTURE, modèle « no linkage »): hpEurope, hpAfrica1 (subdivisé en hspSAfrica, hspWAfrica), hpAfrica2, hpEastAsia (subdivisé en hspAmerind, hspEAsia, hspMaori), hpAsia2 et hpNEAfrica (Figure 12A).

Les souches hpEurope étaient isolées chez des patients vivant en Europe ou dans des pays autrefois colonisés par les Européens (continent américain, Australie....), ainsi qu'en Inde. Le groupe hpEastAsia se divisait en trois sous-populations, hspAmerind retrouvée exclusivement chez des patients amérindiens, hspEAsia regroupant la majorité des isolats d'Asie de l'est et hspMaori retrouvée chez des patients aborigènes vivant en Océanie (Nouvelle-Zélande, Iles Tonga, Philippines..). Le groupe hpAfrica 1 se divisait en deux sous-populations, hspSAfrica retrouvée exclusivement chez des patients vivant en Afrique du Sud, et hspWAfrica regroupant le reste des souches assignées à cette population. Les souches hspWAfrica ont été isolées en Afrique de l'ouest (Burkina Faso et Sénégal) et représentaient la majorité des souches isolées en Afrique du Sud. Cette sous-population a été également retrouvée sur le continent américain. La population hpAfrica 2 était retrouvée uniquement chez des patients d'Afrique du Sud, d'ethnies européenne ou africaine. Le groupe hpAsia2 contenait des souches isolées chez des patients originaires du nord de l'Inde, de la Thaïlande, du Bangladesh et le groupe hpNEAfrica regroupant des souches isolées chez des patients originaires d'Ethiopie, de Somalie, du Soudan et du Nord du Nigéria (Figure 12B).



**Figure 12. Sous-populations et populations génétiques de *H. pylori***

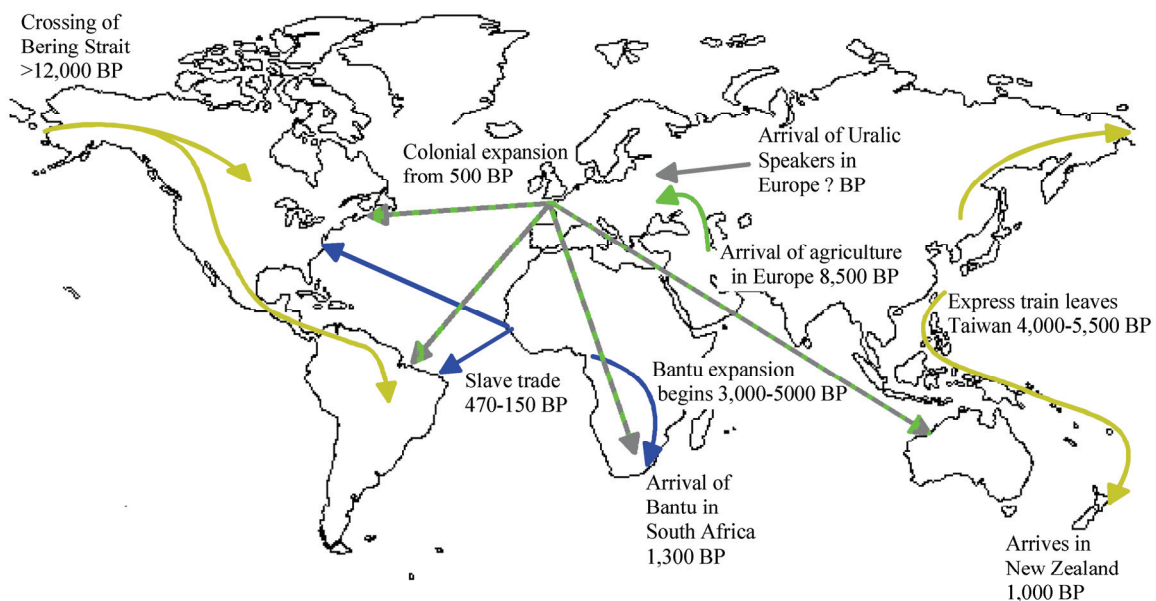
D'après Linz *et al.* (Linz *et al.* 2007)

A. Arbre phylogénétique basé sur la méthode de Neighbour-Joining construit à partir de l'analyse du polymorphisme génétique de 7 gènes de ménage.

B. Distribution géographique des sous-populations et populations.

La structure phyléogéographique de *H. pylori* reflète l'histoire des grandes migrations humaines (Figure 13):

- l'expansion coloniale européenne depuis 500 ans (population hpEurope),
- la migration indo-aryenne en Inde (4000 à 10000 ans avant notre époque) (population hpEurope),
- le commerce triangulaire des esclaves du 16<sup>ème</sup> au 19<sup>ème</sup> siècle (sous-population hspWAfrica retrouvé sur le continent américain),
- les migrations récentes des commerçants chinois en Asie du sud-est (sous-population hspEAsia),
- la colonisation du continent américain à la fin de la dernière période glaciaire, il y a plus de 12 000 ans, par le détroit de Béring (sous-population hspAmerind),
- l'expansion austronésienne en Océanie il y a environ 4500 ans (sous-population hspMaori),
- l'expansion bantou en Afrique probablement du Cameroun et du sud-est du Nigeria, vers – 3000 avant JC, en direction de l'est et puis du sud du continent (hspSAfrica).



**Figure 13. La diversité génétique de *H. pylori* reflète les anciennes grandes migrations humaines**

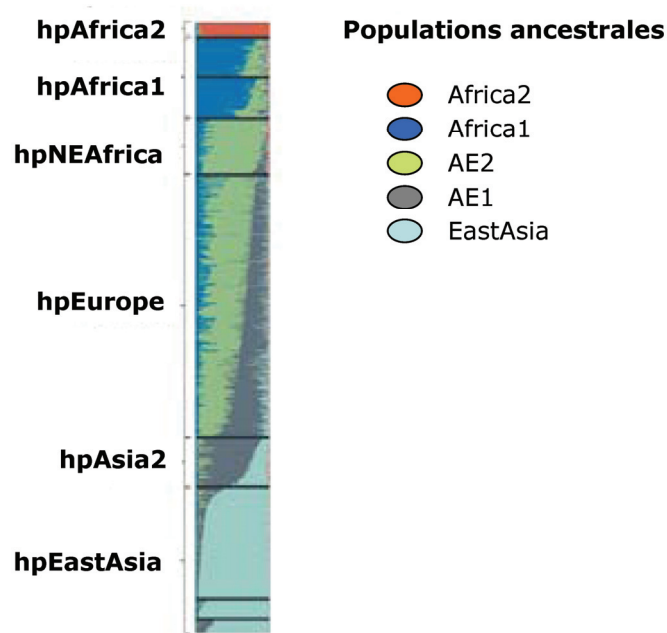
D'après Falush *et al.* (Falush *et al.* 2003)



### 6.2.2.2. Populations génétiques ancestrales

L'analyse des polymorphismes génétiques des 7 gènes de ménage par le modèle « linkage » du logiciel STRUCTURE montre que les populations génétiques dérivent de 6 populations ancestrales : ancestral Europe1 (AE1), ancestral Europe2 (AE2), ancestral EastAsia, ancestral Africa1, ancestral Africa2 et ancestral Sahul (Falush *et al.* 2003; Linz *et al.* 2007). Le terme « ancestral Europe » est mal choisi puisque AE2 aurait émergé dans le Nord-Est de l'Afrique et AE1 dans le centre de l'Asie (Linz *et al.* 2007). Ancestral EastAsia est originaire de l'est de l'Asie, Ancestral Africa1 de l'ouest de l'Afrique et Ancestral Africa2 du sud de l'Afrique. La proportion de nucléotides appartenant à chacune des populations ancestrales varie en fonction des isolats (Figure 14). Les isolats assignés aux populations hpAfrica2 et hpEastAsia dérivent de façon presque exclusive de leur population ancestrale respective, à savoir ancestral Africa2 et ancestral EastAsia. AE1 représente le composant majeur (60-90 %) des souches assignées à la branche hpAsia2, les 40-10 % restants dérivent de la population ancestral EastAsia. Le profil ancestral de la population hpEurope révèle un haut niveau de recombinaison entre AE1 et AE2, témoin probable d'une recolonisation de l'Europe après la période glaciaire.

L'étude des populations ancestrales apporte des détails supplémentaires sur les migrations humaines. Elle a permis de différencier deux populations au Ladakh (nord de l'Inde), l'une bouddhiste et l'autre musulmane. Tandis que les isolats des musulmans dérivent de façon presque exclusive de la population ancestrale AE1, ceux des bouddhistes présentaient un haut niveau de recombinaison entre AE1 et ancestral EastAsia, témoin d'une coévolution des deux populations ancestrales pendant des millénaires avant leur arrivée dans la région. La diversité génétique de *H. pylori* reflétait l'histoire du Ladakh : i) une migration il y a environ 1000 ans de Tibétains qui ont introduit le bouddhisme dans la région, ii) une migration de musulmans en provenance du Pakistan après le 14ème siècle, iii) une coexistence pendant des siècles avec peu de mélange entre les deux communautés du à un isolement culturel et religieux (Wirth *et al.* 2004).



**Figure 14. Proportion de nucléotides appartenant à chacune des cinq populations ancestrales dans chaque population génétique moderne de *H. pylori* déterminée par le logiciel STRUCTURE, modèle « linkage »**

D'après Achtman et Wagner (Achtman and Wagner 2008)

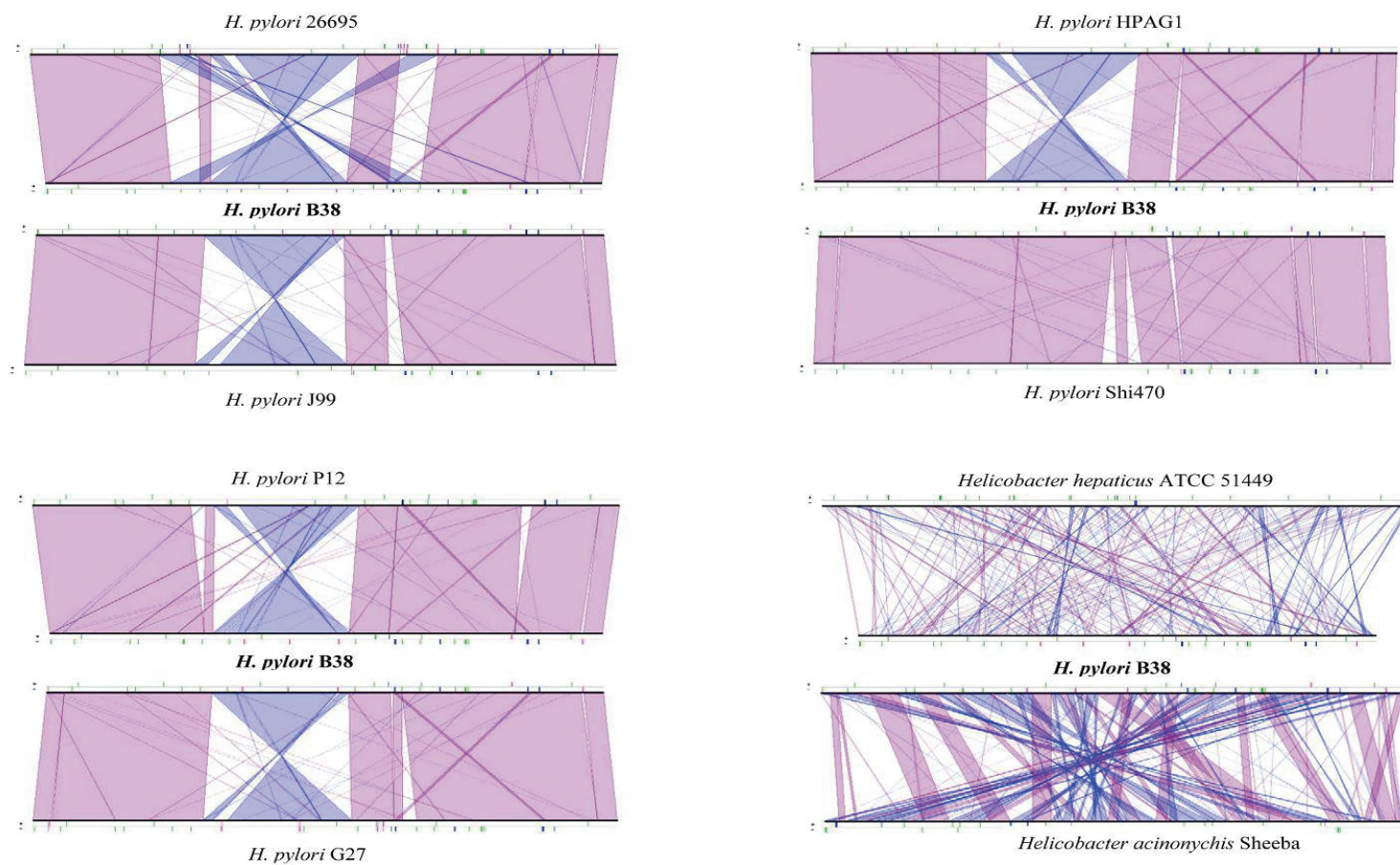
### 6.2.2.3. Coévolution entre l'homme et la bactérie

La plus grande diversité génétique est observée pour les souches africaines. En particulier, certains peuples xhosas d'Afrique du Sud sont porteurs de souches très différenciées qui, à ce jour, n'ont été observées nulle part ailleurs (hpAfrica 2). La diversité génétique de *H. pylori* décroît de façon presque linéaire lorsque la distance géographique à l'Afrique s'accroît. Cette caractéristique est partagée par la diversité génétique humaine et reflète une co-évolution entre l'homme et la bactérie (Linz *et al.* 2007). En se fondant sur le temps moyen nécessaire à un réarrangement génétique et en évaluant le nombre de ces réarrangements séparant la population ancestrale des souches actuelles, l'expansion des souches de *H. pylori* en dehors de l'Afrique s'est produite il y a environ 58 000 ans, date en concordance parfaite avec les datations les plus récentes de l'expansion de l'homme. Ceci implique que *H. pylori* était présent en Afrique avant les grandes migrations humaines suggérant que l'Afrique est à la fois le berceau de l'homme et de *H. pylori* (Linz *et al.* 2007) (Figure 14).

### **6.2.3. Macrodiversité**

#### **6.2.3.1. Organisation génomique : sythénie**

La comparaison des différentes séquences génomiques a révélé que les gènes sont organisés de façon similaire dans les souches de *H. pylori*: seules 10 régions du génome de la souche J99 de taille comprise entre 1 et 83 Kb sont transposées ou inversées par rapport au génome de la souche 26695 (Alm *et al.* 1999). Le génome de la souche B38 (population génétique hpEurope) récemment publié possède une organisation très proche de celui de la souche Shi470 isolé d'un Amérindien (population génétique hspAmerind), une co-linéarité jamais décrite auparavant. En revanche, l'organisation génomique entre les différentes espèces appartenant au genre *Helicobacter* (*H. hepaticus*, *H. acinonychis*) est très différente (Figure 15).



**Figure 15. Synthénie entre le génome de la souche de *H. pylori* B38 et les souches de *H. pylori* 26695, J99, HPAG1, Shi470, P12, G27 et de *H. hepaticus* et de *H. acinonychis***

D'après Thiberge *et al.* (Thiberge *et al.* 2010)

### 6.2.3.2. “Core” génomique et pan-génome

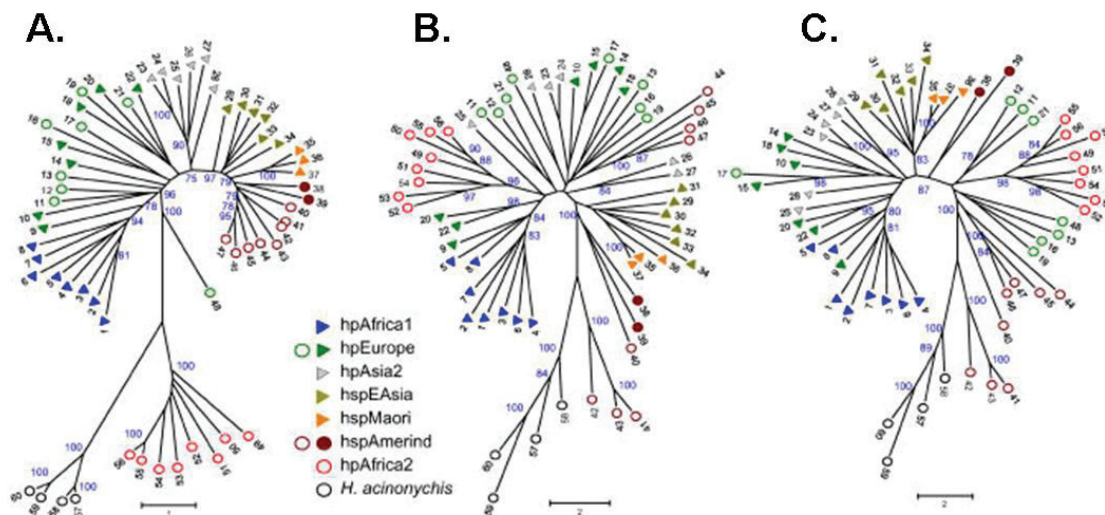
Une analyse génomique à l’aide de puces à ADN a comparé la distribution des gènes identifiés dans le génome des souches 26695 et J99 dans le génome de 15 souches de *H. pylori* d’origine caucasienne. Parmi les 1643 gènes analysés, 1281 (78 %) gènes (appelés gènes ubiquistes) étaient présents dans les 15 souches et formaient le « core » génomique de l’espèce *H. pylori* (Salama *et al.* 2000). Ce nombre est légèrement plus élevé que dans l’étude publiée par Gressman *et al.* qui a extrapolé la taille du « core » génomique à 1111 gènes (Gressmann *et al.* 2005). Selon les différentes études publiées, 20 à 25 % des gènes étaient « souches-spécifiques » ou « non-ubiquistes » avec des variations qualitatives de leur composition (Salama *et al.* 2000; Gressmann *et al.* 2005; Han *et al.* 2007; Romo-Gonzalez *et al.* 2009). Ces différences peuvent s’expliquer par : i) une différence de contenu génétique entre isolats associés à des pathologies gastro-duodénales différentes et issues de populations ethniques et/ou géographiques différentes, ii) la méthodologie choisie pour l’élaboration de la puce à ADN, la standardisation et l’interprétation du signal, iii) les conditions expérimentales et iv) l’expérience du manipulateur.

Les gènes « non-ubiquistes » codent majoritairement pour des protéines de fonctions inconnues, des transposases ou des gènes appartenant au système de R-M. Dans le génome séquencé des souches 26695 et J99 de *H. pylori*, la plupart de ces gènes non-ubiquistes étaient localisés dans 2 grandes régions, le *cag* PAI et la zone de plasticité (PZ), caractérisées par un GC % différent de celui du reste du génome. Le *cag* PAI est principalement transmis sous la forme d’un bloc unique de 27 gènes, tandis que la PZ contient de multiples transposases, endonucléases, suggérant des phénomènes de recombinaison, insertion et délétion.

Les analyses de macrodiversité à l’aide de puces à ADN sont basées sur la présence ou l’absence de gènes déjà connus, ce qui ne permet pas l’identification de gènes additionnels inconnus, et donne par conséquent une image imparfaite de la diversité génétique de cette bactérie. Seul le séquençage complet du génome de multiples souches représentatives de la diversité génétique de *H. pylori* permettra de définir précisément le « core » génomique et le pan-génome (correspondant au « core » génomique et aux gènes auxiliaires). Une analyse de sept génomes totalement séquencés a identifié 1275 gènes toujours présents (Thiberge *et al.* 2010). Ce nombre est légèrement plus élevé que celui récemment publié par McClain *et al.* (1237) qui s’est basé sur l’analyse de 5 génomes (McClain *et al.* 2009) puisqu’il prend en compte des CDSs additionnels détectés par le système MaGe.

### 6.2.3.3. Evolution de la macrodiversité

L'analyse d'une collection de 56 souches de *H. pylori* représentatives des différentes populations et sous-populations génétiques (cf *Partie I. Généralités : 6.2.2. Microdiversité*) ainsi que de quatre souches de *H. acinonychis* à l'aide de puces à ADN a apporté un nouvel éclairage sur l'évolution de la macrodiversité au cours du temps (Gressmann *et al.* 2005). Dans cette étude, 25 % des gènes étaient décrits comme étant non-ubiquistes. La topologie de l'arbre phylogénétique basé sur la méthode de Neighbor-Joining construit à partir des données issues de ces puces à ADN différait de celle basée sur les séquences des 7 gènes de ménage habituellement utilisés (Figure 16).



**Figure 16. Arbres phylogénétiques construits à partir des données de la MLSA et des puces à ADN**

D'après Gressmann *et al* (Gressmann *et al.* 2005)

A. Séquençage de 7 gènes de ménage

B. Les données des puces à ADN excluant le *cag* PAI

C. Les données des puces à ADN incluant le *cag* PAI

Triangles pleins, souches *cag* PAI positives ; cercles vides, souches *cag* PAI négatives ; cercles pleins, souches hspAmerind ne possédant pas l'ORF HP0536 du *cag* PAI

L'analyse de ces arbres permet d'émettre des hypothèses quant à l'origine de ces gènes non-ubiquistes de souches. En particulier, il est intéressant de savoir s'il s'agit d'une acquisition de ces gènes par certaines souches, d'une perte par d'autres, ou d'une combinaison des deux mécanismes. Les gènes variables sont en général de faible taille et répartis sur tout le génome. Ils sont donc potentiellement transférables entre souches par recombinaison homologue. De plus, ils sont retrouvés dans toutes les (sous-) populations génétiques de *H. pylori* et codent pour des fonctions qui n'ont pas de potentiel de sélection positive évident (enzymes de restriction, transposases, protéines de fonction inconnue). L'hypothèse est donc une acquisition de ces gènes par le dernier ancêtre commun puis une perte par certaines souches au cours de l'évolution. Quant au *cag* PAI, il n'est pas uniformément présent dans toutes les populations de *H. pylori*. Etant absent chez l'ensemble des souches assignées à hpAfrica2, le *cag* PAI pourrait avoir été acquis en bloc après la séparation de hpAfrica2 et de l'ancêtre des autres populations de *H. pylori*. Les souches *cag* PAI négatives auraient ensuite perdu l'îlot par délétion ou transformation avec un site vide. Une récente analyse phylogénétique du *cag* PAI de souches assignées aux différentes sous-populations et populations génétiques représentatives de la microdiversité, confirme ce résultat : il a été acquis en bloc avant la sortie de l'homme moderne d'Afrique de l'est il y a 60 000 ans puisque sa microdiversité est corrélée à celle des 7 gènes de ménage et qu'elle diminue au fur et à mesure que l'on s'éloigne du nord-est de l'Afrique (Olbermann *et al.* 2010).

#### **6.2.3.4. Contenu génétique et pathologie gastro-duodénale**

La plupart des études de génomique comparative ont recherché sans succès une association entre le contenu génétique des souches et la pathologie gastro-duodénale associée. Récemment, une association a été mise en évidence avec des souches associées à un lymphome du MALT (Thiberge *et al.* 2010). La méthodologie a été la suivante : i) une étude préliminaire a comparé le génome de 24 souches de *H. pylori* d'origines géographiques variées (Afrique, Asie, Europe) à l'aide de membranes d'hybridation à haute densité contenant l'ensemble des gènes de la souche 26695. Elle a permis d'identifier un sous-jeu de gènes non-ubiquistes restreint (248) qui a servi de base à la fabrication d'un nouveau jeu de membranes d'hybridation. ii) 120 isolées de patients vivant en France et présentant des pathologies variées ont été analysées à l'aide de ces membranes: 33 étaient associées à une gastrite chronique, 27 à un ulcère duodéal, 17 à une métaplasie intestinale et 43 à un lymphome du MALT. L'analyse en composantes principales en utilisant les données relatives aux gènes les plus discriminatifs a permis d'identifier un cluster de 20 souches toutes *cag* PAI

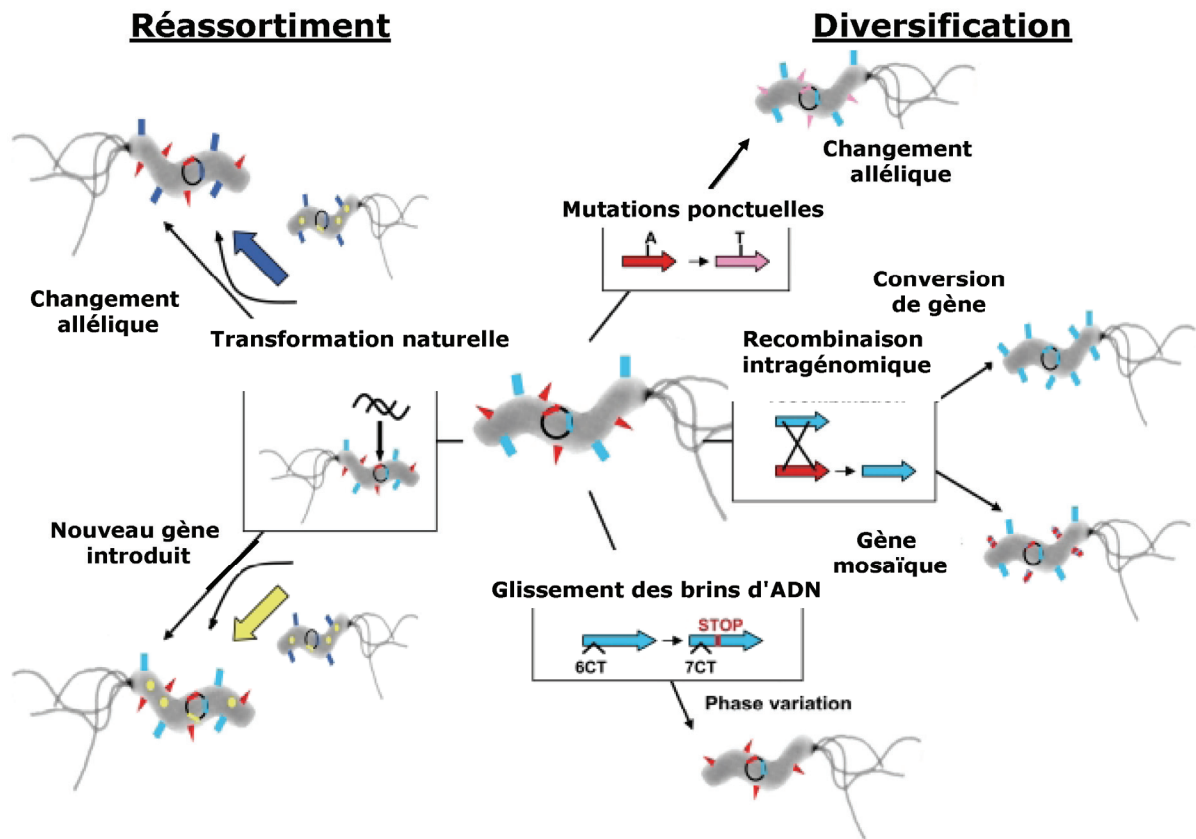
négative, dans lequel 18 souches étaient associées à un lymphome du MALT. Ces 18 souches étaient toutes *vacA s2m2*, c'est à dire sans toxine fonctionnelle. Une souche de ce cluster (B38) a été sélectionnée (absence de plasmide, transformable, adaptable à la souris) pour être séquencée. Cette souche, en plus de l'absence de *cag* PAI et de son statut *vacA s2m2* était dépourvue de 4 gènes codant pour des protéines de membrane externe (*babB*, *babC*, *sabB* et *homB*). Elle possède le plus petit génome jamais décrit (1.57 Mb) (Thiberge *et al.* 2010).

Les souches associées au lymphome du MALT semblent avoir un contenu génétique différent de celui des souches associées aux autres pathologies, contenu qui pourrait jouer un rôle déterminant dans la physiopathologie de la prolifération des lymphocytes tumoraux. Ces résultats confirment ceux de Lehours *et al.* (Lehours *et al.* 2004; Lehours *et al.* 2009), qui sur la même collection de souches, avaient montré un lien entre la survenue d'un lymphome du MALT et certains marqueurs génétiques. Ceci illustre l'importance des interactions entre facteurs de la bactérie, de l'environnement et de l'hôte dans le développement des pathologies gastro-duodénales sévères. Une meilleure compréhension de ces interactions est nécessaire pour mieux appréhender les mécanismes physiopathologiques qui aboutissent au cancer.

#### **6.2.4. Mécanisme générant la diversité génétique**

*H. pylori* est l'une des espèces génétiquement les plus diverses du monde bactérien. Des variations génétiques chez un microorganisme indiquent une sélection constante dans un environnement dynamique. Elles permettent à *H. pylori* de s'adapter à un nouvel hôte après sa transmission, de coloniser l'ensemble de l'estomac et de s'adapter aux conditions changeantes de l'estomac au cours du temps afin notamment d'échapper à la réponse immunitaire de l'hôte. Les mécanismes générant la diversité génétique chez *H. pylori* sont présentés figure 17.





**Figure 17. Mécanismes générant la diversité génétique chez *H. pylori***

D'après Dorer *et al.* (Dorer *et al.* 2009)

Les flèches colorées représentent des gènes. Les triangles, rectangles et cercles colorés représentent les protéines codées par ces gènes. Les mécanismes de diversification incluent des mutations ponctuelles, le glissement des brins d'ADN et les recombinaisons intragénomiques. Les changements alléliques impliquent des mutations non silencieuses, des gènes mosaïques dus à des recombinaisons intragénomiques, la variation de phase par glissement des brins d'ADN. Ils peuvent altérer la fonction et/ou les épitopes antigéniques des protéines correspondantes. L'expression des gènes peut également être régulée par la conversion de gènes due à la recombinaison intragénomique et la variation de phase par glissement des brins d'ADN. Le réassortiment de gènes par transformation naturelle avec de l'ADN génomique exogène contribue également à la diversité génétique.

#### 6.2.4.1. Mutations ponctuelles

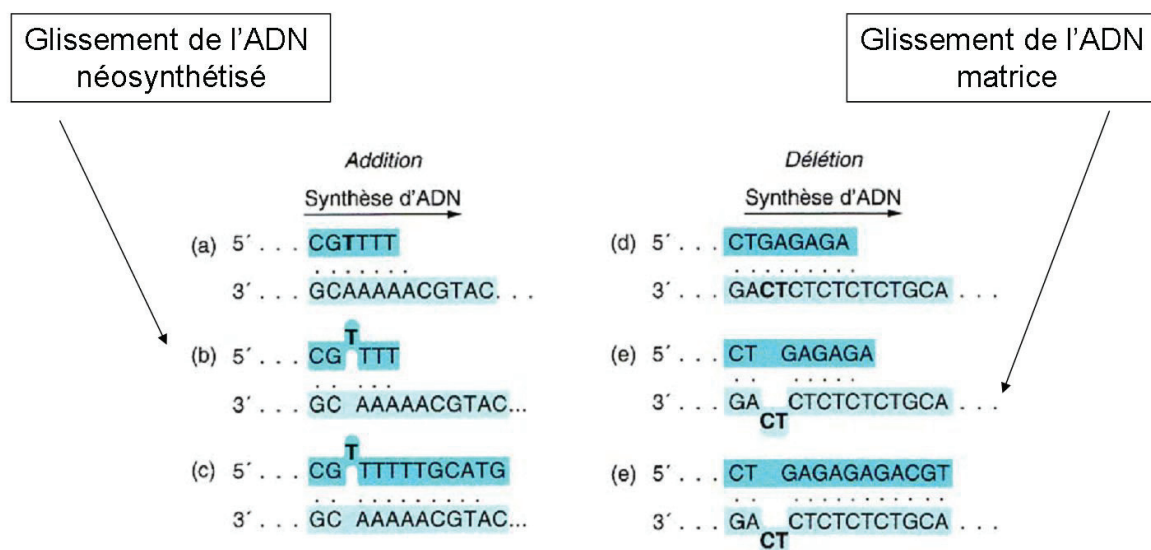
Une mutation est une modification irréversible et héréditaire de l'information génétique. Elle peut résulter d'une substitution, d'une insertion ou d'une délétion d'un ou plusieurs acides nucléiques. Les mutations par substitution peuvent être classées selon leurs modalités de modification du gène en mutation « synonyme », « faux-sens » et « non-sens ». La plupart des mutations chez *H. pylori* sont des mutations synonymes (Alm *et al.* 1999), c'est-à-dire des mutations qui, en raison de la redondance du code génétique ne modifient pas la séquence d'une protéine. Ces mutations peuvent aussi toucher une région non codante de l'ADN. La mutation faux-sens entraîne une modification de l'acide aminé codé. Elle peut avoir une répercussion sur la fonction de la protéine produite par le gène, dans le cas d'un gène codant, ou sur l'affinité pour un facteur de transcription, dans le cas d'une zone promotrice de l'ADN. Les mutations « non-sens » provoquent le remplacement d'un codon codant un acide aminé par un codon stop, entraînant la production d'une protéine tronquée. Une addition ou une suppression de nucléotides non multiple de 3 provoquera un changement de cadre de lecture, générant le plus souvent au moment de la traduction, une protéine tronquée par l'apparition d'un codon stop prématuré.

Le taux de mutation chez *H. pylori* est plus élevé que chez la majorité des autres bactéries. *H. pylori* a été qualifié de bactérie hypermutatrice. Le support génétique de la résistance à la rifampicine chez *H. pylori*, comme chez la plupart des espèces, est le gène *rpoB*, dont le taux de mutation reflète le taux de mutation globale de *H. pylori*. La plupart des souches de *H. pylori* (36 des 39 souches testées) montraient un taux de mutation plus élevé à la rifampicine que le taux de mutation moyen de souches sauvages d'entérobactéries. De plus, environ 25 % des souches présentaient un taux de mutation plus élevé que celui des souches mutantes d'entérobactéries dont le gène *mutS* avait été inactivé (Bjorkholm *et al.* 2001). Le taux de mutation de *H. pylori* a été récemment estimé à  $1.4 \times 10^{-6}$  par nucléotide et par année (Morelli *et al.* 2010).

#### 6.2.4.2. Variation de phase par glissement des brins d'ADN

La comparaison des génomes des souches 26695 et J99 a permis l'identification de 46 gènes potentiellement soumis à une variation de phase sur la base de la présence de régions homopolymériques ou de répétitions de dinucléotides. Dans une collection de 23 souches issues des populations génétiques représentatives de la microdiversité (cf *Partie I. Généralités : 6.2.2. Microdiversité*), 30 gènes présentaient une variation du nombre de répétitions de séquence par glissement des brins d'ADN (« slipped-strand mispairing ») (Figure 18) (Salaun

*et al.* 2004). Le nombre de répétitions n'était pas corrélé à des populations génétiques particulières.



**Figure 18. Glissement des brins d'ADN au niveau d'une séquence répétée durant la réplication**

La variation du nombre de répétitions est à l'origine i) de mutations survenant dans la région promotrice et affectant la transcription du gène situé en aval ou ii) de mutations survenant dans la région codante et affectant la traduction par l'apparition d'un codon stop causant une interruption prématurée (système du « tout ou rien » ou «on/off»). La plupart de ces gènes codaient des protéines impliquées dans la synthèse du LPS ou de protéines de surface cellulaire, mettant en valeur l'importance du mécanisme de variation de phase dans le contrôle de la synthèse de molécules qui interagissent directement avec l'environnement. Les gènes codant pour les adhésines SabB, SabA, BabB et la protéine proinflammatoire OipA sont soumis par exemple, à une variation de phase (Backstrom *et al.* 2004; Kudo *et al.* 2004; Aspholm *et al.* 2006). L'expression des glycosyltransférases et fucosyltransférases impliquées dans la synthèse du LPS est également régulée par un mécanisme de variation de phase et l'expression variable des fucosyltransférases permet à la bactérie d'imiter les antigènes Lewis humains, ce qui permet un échappement au système immunitaire et donc participe à sa persistance au sein de l'estomac (Lozniewski *et al.* 2003). Chez *C. jejuni*, la variation de phase de certains gènes est impliquée dans la variation antigénique à travers la modification

du LPS et de la capsule bactérienne (Linton *et al.* 2001). Les régions homopolymériques chez les bactéries du genre *Helicobacter* sont plus fréquentes que les répétitions de dinucléotides. Cependant, elles sont virtuellement absentes chez *W. succinogenes*, mais cette absence est compensée par une abondance de répétitions de dinucléotides (Baar *et al.* 2003; Eppinger *et al.* 2004). La variabilité antigénique par variation de phase semble être une stratégie globale chez les  $\epsilon$ -proteobacteria, leur permettant une adaptation à leurs hôtes respectifs et donc la persistance.

#### 6.2.4.3. Recombinaison

Le réassortiment de gènes peut résulter de la recombinaison avec de l'ADN exogène provenant d'un autre clone ou d'un clone variant de la même souche. Le point clef est la capacité de la bactérie à récupérer de l'ADN exogène et à l'incorporer dans son génome. *H. pylori* est une bactérie naturellement compétente, c'est-à-dire qu'elle est capable d'internaliser de l'ADN présent dans l'environnement extérieur, puis de l'intégrer dans son génome par recombinaison homologue (Suerbaum *et al.* 1998; Israel *et al.* 2000). De nombreuses barrières empêchent les échanges génétiques entre souches d'*H. pylori* et de façon plus large, entre *H. pylori* et les autres procaryotes [pour une revue (Kraft and Suerbaum 2005)]. A la différence des autres bactéries, *H. pylori* ne semble pas posséder de séquence « uptake » qui est un site de reconnaissance essentiel à l'internalisation d'ADN lors de la transformation de la bactérie puisqu'elle permet à la bactérie de reconnaître de l'ADN apparenté. Cependant, elle est capable de différencier l'ADN apparenté de l'ADN non apparenté et n'internalisera pas dans son génome de l'ADN provenant d'autres espèces du genre *Helicobacter*. La seule exception connue a été l'internalisation d'ADN provenant de *H. acinonychis*. Le mécanisme d'internalisation de l'ADN par *H. pylori* n'est pas complètement compris mais le locus *comB* codant pour un appareil de sécrétion de type IV semble jouer un rôle majeur (Kraft and Suerbaum 2005). Bien que le transfert de matériel génétique par conjugaison chez *H. pylori* (Kuipers *et al.* 1998) ait été décrit et que la présence de phages ait été observée en microscopie électronique (transduction), le rôle de ces deux mécanismes dans la diversité génétique de cette bactérie reste à démontrer

Un élément clef est le système de R-M. Il protège l'ADN chromosomique de l'ADN étranger, le plus commun étant les bactériophages. (Kong *et al.* 2000). Le système R-M associe systématiquement deux enzymes, l'une de coupure et l'autre de protection : i) une endonucléase est capable de cliver un fragment d'ADN au niveau d'une séquence de nucléotides caractéristique, appelée site de restriction, ii) une méthylase qui reconnaît

également le site de restriction; elle ne coupe pas l'ADN mais le modifie en lui rajoutant un groupement méthyle sur un ou plusieurs nucléotides du site, empêchant la coupure par l'enzyme de restriction de son propre ADN. Chez la souche B38 récemment publiée (Thiberge *et al.* 2010), le système R-M comprenait 63 CDSs dont 33 potentiellement actifs. Bien que la signification biologique de cette richesse en systèmes de R-M soit encore peu claire, cette diversité serait due à un gain continu de nouveaux systèmes de R-M et à une inactivation de systèmes plus anciens, diminuant le risque d'inactivation par mutation (Lin *et al.* 2001).

Plus récemment, un protocole de transformation *in vitro* de souches de *H. pylori* avec de l'ADN exogène a été développé. La longueur moyenne des fragments incorporés dans le génome était comprise entre 1294 et 3853 pb, une taille plus petite que chez la plupart des autres espèces bactériennes (Kulick *et al.* 2008). Dans environ 10 % des clones recombinants, le fragment d'ADN intégré était interrompu par de courtes séquences d'ADN non remplacées de la souche réceptrice (« interspersed sequences of the recipient » ou ISR) avec une longueur moyenne de 82 pb. Ce mécanisme contribue à la diversité génétique de cette bactérie. Dix-huit gènes candidats ont été inactivés dans le but d'identifier les gènes impliqués dans le contrôle de la taille des fragments intégrés. L'inactivation de la glycosylase MutY augmentait la longueur des fragments intégrés mais n'avait pas d'effet sur la fréquence des ISR. Une surexpression de *mutY* augmentait de façon significative la fréquence des ISR, indiquant que MutY, bien que non indispensable à la formation des ISR, intervient dans le système les générant.

#### 6.2.4.4. Systèmes de réparation de l'ADN

Plusieurs grands systèmes de réparation de l'ADN existent au sein des procaryotes. Ils permettent (Figure 19):

- la correction des mutations spontanées (réparation des mésappariements ou « mismatch repair »),
- l'inhibition de la variation du nombre de répétitions de séquences par glissement des brins d'ADN (réparation des mésappariements),
- la réparation des dommages liés au radicaux libres et aux agents alkylants (réparation par excision de nucléotides ou « nucleotide excision repair » ; réparation par recombinaison ; réparation par excision de base ou « base excision repair »),
- la prévention des recombinaisons illégitimes (réparation des mésappariements ; réparation par recombinaison).

Tous ces événements surviennent au cours de la réplication de l'ADN et du métabolisme cellulaire. Sans ces mécanismes de réparation de l'ADN, l'accumulation de mutations délétères peut devenir incompatible avec la survie de la bactérie. L'absence de certains homologues de gènes impliqués dans la réparation de l'ADN chez *H. pylori* et communs à la plupart des espèces bactériennes contribue à la grande diversité génétique de la bactérie (Alm and Trust 1999).

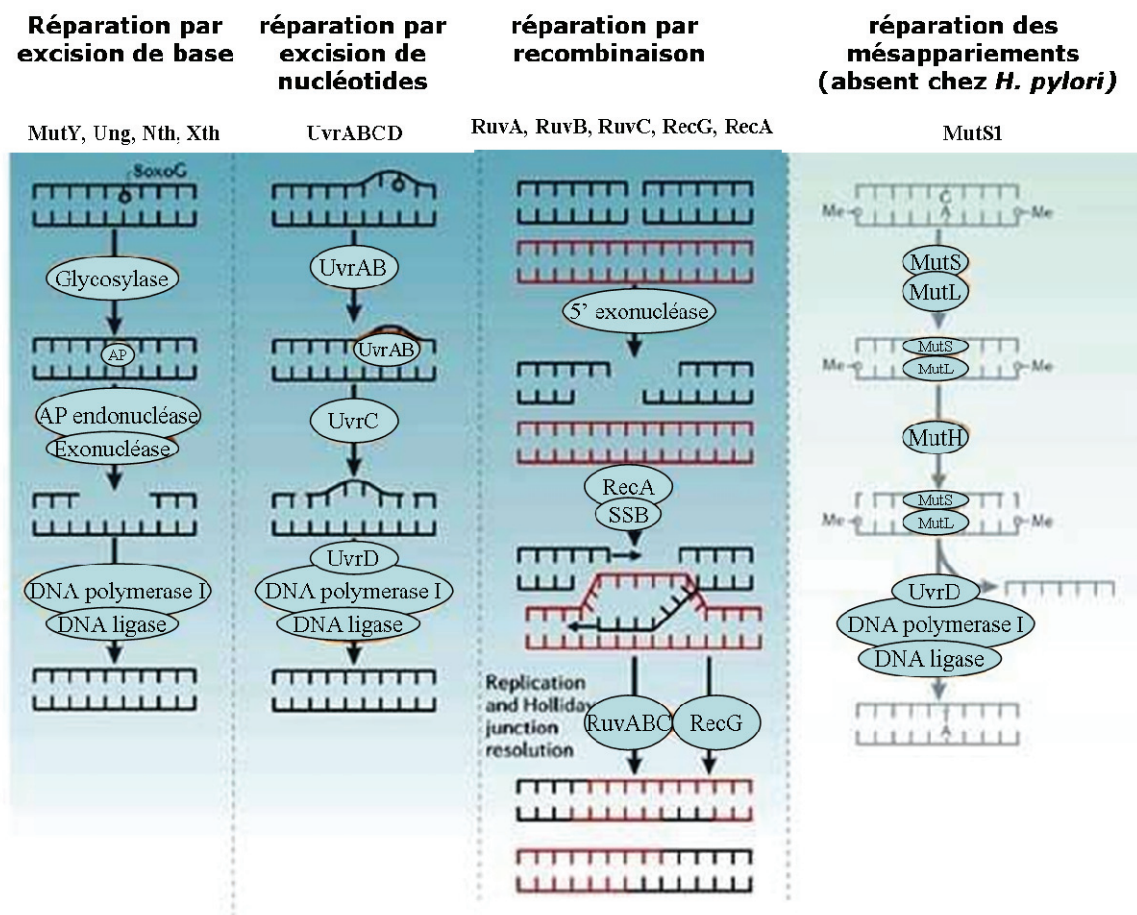


Figure 19. Systèmes de réparation de l'ADN au sein des procaryotes.

D'après Kang et Blaser (Kang and Blaser 2006)

#### 6.2.4.4.1. Réparation des mésappariements

Des analyses des séquences génomiques de souches de *H. pylori* ont montré que *H. pylori* ne possède pas les homologues de la plupart des gènes permettant la réparation de l'ADN : aucun composant putatif du système de réparation des mésappariements (SRM) n'a pu être identifié (système *mutS1/mutL/mutH* retrouvé chez *Escherichia coli*) à part UvrD

(Humbert *et al.* 1995; Schofield and Hsieh 2003). *H. pylori* génère de façon continue des mutations, des recombinaisons, des variations de phase, en raison d'un SMR incomplet.

#### 6.2.4.4.2. Réparation par excision de nucléotides

*H. pylori* possède des homologues des 4 composants du système de réparation par excision de nucléotides (NER), UvrA, UvrB, UvrC et UvrD (Kang and Blaser 2006). Ce système reconnaît et répare les diverses lésions de l'ADN résultant d'une exposition aux UV, aux agents alkylants ou intercalants. Des mutants NER ont une sensibilité augmentée aux lésions de l'ADN (Abril *et al.* 1992). L'environnement de *H. pylori* étant riche en radicaux libres (Normark *et al.* 2003), ce système est important pour le réparer et pour limiter les mutations. La relative homologie entre *H. pylori* UvrD et UvrD des bactéries extrémophiles, *Thermotoga maritima* et *Aquifex aeolicus*, pourrait refléter une évolution convergente d'organismes habitant un environnement génotoxique (Kang and Blaser 2006).

#### 6.2.4.4.3. Réparation par excision de base

Le système de réparation par excision de base (BER) implique des glycosylases telles que MutY, MutM, Ung, des AP endonucléases telles que Xth et Nfo, qui reconnaissent et réparent un dommage au niveau d'une base de l'ADN. *H. pylori* possède des homologues de l'ensemble de ces enzymes sauf Nfo (Kang and Blaser 2006). Le système BER est crucial pour la réparation de dommages oxydatifs causés à une ou plusieurs bases de l'ADN (Huang *et al.* 2006). Comme ce système peut reconnaître et corriger ces mutations, il pourrait jouer un rôle très important pour préserver l'intégrité du génome de *H. pylori* qui est déficient en SMR.

Chez des organismes autres que *H. pylori*, il a été montré que les souspopulations mutantes sont les produits d'une adaptation à un environnement difficile et que l'accumulation de mutations dans des conditions favorables désavantage progressivement le phénotype mutant. Les glycosylases pourraient jouer un rôle majeur dans cet équilibre par un mécanisme de variation de phase (Kang and Blaser 2006). Des données génétiques et expérimentales montrent un mécanisme de variation de phase pour le gène *mutY* (Huang *et al.* 2006), pouvant générer une inactivation de l'expression du gène (phénotype « off ») et donc une hypervariabilité génétique. Une population bactérienne pourrait donc vivre dans un équilibre dynamique, avec deux souspopulations de phénotype génétiquement hypervariable ou non, la balance étant déterminée par le bénéfice ou le coût résultant de la diversité génétique.

#### 6.2.4.4. Réparation par recombinaison

L'annotation des génomes de *H. pylori* a permis d'identifier une quantité limitée de gènes liés au système de réparation par recombinaison, parmi lesquels un homologue de RecA (Alm and Trust 1999). Certains systèmes de recombinaison connus chez *E. coli* sont soit absents du génome de *H. pylori*, soit présents de façon rudimentaire. *H. pylori* ne possède pas les homologues des recombinases, TecBCD, RecF, RecO, RecQ, SbcC et SbcD, suggérant une susceptibilité plus élevée aux agents responsables de lésions de l'ADN et aux réarrangements génétiques aberrants. Cependant, *H. pylori* possède les homologues RecG, RuvA et RuvB.

Aucun composant putatif du système de réparation des mésappariements n'a pu être identifié (système *mutS1/mutL/mutH* retrouvé chez *E. coli*). Toutefois, la protéine codée par l'ORF HP0621 présente une homologie avec les protéines MutS, ce qui pose la question de sa fonction. Ne présentant pas toutes les caractéristiques des protéines MutS1, elle a été rattachée à la famille MutS2. Ces protéines ATPasiques participent au contrôle du crossing-over au cours de la méiose chez les eucaryotes, mais elles n'ont pas de fonction connue chez les procaryotes. Pour tester un éventuel rôle d'HpMutS2 dans le SMR, le taux de mutation du gène conférant la résistance à la rifampicine, un gène essentiel à la bactérie, a été estimé chez une souche de *H. pylori* sauvage et chez une souche dont le gène HP0621 a été inactivé (Pinto *et al.* 2005). La même expérience a été réalisée pour un gène non essentiel (gène conférant la résistance au métronidazole). Dans un cas comme dans l'autre, aucune différence dans le taux de mutation n'a été observée entre les souches sauvages et mutées, indiquant que MutS2 ne participe pas au SMR. Des expériences de transformation de *H. pylori* par des séquences d'ADN ayant des degrés d'homologie variés avec son génome ont été réalisées. Étonnamment, les recombinaisons homologues sont 3 à 21 fois plus fréquentes dans les souches mutées pour HpMutS2 que dans les souches sauvages, suggérant que cette protéine est impliquée dans l'inhibition de la recombinaison homologue. De manière intéressante, un plus faible degré d'homologie des régions de recombinaison (jusque 26 % de divergence) n'affecte pas l'efficacité de ce mécanisme. Il a de plus été montré que HpMutS2 inhibe les échanges de brins d'ADN catalysés par la protéine RecA de façon indépendante du degré d'homologie entre les séquences, confirmant que HpMutS2 agit sans distinction sur des séquences homologues et hétérologues. MutS2 pourrait donc contrôler la plasticité du génome en régulant l'intégration d'ADN exogène dans le chromosome bactérien et en limitant la réorganisation de séquences à l'intérieur du chromosome. La fonction exacte de HpMutS2 et la nécessité de cofacteurs restent à déterminer.



### 6.3. Changements génétiques pendant l'infection chronique *in vivo*

#### 6.3.1. Au niveau microdiversité

*H. pylori* est une des espèces bactériennes présentant un des plus hauts taux de recombinaison (Vos and Didelot 2009). La dynamique de la recombinaison a été étudiée *in vivo* en analysant la diversité génétique de 26 paires d'isolats prélevées de façon itérative à un intervalle moyen de 1.8 ans chez des patients vivant aux Etats-Unis et en Colombie (Falush *et al.* 2001). Pour chacun de ces isolats, 10 fragments de gènes représentant une longueur totale de 4658 pb ont été séquencés. Treize paires d'isolats avaient une séquence génomique identique sur l'ensemble des loci. Les 13 autres paires d'isolats se différenciaient deux à deux par un polymorphisme touchant un seul nucléotide sur un locus (3 paires) ou de multiples nucléotides sur un ou plusieurs loci (10 paires). Afin de modéliser l'évolution du génome *in vivo*, un algorithme basé sur des méthodes bayésiennes a été développé. Le génome de *H. pylori* se caractérise par un taux de recombinaison exceptionnellement élevé et par une longueur moyenne des fragments incorporés plus courte que pour les autres bactéries (417 pb versus 14 000 pb pour *E. coli*, 6000 pb pour *S. pneumoniae*). Il a été estimé par ces mêmes auteurs que 50 % du génome était renouvelé par recombinaison pendant une période comprise entre 40 et 2000 ans. A titre de comparaison, 10 à 100 millions d'années sont nécessaires pour remplacer 60 % du génome de *E. coli*. Récemment, les taux de mutation et de recombinaison ont été estimés par séquençage de 78 fragments d'ADN d'une longueur moyenne de 39 300 pb chez 97 isolats. Ils correspondaient à 34 paires d'isolats prélevées à un intervalle moyen de 1.8 an et à 29 isolats prélevés chez 29 individus appartenant à 10 familles. En utilisant un algorithme basé sur des méthodes bayésiennes, les recombinaisons avaient entraîné 3 fois plus de modifications de nucléotides que ne l'avaient fait les mutations (Morelli *et al.* 2010). La longueur moyenne des fragments incorporés était de 1247 pb, soit une longueur similaire à celle observée dans un protocole de transformation *in vitro* avec de l'ADN exogène (Kulick *et al.* 2008). La longueur plus courte observée par Falush *et al.* pourrait s'expliquer par le nombre limité d'évènements de recombinaison étudiés.

Récemment, le génome de 5 paires d'isolats de patients colombiens (intervalle d'isolement 3 à 16 ans) a été totalement séquencé. Les gènes codant pour les protéines OMP (famille *hop*) montraient le plus haut niveau de recombinaison quelque soit le patient. La sous-famille Hop regroupent de nombreuses protéines impliquées dans l'adhésion de la bactérie à l'épithélium gastrique (BabA, SabA...). Un mécanisme de sélection positive favorisant la survie des clones recombinants est fortement suspecté (Kennemann *et al.* 2011).

### 6.3.2. Au niveau macrodiversité

La première étude à avoir investigué les changements génétiques au cours d'une infection chronique est celle réalisée par Israel en 2001 (Israel *et al.* 2001). A partir d'isolats du même patient porteur de la souche J99 prélevés à six ans d'intervalle, les contenus génétiques des plus récents isolats ont été comparés à celui de la souche J99, dont le génome a été totalement séquencé (Alm *et al.* 1999). Ils ont montré que de nombreux changements sont apparus au cours de la colonisation, en particulier dans des régions hautement spécifiques du génome, le *cag* PAI et la PZ. Cependant, une des limites de cette étude est le faible nombre d'informations quantitatives obtenues sur la fréquence de ces changements dû au fait qu'un seul patient a été investigué et que très peu de ces changements génétiques détectés par puce à ADN ont été confirmés par séquençage. Plus récemment, le contenu génétique de 21 des 26 paires d'isolats préalablement caractérisées d'un point de vue microdiversité (Falush *et al.* 2001) a été investigué par puce à ADN (Kraft *et al.* 2006). Les gains ou les pertes de gènes étaient beaucoup plus rares (650 fois moins) que les remplacements alléliques par recombinaison homologue. Ce dernier mécanisme semble prépondérant pour l'adaptation de la bactérie aux changements dynamiques de l'environnement. De façon concordante avec l'étude sur la J99, la plupart des pertes et des gains de gènes concernaient le *cag* PAI et la PZ.

### 6.3.3. Rôle de la variabilité génétique dans la colonisation de l'estomac

Le modèle proposé par Kang et Blaser (Kang and Blaser 2006) est basé sur une co-évolution de la bactérie et de son hôte. Dans le modèle le plus simple, la bactérie produit des protéines effectrices telle la protéine CagA qui engendrent une réponse de l'hôte, diminuant la production de ces protéines par un système de rétrocontrôle négatif, et entraînant un équilibre stable entre l'hôte et la bactérie. Bien qu'une population clonale soit capable de persister chez l'hôte, la totalité de l'estomac ne sera pas colonisée (Figure 20A). La variabilité génétique serait un mécanisme clef pour permettre à la bactérie de coloniser de façon plus large l'estomac. La génération de clones variants entraîne une production variable de protéines effectrices, engendrant une réponse de l'hôte elle-même variable. Un équilibre dynamique entre les variants et l'hôte se met en place, constamment remanié et recréé. La capacité de colonisation de l'estomac est alors augmentée (Figure 20B). Les protéines effectrices peuvent chez un hôte entraîner la libération de molécules induisant des lésions de l'ADN telles que les radicaux libres. Ces molécules génèrent donc un stress pour la bactérie. Elle réagit en diversifiant son génome, ce qui permet d'augmenter la proportion d'estomac colonisé (Figure 20C). *H. pylori* possède des gènes impliqués dans la réparation de l'ADN soumis à une

variation de phase tel que *mutY*. Une absence de traduction de ces gènes (phénotype « off ») va entraîner la sélection de bactéries génétiquement hypervariables et donc une variabilité plus importante des protéines effectrices. L'augmentation de la variabilité permet une adaptation supérieure aux changements dynamiques de l'environnement. La population bactérienne peut ainsi coloniser de façon maximale l'estomac (Figure 20D). Bien que ce modèle se focalise sur les recombinaisons intragénomiques, des modèles parallèles peuvent être construits pour des recombinaisons intergénomiques. En conclusion, les auteurs avancent l'hypothèse que l'extraordinaire plasticité du génome de *H. pylori* est la clef pour comprendre l'extraordinaire capacité d'adaptation de cette bactérie à une multitude d'hôtes (plus de 50 % de la population mondiale est infectée), la colonisation de l'ensemble de l'estomac et la persistance de l'infection pendant des décennies.

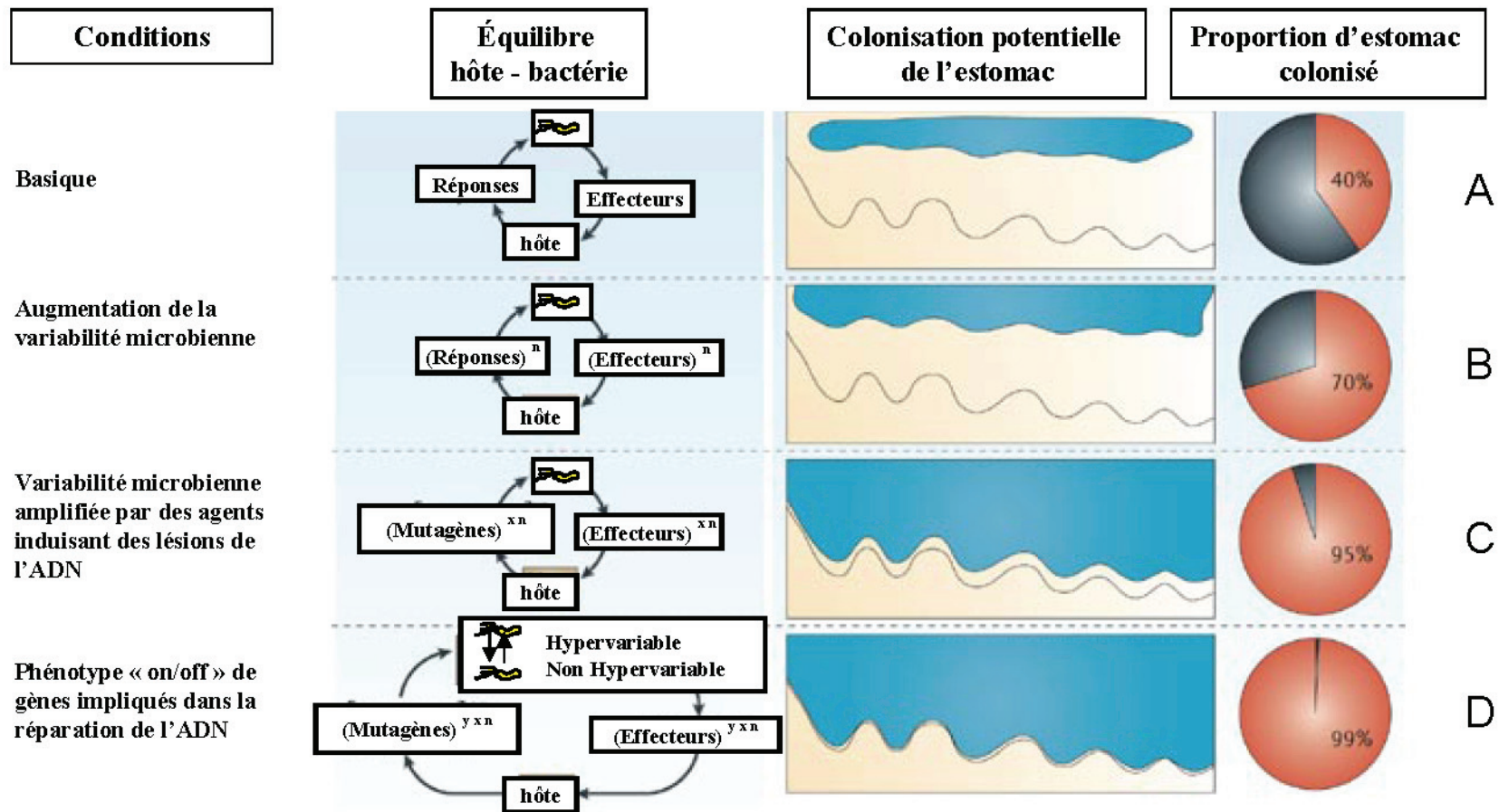


Figure 20. Rôle de la variabilité génétique dans la colonisation de l'estomac

D'après Kang et Blaser (Kang and Blaser 2006)

**PARTIE II. TRAVAIL EXPERIMENTAL**

## 1. PRESENTATION DES TRAVAUX DE RECHERCHE

A l'initiation de ce travail de recherche, les connaissances concernant le fond génétique et les facteurs de virulence des souches de *H. pylori* isolées en Afrique, en Océanie et en Asie du sud-est étaient limitées.

En 2002, une étude associant les Instituts Pasteur du Cambodge (Asie du sud-est), de Nouvelle-Calédonie (Océanie) et de Paris a été mise en place. L'objectif de ce travail était d'étudier: i) la pathologie gastro-duodénale associée à l'infection à *H. pylori* dans ces pays, ii) la sensibilité aux antibiotiques des souches isolées, iii) la distribution des gènes non-ubiquistes de *H. pylori* dans le génome de souches cliniques isolées de patients de diverses origines géographiques et ethniques par la méthode d'hybridation sur membranes à haute densité, iv) les facteurs de virulence et leur lien avec la pathologie gastro-duodénale. Ce projet a fait l'objet de mon DEA en 2004. Ma contribution avait consisté en un transfert de technologies au service de bactériologie de l'Institut Pasteur du Cambodge (culture bactérienne, identification, étude de la sensibilité aux antibiotiques et extraction de l'ADN) et à étudier la distribution des gènes non-ubiquistes de *H. pylori* au sein des isolats cliniques par hybridation sur membranes à haute densité à l'Institut Pasteur de Paris (fabrication et hybridation des membranes avec l'ADN génomique des souches cliniques isolées et analyse des résultats en collaboration avec Jean-Michel Thiberge de l'équipe du Professeur Agnès Labigne) (article en cours d'écriture et non inclus dans cette thèse).

En 2003, l'équipe du Professeur Mark Achtman (Institut Max Planck) montrait la très forte structure phylogéographique du génome de *H. pylori* (Falush *et al.* 2003). En 2007, cette même équipe montrait que *H. pylori* était déjà présent dans l'estomac des *Homo sapiens* il y a environ 58 000 ans, avant les grandes migrations de l'homme moderne, partant de l'Afrique de l'est vers l'Asie et l'Europe. La diversité génétique de *H. pylori* est maximale pour les souches isolées au nord-est de l'Afrique et décroît au fur et à mesure que la population se trouve éloignée de cette région, une caractéristique partagée par l'espèce humaine (Linz, Balloux *et al.* 2007). Ces données montrent qu'au sein d'une humanité originelle vivant en Afrique, la population ancestrale de *H. pylori* a colonisé l'homme suffisamment longtemps dans son berceau africain pour développer une importante diversité génétique. Une partie de la population humaine a ensuite migré depuis l'Afrique de l'est vers d'autres régions du globe, emportant avec elle une partie de la diversité génétique globale de la population ancestrale de *H. pylori*. Les bactéries présentes dans l'estomac des hommes se sont ensuite diversifiées de façon différente selon les populations en raison de leur isolement, de la dérive génétique et de

processus de sélection. Cette différenciation a généré un certain nombre de populations génétiques de *H. pylori* qui sont caractéristiques de larges zones géographiques, se superposant à celle des différentes populations humaines (Li *et al.* 2008). L'analyse de la diversité génétique de *H. pylori* a permis de retracer les mouvements des populations au cours des grandes migrations humaines (cf *Partie I. Généralités : chapitre 6.2.2. Microdiversité*).

Au total, 11 populations et sous-populations génétiques différentes de *H. pylori* ont été définies par analyse du polymorphisme génétique de 7 gènes de ménage (*atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI*, *yphC*) à l'aide de méthodes bayésiennes (logiciel STRUCTURE, modèle « no linkage ») (Annexe II) (Achtman *et al.* 1999; Falush *et al.* 2003; Devi *et al.* 2007; Linz *et al.* 2007; Moodley *et al.* 2009; Tay *et al.* 2009):

- hpEurope (présent en Europe, au Moyen-Orient, en Iran, en Inde et en Asie du sud-est),
- hpAfrica2 (présent en Afrique du Sud),
- hpNEAfrica (présent dans le nord-est de l'Afrique),
- hpAfrica1, subdivisée en hspSAfrica (présent en Afrique du Sud) et hspWAfrica (présent en Afrique de l'ouest et en Afrique du Sud),
- hpEastAsia, subdivisée en hspAmerind (présent chez les Amérindiens), hspEAsia (présent en Asie de l'est et du sud-est) et hspMaori (présent en Océanie),
- hpSahul (présent chez les Aborigènes d'Australie et de Nouvelle-Guinée).

Cependant, l'échantillon restreint de souches issues d'Afrique, d'Asie et d'Océanie n'avait pas permis d'explorer la diversité génétique de *H. pylori* dans ces parties du monde et d'y préciser les détails des migrations humaines. Afin de répondre à ces questions, nous avons mis en place une étude de la diversité génétique de souches de *H. pylori* en se basant sur les Instituts membres du Réseau International des Instituts Pasteur (Algérie, Cambodge, Grèce, Madagascar, Nouvelle-Calédonie, Sénégal) et l'Institut Louis Malardé à Tahiti. Ce travail a été réalisé en collaboration avec l'Institut Pasteur de Paris (Agnès Labigne, Sylvain Brisse, Michel Huerre) et l'Institut Max Planck à Berlin (Mark Achtman) dans le cadre d'un financement de l'Agence Nationale de la Recherche (Projet HELDIVNET 2006). Les deux principaux facteurs de virulence de *H. pylori*, CagA et VacA, ont été également investigués par des méthodes de biologie moléculaire avec principalement une étude du polymorphisme génétique du gène *vacA* et de l'extrémité 3' du gène *cagA* (financement de l'Institut Pasteur de Paris, Projet ACIP *Helicobacter pylori* 2006).

Mon rôle a été:

- de rédiger le projet en collaboration avec le Professeur Agnès Labigne (Institut Pasteur de Paris) et le Docteur Benoit Garin (Institut Pasteur de Madagascar) et de le soumettre aux appels d'offre,
- de coordonner la mise en place d'un réseau associant microbiologistes, anatomo-pathologistes et gastroentérologues sur chacun des sites,
- de mettre en place les procédures d'étude de cette bactérie au sein des laboratoires de bactériologie (culture, identification, étude de la sensibilité aux antibiotiques et extraction de l'ADN). Pour les laboratoires ne maîtrisant pas ces techniques, un stage a été organisé pour un technicien au sein du laboratoire de bactériologie de l'Hôpital Cochin à Paris (Dr Josette Raymond).
- de mettre en place les fiches de recueil des renseignements cliniques en collaboration avec les gastroentérologues,
- de mettre en place les fiches de recueil des renseignements anatomopathologiques en collaboration avec le Dr Michel Huerre (Unité de Recherche et d'Expertises en Histotechnologie et Pathologie, Institut Pasteur, Paris),
- d'organiser une seconde lecture de l'examen anatomopathologique sur l'ensemble des biopsies gastriques pour standardiser les résultats chez le Dr Michel Huerre,
- d'organiser la soumission du projet aux différents comités nationaux d'éthique,
- de soumettre le projet au Comité national Sénégalais d'éthique et à la Direction médicale de l'Institut Pasteur de Paris,
- de gérer au quotidien l'étude sur le site du Sénégal où j'exerce une activité de biologiste médical,
- d'organiser le transfert des ADN de l'ensemble des sites vers la Plateforme de Génotypage des Pathogènes et Santé Publique à l'Institut Pasteur de Paris (Sylvain Brisse et Valérie Caro) et vers l'Institut Pasteur de Dakar,
- de réaliser l'amplification et le séquençage des 7 gènes de ménage précédemment décrits en collaboration avec Jean-Michel Thiberge (Institut Pasteur de Paris) et Fatou Bintou Dieye (Institut Pasteur de Dakar),
- d'analyser les résultats en collaboration avec Bodo Linz, spécialiste en génétique de population, chercheur à l'Institut Max Planck jusqu'en 2008 puis à l'Université Park en Pennsylvanie,



- de réaliser l'étude du polymorphisme génétique du gène *vacA* et de l'extrémité 3' du gène *cagA* à l'Institut Pasteur de Dakar et l'analyse en collaboration avec Josette Raymond (Unité Postulante Pathogénèse de *Helicobacter*, Institut Pasteur de Paris).

Au total, plus de 800 patients ont été inclus sur l'ensemble des sites et 512 souches ont été collectées. L'examen anatomopathologique et la culture bactériologique des biopsies gastriques avec étude de la sensibilité aux antibiotiques des souches isolées par la méthode du E-test (amoxicilline, clarithromycine, métronidazole et lévofloxaïne) ont été pris en charge par le budget de l'étude, permettant la mise en place d'un traitement antibiotique adapté si besoin par le clinicien. Parmi ces souches, 300 ont été choisies comme étant représentatives de la pathologie gastro-duodénale associée et/ou de l'ethnie d'origine des patients afin d'étudier le polymorphisme génétique des 7 gènes de ménage et des facteurs de virulence *cagA* et *vacA* de *H. pylori*. Les résultats concernant les études menées en Nouvelle-Calédonie (**Article n°1**), au Cambodge (**Articles n°2 et 3**) et au Sénégal (**Article n°4**) seront présentés dans cette thèse. La mise en place de l'étude sur les sites d'Algérie, de Madagascar et Tahiti a été retardée, ne permettant pas d'y inclure les résultats. Ils seront publiés ultérieurement.

Les objectifs étaient:

- de documenter les migrations humaines en Océanie et en Asie du sud-est,
- de rechercher un lien entre les facteurs de virulence et la pathologie gastro-duodénale au Sénégal et au Cambodge.

Les résultats de ces travaux de recherche sont présentés dans les quatre articles suivants. Ils seront ensuite discutés.

**Article n°1:** Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, Wu JY, Maady A, Bernhöft S, Thiberge JM, Phuanukoonnon S, Jobb G, Siba P, Graham DY, Marshall BJ, Achtman M. The peopling of the Pacific from a bacterial perspective. *Science*. 2009. 323: 527-30.

**Article n°2:** Breurec S, Guillard B, Hem S, Brisse S, Dieye FB, Huerre M, Oung C, Sreng Tan T, Raymond J, Thiberge JM, Vong S, Monchy D, Linz B. Evolutionary history of *Helicobacter pylori* sequences reflects past human migrations in Southeast Asia. *PLoS ONE*. 2011. 6: e22058.

**Article n°3:** Breurec S, Guillard B, Hem S, Brisse S, Dieye FB, Huerre M, Monchy D, Oung C, Sreng Tan T, Thiberge JM, Vong S, Raymond J, Linz B. Host Range Expansion of Western CagA among East Asian *Helicobacter pylori* Strains in Cambodia. *Infection, Genetics and Evolution*. 2011 (Article sous presse).

**Article n°4:** Breurec S, Michel R, Seck A, Brisse S, Côme D, Dieye FB, Garin B, Huerre M, Mbengue M, Fall C, Sgouras DN, Thiberge JM, Dia D, Raymond J. Clinical Relevance of *cagA* and *vacA* Gene Polymorphisms in *Helicobacter pylori* Isolates from Senegalese Patients. *Clinical Microbiology and Infection*. 2011 (Article sous presse).

**Article n°1: The peopling of the Pacific from a bacterial  
perspective**

Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, Wu JY,  
Maady A, Bernhöft S, Thiberge JM, Phuanukoonnon S, Jobb G, Siba  
P, Graham DY, Marshall BJ, Achtman M.

Science. 2009. 323: 527-30.

## Contexte

Deux vagues de migrations humaines ont été décrites en Océanie:

- une migration préhistorique de l'Asie vers le Sahul (Sahul = continent constitué de l'Australie, la Nouvelle Guinée et la Tasmanie réunis par le faible niveau des mers, au moment d'une période glaciaire). Cependant, le nombre présumé de migrations ainsi que la période de migration sont le sujet de débats au sein de la communauté scientifique.
- une deuxième vague de migration plus tardive, de peuples aborigènes parlant des langues austronésiennes vers la Mélanésie puis la Polynésie, sans consensus sur l'origine géographique de cette migration.

Dans le cadre du Projet HELDIVNET 2006, une collaboration a été mise en place entre plusieurs Instituts dont l'objectif était de retracer les migrations humaines en Océanie:

- l'Institut Pasteur de Nouvelle-Calédonie, responsable de l'inclusion de souches isolées chez des Mélanésiens de Nouvelle-Calédonie ainsi que chez des Polynésiens de Wallis et Futuna vivant en Nouvelle-Calédonie (Responsable Dr Benoit Garin).
- l'Institut Max Planck (Allemagne), responsable de l'inclusion de souches isolées chez des Polynésiens aux Iles Tonga/Samoa, aux Philippines, en Nouvelle-Zélande et chez des Mélanésiens de Nouvelle-Guinée.
- l'Université de Houston (Etats-Unis), responsable de l'inclusion de souches isolées chez des aborigènes à Taiwan,
- l'Université de Perth (Australie), responsable de l'inclusion de souches isolées chez des aborigènes australiens.

Ma contribution a porté sur l'étude de cette seconde migration avec : i) l'amplification et le séquençage des 7 gènes de ménage précédemment décrits à partir de l'ADN des souches isolées à l'Institut Pasteur de Nouvelle-Calédonie, ii) l'analyse des résultats en collaboration avec l'équipe de l'Institut Max Planck à Berlin.

## Matériels et méthodes

Au total, 254 souches ont été isolées chez des aborigènes provenant d'Australie (n=92), des hauts plateaux de Nouvelle-Guinée (n=27), de Taiwan (n=74), d'une île du détroit de Torres (n=2), des Philippines (n=13), de Nouvelle-Calédonie (n=13), de Wallis et Futuna (n=5), des Iles Tonga/Samoa (n=10) et de Nouvelle-Zélande (n=18). Les 7 gènes de ménage précédemment décrits (*atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI*, *yphC*) ont été amplifiés et séquencés, puis les souches ont été assignées à des (sous-) populations génétiques à l'aide de méthodes bayésiennes (Logiciel STRUCTURE, modèle « no linkage ») (Annexe II).

La détermination du nombre présumé de migrations et l'estimation de la période de migration ont été réalisées avec les logiciels IMA et CLONAL FRAME. Le logiciel IMA simule un modèle de séparation géographique initiale entre deux populations suivie par des migrations occasionnelles dans les deux directions. Il permet ainsi d'estimer les quantités et la direction des flux de gènes entre deux populations génétiques. CLONAL FRAME est un logiciel d'analyse phylogénétique utilisant des méthodes bayésiennes avec un modèle de coalescence neutre. Il est parfaitement adapté à l'étude de la diversité génétique de *H. pylori* puisqu'il prend en compte les événements de recombinaison.

## **Résultats et Discussion**

### **Première vague de migration**

Durant la dernière glaciation (entre 10 000 et 41 000 ans avant JC), l'Asie du sud-est insulaire n'était pas l'ensemble archipélagique qui existe aujourd'hui mais constituait un plateau continental, celui du Sunda, prolongement en forme de péninsule du continent asiatique. En effet, le niveau de la mer était beaucoup plus bas qu'aujourd'hui (100 à 150 mètres par rapport au niveau actuel). L'Australie, la Nouvelle-Guinée et la Tasmanie formaient un seul continent, appelé Sahul. Il semble que les hommes aient navigué sur de petites distances à partir des îles de l'actuelle Indonésie pour atteindre le Sahul. Puis, par voie terrestre, ils se seraient dispersés sur l'ensemble du continent.

L'analyse de la diversité génétique des souches de *H. pylori* a permis de définir une nouvelle population génétique, hpSahul, incluant 50 souches. Cette population est considérée comme le marqueur de cette migration humaine préhistorique puisqu'elle est retrouvée uniquement chez les peuples vivant en Australie et en Nouvelle-Guinée (Figure 1A, p528 de l'article). Le peuplement de l'Australie et de la Papouasie-Nouvelle-Guinée est lié à un événement migratoire unique par un groupe d'individus resté dans un isolement substantiel ou total jusqu'à une époque récente (Figure 2A, p529 de l'article). Ceci recoupe parfaitement les données d'une analyse récente de l'ADN mitochondrial (ADNmt) et du chromosome Y d'aborigènes australiens, de Mélanésiens et de néo-guinéens, mais également certaines données archéologiques.

Le taux de mutation de *H. pylori* étant inconnu, la datation moléculaire a consisté à coupler l'arbre phylogénétique obtenu par les logiciels IMA et CLONAL FRAME, décrivant les relations de parenté entre isolats appartenant aux différentes (sous-) populations génétiques de *H. pylori*, avec des points de calibration. Ils correspondent aux dates présumées de séparation

des différentes (sous-) populations génétiques et permettent d'estimer l'âge absolu de divergence entre la population hpSahul et les autres populations d'Asie (hpAsia2, hspAmerind, hspEAsia, hspMaori) (Figure 2B, Figure 3, p529 de l'article). Cette migration a été datée entre 31 000 à 37 000 ans avant notre époque par l'équipe de l'Institut Max Planck, ce qui est en accord avec l'âge estimé de l'haplogroupe Q de l'ADNmt, commun aux peuples du Sahul. Ceci indique que la date présumée des plus anciens restes humains, entre 40 000 et 50 000 ans avant notre époque, est fautive.

### **Deuxième migration**

La deuxième grande migration de peuples en Océanie est le peuplement du Pacifique survenu il y a environ 5000 ans. Son origine est le sujet de débats au sein de la communauté scientifique. Les langues austronésiennes s'étendent sur une aire géographique allant de Madagascar à l'ouest à l'île de Pâques à l'est, et de Taiwan et Hawaï au nord à la Nouvelle-Zélande au sud. La génétique linguistique, qui vise à reconstituer la filiation à travers le temps d'un ensemble de langues apparentées entre elles, tend à situer Taiwan comme le berceau de ces langues, théorie appuyée de façon inconstante par l'étude des marqueurs humains traditionnels.

En analysant le polymorphisme génétique des 7 gènes de ménage selon la méthodologie précédemment décrite, 77 des 254 souches étaient assignées à la sous-population génétique hspMaori (population hpEAsia). Cette population est considérée comme étant un marqueur de cette migration puisqu'elle est retrouvée uniquement chez les peuples vivant à Taiwan, en Mélanésie et en Polynésie (Figure 1A, p528 de l'article). Les autres haplotypes retrouvés en Océanie (Tableau S1 de l'article) étaient assignés aux populations hpEurope, hspEAsia et hpAfrica1, témoins de migrations récentes.

L'analyse de la diversité génétique de *H. pylori* est en faveur d'une migration à partir de Taiwan :

- la sous-population hspMaori est présente chez 5 des 6 tribus incluses à Taiwan avec des fréquences allant de 44 à 100 %,
- la diversité génétique des souches est significativement plus élevée à Taiwan ( $\Pi_{95} = 1.79$  to  $1.82$  %) qu'en dehors de Taiwan ( $\Pi_{95} = 1.58$  to  $1.62$  %) (Logiciel DnaSP4), indiquant que cette migration a emporté avec elle une partie de la diversité génétique globale de la population hspMaori présente à Taiwan,

- l'analyse phylogénétique des souches hspMaori par le logiciel CLONAL FRAME montre une migration séquentielle à partir de Taiwan, vers les Philippines, la Mélanésie et la Polynésie (Figure 1B, p528 de l'article),

- cette migration a été unidirectionnelle après analyse par le logiciel IMA (Figure 3A, p529 de l'article).

La sous-population hspMaori est par conséquent le reflet de l'expansion unidirectionnelle en Océanie de peuples parlant des langages austronésiens, hébergeant *H. pylori* dans leur estomac, à partir de Taiwan, via les Philippines, il y a environ 5000 ans.

### **Conclusion**

Cette étude a permis de retracer deux vagues de migration en Océanie, une migration préhistorique vers la Nouvelle-Guinée et l'Australie et une migration plus récente à partir de Taiwan dans le Pacifique. Elle confirme la résolution plus élevée de la diversité génétique de *H. pylori* pour retracer les anciennes migrations humaines par comparaison aux marqueurs génétiques humains traditionnels.

G and H, and table S5). Because *mil/SIP2* is expressed in the mesoderm just lateral to the midline (4), *mil/SIP2* is proposed to function in mesoderm over the YSL. Further, cardia bifida in the *spns2<sup>ko157</sup>* mutant was restored by the injection of *spns2* mRNA but not *spns2(R153S)* mRNA into the YSL at shield stage (Fig. 4, I to L, and table S6).

The rescue frequency by injection of *spns2* mRNA into the YSL was slightly lower than for injection into the blastomere (tables S2 and S6). One explanation is that *spns2* mRNA injected into the blastomere at the one-cell stage is widely distributed in the YSL because the YSL is constituted by marginal blastomeres collapsing onto the yolk around the 1000-cell stage. Another explanation is that the function of Spns2 in embryonic tissues as well as in the YSL may be partly required for the migration of myocardial precursors. Furthermore, transplantation analysis showed that Spns2 at least functions in a cell-nonautonomous manner, because *ko157*-derived donor cells were incorporated into single beating hearts of wild-type recipients, and wild type-derived donor cells were incorporated into one of two beating hearts of *ko157* recipients (movies S1 to S3). One attractive interpretation is that Spns2 in the YSL regulates the SIP export from the yolk to the embryonic body, leading to the activation of Mil/SIP<sub>2</sub> in mesoderm just lateral to the midline (fig. S1). Recent reports have pointed out the importance of *ferroportin1* (*fpn1*) as a transporter of iron from the yolk to the embryonic body (21) and the clinical relevance to hypochromic anemia and hemochromatosis in humans (22, 23).

By investigating characteristic features of the zebrafish *spns2<sup>ko157</sup>* mutant and analyzing the biological activity of Spns2, we have demonstrated that Spns2 functions as a SIP transporter and that Spns2 in the extraembryonic YSL is a prerequisite for the migration of myocardial precursors, presumably mediated by the SIP-Mil/SIP<sub>2</sub> pathway. The identification of Spns2 not only contributes to our understanding of the molecular mechanism of biological SIP delivery, but may also elucidate the physiological importance of Spns2 in autoimmune disease (24), cardiovascular diseases, and cancer (25) in which SIP plays a central role.

#### References and Notes

1. D. Y. Stainier, *Nat. Rev. Genet.* **2**, 39 (2001).
2. J. J. Schoenebeck, D. Yelon, *Semin. Cell Dev. Biol.* **18**, 27 (2007).
3. J. N. Chen, M. C. Fishman, *Trends Genet.* **16**, 383 (2000).
4. E. Kupperman, S. An, N. Osborne, S. Waldron, D. Y. Stainier, *Nature* **406**, 192 (2000).
5. M. J. Lee *et al.*, *Science* **279**, 1552 (1998).
6. T. Hla, *Pharmacol. Res.* **47**, 401 (2003).
7. S. Spiegel, S. Milstien, *Nat. Rev. Mol. Cell Biol.* **4**, 397 (2003).
8. Y. A. Hannun, L. M. Obeid, *Nat. Rev. Mol. Cell Biol.* **9**, 139 (2008).
9. T. Matsui *et al.*, *Nat. Clin. Pract. Cardiovasc. Med.* **4** (suppl. 1), S77 (2007).
10. Y. Nakano *et al.*, *Mol. Cell. Biol.* **21**, 3775 (2001).
11. R. M. Young *et al.*, *Dev. Dyn.* **223**, 298 (2002).
12. Y. Huang, M. J. Lemieux, J. Song, M. Auer, D. N. Wang, *Science* **301**, 616 (2003).
13. N. Kobayashi *et al.*, *J. Lipid Res.* **47**, 614 (2006).
14. P. Mitra *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16394 (2006).
15. K. Sato *et al.*, *J. Neurochem.* **103**, 2610 (2007).
16. E. B. Neufeld *et al.*, *J. Biol. Chem.* **276**, 27584 (2001).
17. Y. M. Lee, K. Venkataraman, S. I. Hwang, D. K. Han, T. Hla, *Prostaglandins Other Lipid Mediat.* **84**, 154 (2007).

18. Y. Kikuchi *et al.*, *Genes Dev.* **15**, 1493 (2001).
  19. T. Dickmeis *et al.*, *Genes Dev.* **15**, 1487 (2001).
  20. T. Sakaguchi, Y. Kikuchi, A. Kuroiwa, H. Takeda, D. Y. Stainier, *Development* **133**, 4063 (2006).
  21. A. Donovan *et al.*, *Nature* **403**, 776 (2000).
  22. O. T. Njajou *et al.*, *Nat. Genet.* **28**, 213 (2001).
  23. G. Montosi *et al.*, *J. Clin. Invest.* **108**, 619 (2001).
  24. M. Sekiguchi *et al.*, *J. Immunol.* **180**, 1921 (2008).
  25. S. Milstien, S. Spiegel, *Cancer Cell* **9**, 148 (2006).
26. We thank M. Sone and M. Minamimoto for technical assistance; S. Endo and R. Hanaoka for the mutagenesis screening; T. Kitaguchi for the establishment of *cmlc2:mRFP*; M. Masuda, Y. Igarashi, A. Kihara, S. Mitsutake, K. Shioya, I. B. Dawid, and Y. Kaziro for support and suggestions; M. Hibi, J. S. Gutkind, and M. Tsang for critical reading and valuable suggestions; D. Y. Stainier for discussions and sharing of unpublished data; and H. Okamoto, K. Kawakami, Y. Kikuchi, H. Yanagisawa, D. Yamamoto, R. Y. Tsien, and H.-J. Tsai for fish and reagents. Supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan; the National BioResource Project of Japan; the Japan Society for the Promotion of Science; the Program for the Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation; the Ministry of Health, Labor and Welfare of Japan; and the Takeda Science Foundation. The sequences of zebrafish *spns2*, human *Spns2*, and mouse *Spns2* have been deposited in the DNA Data Bank of Japan [accession numbers: AB441164 (zebrafish), AB441165 (human), AB441166 (mouse)]. The authors are filing a patent based on the results reported in this paper.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/1167449/DC1  
Materials and Methods

Figs. S1 to S19

Tables S1 to S6

Movies S1 to S3

References

20 October 2008; accepted 3 December 2008

Published online 11 December 2008;

10.1126/science.1167449

Include this information when citing this paper.

## The Peopling of the Pacific from a Bacterial Perspective

Yoshan Moodley,<sup>1\*†</sup> Bodo Linz,<sup>1\*‡</sup> Yoshio Yamaoka,<sup>2\*</sup> Helen M. Windsor,<sup>3</sup> Sebastien Breurec,<sup>4,5</sup> Jeng-Yih Wu,<sup>6</sup> Ayas Maady,<sup>7</sup> Steffie Bernhöft,<sup>1</sup> Jean-Michel Thiberge,<sup>8</sup> Suparat Phuanukoonnon,<sup>9</sup> Gangolf Jobb,<sup>10</sup> Peter Siba,<sup>9</sup> David Y. Graham,<sup>2</sup> Barry J. Marshall,<sup>3</sup> Mark Achtman<sup>1,11§</sup>

Two prehistoric migrations peopled the Pacific. One reached New Guinea and Australia, and a second, more recent, migration extended through Melanesia and from there to the Polynesian islands. These migrations were accompanied by two distinct populations of the specific human pathogen *Helicobacter pylori*, called hpSahul and hspMaori, respectively. hpSahul split from Asian populations of *H. pylori* 31,000 to 37,000 years ago, in concordance with archaeological history. The hpSahul populations in New Guinea and Australia have diverged sufficiently to indicate that they have remained isolated for the past 23,000 to 32,000 years. The second human expansion from Taiwan 5000 years ago dispersed one of several subgroups of the Austronesian language family along with one of several hspMaori clades into Melanesia and Polynesia, where both language and parasite have continued to diverge.

After modern humans dispersed “out of Africa” about 60,000 years ago (60 ka) (1), they reached Asia via a southern coastal route (2). That route extended along the Pleistocene landmass, known as Sundaland (i.e., the Malay peninsula, Sumatra, Java, Borneo, and

Bali), that was joined to the Asian mainland as a result of low sea levels during the last ice age (12 to 43 ka) (3). Low sea levels also meant that Australia, New Guinea, and Tasmania were connected in a continent called Sahul, separated from Sundaland by a few narrow deep-sea channels. It

seems Sahul was colonized only once, ~40 to 50 ka (3, 4), although backed-blade stone tool technology and the dingo appear to have been introduced from India at a later date (5, 6).

<sup>1</sup>Max-Planck-Institut für Infektionsbiologie, Department of Molecular Biology, Charitéplatz 1, 10117 Berlin, Germany. <sup>2</sup>Department of Medicine-Gastroenterology, Baylor College of Medicine and Michael E. DeBakey VA Medical Center, Houston, TX 77030, USA. <sup>3</sup>Microbiology and Immunology M502, School of Biomedical, Biomolecular and Chemical Sciences, University of Western Australia, Australia 6009. <sup>4</sup>Institut Pasteur, BP 220, Dakar, Sénégal. <sup>5</sup>Institut Pasteur de Nouvelle-Calédonie, BP61, 98845 Noumea, New Caledonia. <sup>6</sup>Department of Gastroenterology, Kaohsiung Medical University, 100 Shih-Chuan 1st Road, Kaohsiung 80708, Taiwan. <sup>7</sup>Department of Endoscopy, Republic Hospital No. 1, Kyzyl City 667003, Republic of Tuva, Russia. <sup>8</sup>Institut Pasteur, Genotyping of Pathogens and Public Health, 28 rue du Dr Roux, 75724 Paris, France. <sup>9</sup>Papua New Guinea Institute of Medical Research, Post Office Box 60, Goroka, EHP, 441 Papua New Guinea. <sup>10</sup>Fritz-Kortner-Bogen 36, 81739 Munich, Germany. <sup>11</sup>Environmental Research Institute and Department of Microbiology, University College Cork, Cork, Ireland.

\*These authors contributed equally to this work.

†Present address: Austrian Academy of Sciences, Konrad Lorenz Institute for Ethology, Savoyenstrasse 1A, A-1160, Vienna, Austria.

‡Present address: Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA.

§To whom correspondence should be addressed. E-mail: m.achtman@ucc.ie



Human genetic data are compatible with these interpretations, but have not provided the details. Redd and Stoneking identified multiple mitochondrial DNA (mtDNA) lineages among New Guinea peoples with coalescence times of 80,000 to 122,000 years (80 to 122 ky), predating the out-of-Africa migrations (5). In subsequent analyses, Australian aboriginals and Melanesians fell into multiple, distinct mtDNA haplogroups interdispersed among lineages from East Asia and India (4), with one exception: haplogroup Q, which had a coalescent estimate of 32 ka and contained both Australian and Melanesian lineages. Y-chromosome markers yielded one lineage for Australians and a second one for Melanesians (4). Australia and New Guinea remained connected by a land bridge until sea levels rose ~8 to 12 ka, and it is surprising that the native inhabitants of Sahul are not genetically associated except for haplogroup Q.

Subsequent prehistoric migrations to island East Asia and the Pacific have been designated differently depending on whether they were traced by language, archaeological remains, or genetic studies. Most of the native Pacific languages from near the African coast (Madagascar) through to Polynesia are Malayo-Polynesian, a subgroup of the Austronesian language family (7). The nine other subgroups of Austronesian are only spoken in Taiwan, suggesting that Taiwan is the origin of Austronesian (7). In support of this interpretation, agriculturists spread from Taiwan via insular and coastal Melanesia into the Pacific, as marked by the Lapita cultural complex, including red-slipped pottery, Neolithic tools, chickens, pigs, and farm-

ing (8). A human genetic marker of this route of spread is the “Polynesian” mtDNA HV1 motif of lineage B4a1a, which is found at high frequency among native Taiwanese (9), Melanesians, and Polynesians (10, 11).

We attempted to trace human prehistory in the Pacific by analyzing the distribution of a bacterial parasite of humans, *Helicobacter pylori*. *H. pylori* accompanied modern humans during their migrations out of Africa (12). Subsequent founder effects, plus geographic separation, have resulted in populations of bacterial strains specific for large continental areas. Thus, Africans are infected by the *H. pylori* populations hpAfrica1 and hpAfrica2, Asians are infected by hpAsia2 and hpEastAsia, and Europeans are infected by hpEurope (12, 13). It seemed possible that the distribution of *H. pylori* genotypes among native inhabitants might provide insights into migrations throughout the Pacific. We cultivated 212 bacterial isolates from gastric biopsies or mucus obtained from aboriginals in Taiwan and Australia, highlanders in New Guinea, as well as Melanesians and Polynesians in New Caledonia (table S1). Concatenated sequences of seven gene fragments (3406 base pairs, of which half are polymorphic) from these isolates yielded 196 unique haplotypes. These were compared with 99 unique haplotypes from 100 Europeans in Australia and 222 other unique haplotypes from Asia and the Pacific, including 15 haplotypes from Chinese inhabitants of Taiwan, as well as ~1700 haplotypes from other sources.

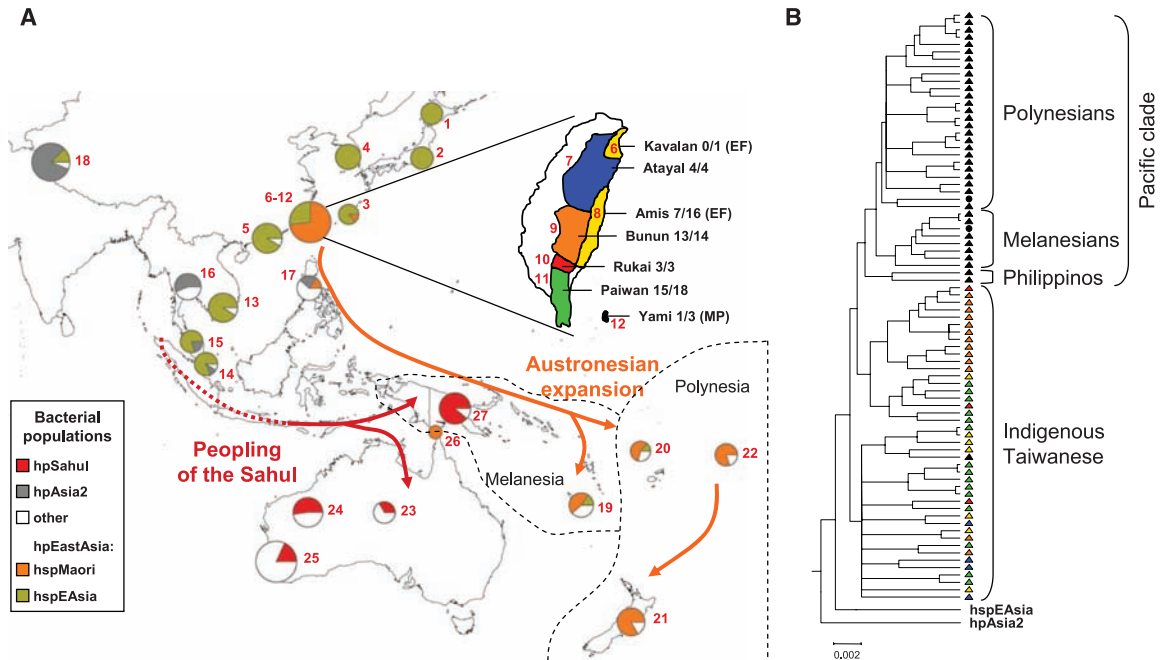
According to Bayesian assignment analysis, our samples from native inhabitants yielded 50

unique haplotypes that formed a distinct biogeographic group called hpSahul (14). Twenty-eight percent (26 of 92) of the haplotypes from aboriginals in Australia and 89% (24 of 27) of the haplotypes from highlanders in New Guinea were hpSahul (Fig. 1A). One hpSahul haplotype was found among 99 haplotypes from Europeans in Australia and none among the other haplotypes from elsewhere.

hspMaori is a subpopulation of hpEastAsia, isolated from Polynesians (Maoris, Tongans, and Samoans) in New Zealand (13) and three individuals in the Philippines and Japan. hspMaori isolates have not previously been isolated from other individuals, including the 15 Chinese inhabitants of Taiwan (12). Fifty-four of the 196 unique haplotypes from native inhabitants were hspMaori (14), and all came from Austronesian sources. These included native Taiwanese (43 of 59, 73%), Melanesians (6 of 13, 46%), and Polynesians (3 of 5, 60%) in New Caledonia, and two inhabitants of the Torres Straits islands that lie between Australia and New Guinea and which have been visited extensively by Polynesians (Fig. 1A and table S1). These observations suggest that hspMaori is a marker for the entire Austronesian expansions rather than only for Polynesians. The remaining unique haplotypes from native inhabitants were hpEurope, hspEAsia, and hpAfrica1, which can be attributed to very recent human travels.

If Taiwan were the source of the Austronesian expansions, hspMaori haplotypes would be expected to be widespread among aboriginal Taiwanese tribes. Indeed, hspMaori was isolated

**Fig. 1. (A)** The distribution of *H. pylori* populations in Asia and the Pacific. The proportions of haplotypes at each sampling location (red numbers; table S1) that were assigned to different bacterial populations are displayed as pie charts whose sizes indicate the numbers of haplotypes. The geographic location of Melanesia and Polynesia is depicted. The term “Austronesia” refers to the entire region inhabited by Austronesian-speaking people from Madagascar through to the Easter Islands. **(Inset)** A detailed map of Taiwan showing the distribution of aboriginal tribes. The names of the tribes plus the proportion of hspMaori haplotypes among all haplotypes are shown in black at the right. The language-family designations are the same as the tribal names except where indicated by parentheses (EF, East Formosan; MP, Malayo-Polynesian). **(B)** Phylogenetic relationships among hspMaori strains (80% consensus of 100 ClonalFrame analyses). One haplotype each of



hpAsia2 and hspEAsia was used to root the tree. Strains are color-coded according to Austronesian language family in (A). Two black circles within the Pacific clade indicate haplotypes isolated from the Torres Strait islands, and a black triangle among indigenous Taiwanese indicates an hspMaori haplotype from Yami.

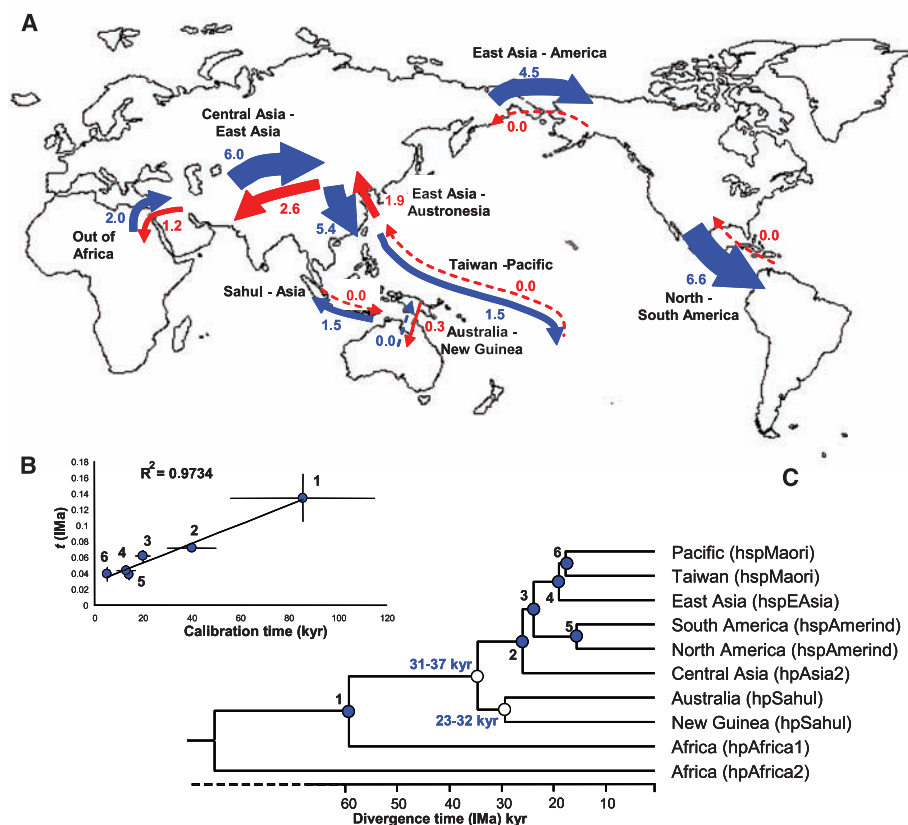
frequently (44 to 100%) from five of the six tribes sampled (Fig. 1A). Taiwan should also harbor the greatest diversity, and the branching order within a phylogenetic tree should reflect the direction of subsequent migrations. The phylogenetic analyses showed that genetic diversity was significantly higher in Taiwanese hspMaori ( $\Pi_{95} = 1.79$  to 1.82%) than in non-Taiwanese hspMaori ( $\Pi_{95} = 1.58$  to 1.62%). All non-Taiwanese hspMaori haplotypes form a single clade, the Pacific clade, which originates from one of several clades among indigenous Taiwanese haplotypes (Fig. 1B). The sequence of branching events within the Pacific clade is consistent with sequential migrations from Taiwan via the Philippines and island Melanesia to Polynesia (Fig. 1B). These results

also support an association between language and haplotype group. The indigenous Taiwanese haplotypes were isolated from tribes that speak 5 of the 10 subgroups of the Austronesian family of languages, whereas the Pacific clade was isolated from individuals that speak variants of Malayo-Polynesian. The sole exception to these generalizations was one haplotype from the Yami of Lanyu, a small island off the coast of Taiwan, where the language is a variant of Malayo-Polynesian but the haplotype clustered with the indigenous Taiwanese haplotypes. Together, these observations provide support for a Taiwanese source of the Austronesian expansions.

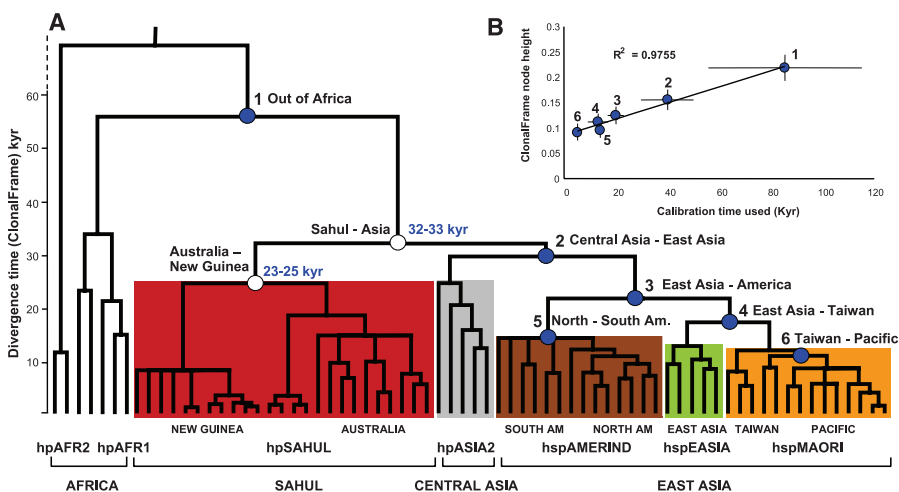
Using the isolation with migration model (IMa), we calculated the magnitude of migrations

in both directions after the initial split between the Taiwan and Pacific clades of hspMaori (15). IMa uses sequence data within a probabilistic framework to simulate a model of initial geographic separation between two populations followed by occasional migration in both directions. Because homologous recombination is frequent within *H. pylori* (13, 16), we excluded blocks of sequences that had a high likelihood of recombination (14). The calculations indicated that migrations subsequent to the initial split were unidirectional, from Taiwan to the Pacific (Fig. 2A).

Other splits between pairs of *H. pylori* populations were also unidirectional: for example, the Amerind colonization of the Bering Strait and the subsequent colonization of the South



**Fig. 2.** Global patterns of migration between eight pairs of *H. pylori* populations as calculated by the isolation with migration model (IMa). (A) Map. The magnitudes of migration are denoted by numbers and arrow thickness and their direction is indicated in blue or red. (B) Graph showing a linear relation between the calibration time (table S2) of six events (filled blue circles) that are dated by archaeological estimates and the estimated time (*t*). (C) Population tree reconstructed from a consensus of 1000 bootstrap samples from the range of calculated *t* values to determine the ages of nodes (thousands of years, kyr) associated with the peopling of the Sahul (unfilled circles). Ages (in light blue) are the 95% confidence limits of estimated coalescence times obtained by applying global rate minimum deformation (GRMD) rate-smoothing, as implemented in Treefinder, to the range of *t* values within the limits of calibration dates.



**Fig. 3.** Global phylogeny of *H. pylori* as calculated by a haplotype approach based on the 80% consensus of 100 ClonalFrame analyses. (A) Phylogenetic tree of divergence time, as indicated by node height versus geographic sources (bottom line) and population assignments (second line). Detailed sources of clades within populations are indicated in the third line from the bottom. Node heights were used to date the two hpSahul nodes (unfilled circles) based on six calibration times (filled blue circles, table S2). Age ranges (light blue numbers) are the 95% confidence limits of estimated coalescence times obtained with GRMD rate-smoothing over the range of node height values and calibration time limits. hpAFR2, hpAfrica2; hpAFR1, hpAfrica1; AM, America. (B) Graph showing a linear relation of calibration time with the range of heights for each node.

America from North America. However, migrations out of Africa, from Central to East Asia, and from East Asia to Taiwan were followed by appreciable levels of return migration (Fig. 2A).

Molecular mutation rates are unknown for most bacteria, so we cannot directly use IMA data to calculate the dates of initial splits. Instead, we calibrated against known dates for splits among human populations. The archaeologically attributed split between Taiwan and the Pacific Clade is 5 ka (8). Five other calibration dates are presented in table S2. The time when populations split ( $t$ ) calculated by IMA varied linearly with the calibration dates (Fig. 2B). We used random values within the range of five  $t$  values that were calculated for each split between all pairs of populations (table S2) to construct 1000 bootstrap trees using Treefinder (17). These trees were then used to calculate the age of the Sahulian migration by rate-smoothing within the limits of the six calibration dates (14).

The dates and numbers of migrations to the Sahul are controversial. According to our IMA calculations, the population split leading to hpSahul postdated the out-of-Africa migrations but predated the splits that resulted in hpAsia2 (found in Central Asia) and hpEastAsia [East Asia (hspEAsia); the Pacific (hspMaori); the Americas (hspAmerind)]. The 95% confidence limits of the date of the split between hpSahul and the Asian populations were estimated as 31 to 37 ka and the split between hpSahul in New Guinea and Australia as 23 to 32 ka. The combined data presented here indicate that hpSahul migrated only once from Asia toward Sahul, and once between New Guinea and Australia, and subsequent migration did not occur from Australia to New Guinea (Fig. 2A).

To verify the use of IMA for dating of population splits in a bacterial species like *H. pylori*, we also used a haplotype-based coalescent approach, which accounts for recombination with unrelated sources of DNA, as implemented in the program ClonalFrame (18). ClonalFrame generated a haplotype tree whose branch order agreed with the population tree generated by IMA (Fig. 3A). It also assigned individual haplotypes to clades that are congruent with the population assignments, including the separation between hpSahul and other populations. The observation that all hpSahul strains clustered in a monophyletic clade verifies a single colonization event and confirms that modern Asians and the inhabitants of the Sahul have undergone independent evolutionary trajectories since they first split. The two hpSahul clades in New Guinea and Australia are also distinct, confirming a lack of migration between the two areas.

Similarly to the IMA analyses, we observed a linear relation between the calibration dates and time of splitting calculated by ClonalFrame as node heights (Fig. 3B). Applying the same rate-smoothing calibration method as above, we estimated that hpSahul split from the Asian population 32 to 33 ka. Subsequently, hpSahul

from New Guinea and Australia split 23 to 25 ka. Both estimates overlap with the range of IMA estimates (31 to 37 ka and 23 to 32 ka, respectively). The date of origin of hpSahul is comparable to the estimated age of 32 ka for the Q mtDNA haplogroup (4), but less than the 40 to 50 ky associated with the oldest archaeological finding of human artefacts in Australia (3).

Our results lend support for two distinct waves of migrations into the Pacific. First, early migrations to New Guinea and Australia accompanied by hpSahul and second, a much later dispersal of hspMaori from Taiwan through the Pacific by the Malayo-Polynesian-speaking Lapita culture. Each sampling area yielded either hpSahul or hspMaori haplotypes, but not both. The lack of overlap between these populations may reflect differential fitness of the parasite, as has been inferred for the modern replacement of hspAmerind haplotypes by European and African *H. pylori* in South America (19, 20). Alternatively, hpSahul and hspMaori may still coexist in unsampled islands of East Asia, Melanesia, and coastal New Guinea, where their identification might help to unravel the details of human history in those areas.

#### References and Notes

- H. Liu, F. Prugnolle, A. Manica, F. Balloux, *Am. J. Hum. Genet.* **79**, 230 (2006).
- V. Macaulay et al., *Science* **308**, 1034 (2005).
- K. O. Pope, J. E. Terrell, *J. Biogeography* **35**, 1 (2008).
- G. Hudjashov et al., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 8726 (2007).
- A. J. Redd, M. Stoneking, *Am. J. Hum. Genet.* **65**, 808 (1999).
- P. Savolainen, T. Leitner, A. N. Wilton, E. Matisoo-Smith, J. Lundeberg, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12387 (2004).
- R. D. Gray, F. M. Jordan, *Nature* **405**, 1052 (2000).

- M. J. T. Spriggs, in *Prehistoric Mongoloid Dispersals*, T. Akazawa, E. Szathmari, Eds. (Oxford Univ. Press, Oxford, 1996), pp. 322–346.
- J. A. Trejaut et al., *PLoS Biol.* **3**, e247 (2005).
- T. Melton et al., *Am. J. Hum. Genet.* **57**, 403 (1995).
- B. Sykes, A. Leiboff, J. Low-Beer, S. Tetzner, M. Richards, *Am. J. Hum. Genet.* **57**, 1463 (1995).
- B. Linz et al., *Nature* **445**, 915 (2007).
- D. Falush et al., *Science* **299**, 1582 (2003).
- Materials and methods and supplementary tables, figures, and scripts are available on Science Online.
- J. Hey, R. Nielsen, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2785 (2007).
- D. Falush et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 15056 (2001).
- G. Jobb, A. von Haeseler, K. Strimmer, *BMC Evol. Biol.* **4**, 18 (2004).
- X. Didelot, D. Falush, *Genetics* **175**, 1251 (2007).
- Y. Yamaoka et al., *FEBS Lett.* **517**, 180 (2002).
- C. Ghose et al., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15107 (2002).
- We gratefully acknowledge C. Stamer for technical assistance, F. Balloux and D. Falush for helpful discussions, and J. Hey for advice on IMA. Support was provided by grants from the ERA-NET PathoGenoMics (project HELDIVNET) to M.A. and S.B., the Science Foundation of Ireland (05/FE1/B882) to M.A., the NIH (grant R01 DK62813) to Y.Y., and the Institut Pasteur and the Institut de Veille Sanitaire to J.-M.T. This publication made use of the *Helicobacter pylori* Multi Locus Sequence Typing Web site (<http://pubmlst.org/helicobacter/>) developed by K. Jolley and sited at the University of Oxford. Each strain has an ID number, and the strains newly isolated here have the continuous block of IDs from 930 to 1242. The development of this site has been funded by the Wellcome Trust and European Union.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/323/5913/527/DC1](http://www.sciencemag.org/cgi/content/full/323/5913/527/DC1)  
Methods  
Fig. S1  
Tables S1 and S2  
References

18 September 2008; accepted 11 December 2008  
10.1126/science.1166083

## Rapid Membrane Disruption by a Perforin-Like Protein Facilitates Parasite Exit from Host Cells

Björn F. C. Kafsack,<sup>1,2</sup> Janethe D. O. Pena,<sup>3</sup> Isabelle Coppens,<sup>2</sup> Sandeep Ravindran,<sup>4</sup> John C. Boothroyd,<sup>4</sup> Vern B. Carruthers<sup>1\*</sup>

Perforin-like proteins are expressed by many bacterial and protozoan pathogens, yet little is known about their function or mode of action. Here, we describe *Toxoplasma* perforin-like protein 1 (TgPLP1), a secreted perforin-like protein of the intracellular protozoan pathogen *Toxoplasma gondii* that displays structural features necessary for pore formation. After intracellular growth, TgPLP1-deficient parasites failed to exit normally, resulting in entrapment within host cells. We show that this defect is due to an inability to rapidly permeabilize the parasitophorous vacuole membrane and host plasma membrane during exit. TgPLP1 ablation had little effect on growth in culture but resulted in a reduction greater than five orders of magnitude of acute virulence in mice. Perforin-like proteins from other intracellular pathogens may play a similar role in microbial egress and virulence.

**P**erforin (PF) and members of the membrane attack complex (MAC) (complement proteins C6 to C9) are pore-forming proteins of the innate and adaptive immune re-

sponse that constitute the founding members of the MACPF domain family (1). Recent studies (2, 3) have suggested a shared mechanism of pore formation between the MACPF domain and



## Supporting Online Material for

### **The Peopling of the Pacific from a Bacterial Perspective**

Yoshan Moodley, Bodo Linz, Yoshio Yamaoka, Helen M. Windsor, Sebastien Breurec,  
Jeng-Yih Wu, Ayas Maady, Steffie Bernhöft, Jean-Michel Thiberge, Suparat  
Phuanukoonnon, Gangolf Jobb, Peter Siba, David Y. Graham, Barry J. Marshall,  
Mark Achtman\*

\*To whom correspondence should be addressed. E-mail: [m.achtman@ucc.ie](mailto:m.achtman@ucc.ie)

Published 23 January 2009, *Science* **323**, 527 (2009)

DOI: 10.1126/science.1166083

#### **This PDF file includes:**

Methods

Fig. S1

Tables S1 and S2

References

## Methods

**Bacterial isolates and DNA sequencing.** 212 *H. pylori* isolates were isolated after informed consent from stomach biopsies obtained during diagnostic gastroendoscopy, or using a string test, of native inhabitants in Taiwan, Papua New Guinea, New Caledonia, and Australia. 100 additional isolates were obtained from Europeans in Australia. These studies were approved by local ethics committees in Perth and Alice Springs, Australia, Goroka, Papua New Guinea; Taiwan; Houston, Texas; Noumea, New Caledonia; and Berlin, Germany. From each isolate, the forward and reverse strands of fragments of *atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI*, *yphC* were sequenced as described (*s1*). All sequences, primer combinations, PCR conditions and information on isolates are publicly available at <http://pubmlst.org/helicobacter>, where the new isolates described here are listed as id numbers 930-1241. The sequences from all seven housekeeping gene fragments were concatenated to form a 3,406 bp haplotype of which 1,695 bp are polymorphic. 17 haplotypes were identified that were identical to another haplotype; these “duplicate” haplotypes were ignored in subsequent calculations. The unique haplotypes are summarized in Table S1, which contains 196 novel haplotypes from native inhabitants, 99 from European Australians plus 222 previously described haplotypes (*s2*) from other inhabitants of Asia and the Pacific.

**Population assignments.** For the assignment to bacterial populations (Fig. S1), the 295 new haplotypes were combined with 769 haplotypes that had been analysed previously (*s2*). Individual haplotypes were assigned to bacterial populations using the “no admixture” model of the program STRUCTURE 2.0, as described (*s2*). The number of bacterial populations,  $K$ , was determined by comparing the posterior probability from multiple runs assuming  $4 \leq K \leq 9$ .  $K = 7$  yielded the highest probability and consistent strain assignment to populations whereas  $K > 7$  either produced hypothetical populations to which no isolates were assigned or split existing populations inconsistently. Subsequent runs on the hpEastAsia population alone separated isolates into the three subpopulations hspMaori, hspEAsia and hspAmerind, in agreement with prior analyses (*s2,s3*).

The number of ancestral populations was determined using the “linkage model” of STRUCTURE 2.0 as described previously (*s2*). When run with  $K=6$ , Structure identified the same five ancestral populations that have been described previously as ancestral Europe1 (AE1), ancestral Europe2 (AE2), ancestral EastAsia, ancestral Africa1 and ancestral Africa2. It also identified an additional, sixth ancestral population that was designated ancestral Sahul because it contributed ancestry solely to isolates of the population hpSahul. The individual nucleotides of hpSahul isolates from New Guinea were almost 100% derived from ancestral Sahul, whereas 95% to 55% of the nucleotides in hpSahul haplotypes from Australia were derived from ancestral Sahul and the remainder were derived from AE1 and AE2 (not shown).

**The Austronesian expansions.** All 80 hspMaori haplotypes from the Pacific (Table S1) were subjected to phylogenetic reconstruction using ClonalFrame (*s4*), except for one isolate from Japan. ClonalFrame estimates the clonal genealogy of a set of DNA sequences by stochastically simulating both mutation and homologous recombination

events using a neutral coalescent model within a Bayesian framework. Parameter space was explored by a Markov chain Monte Carlo (MCMC) simulation of 150,000 iterations, of which the first 50,000 were considered as burn-in. The genealogy was sampled every 100 iterations after the initial burn-in phase. This analysis was repeated 100 times, and an 80% consensus of all the sampled genealogies was computed using Treefinder (s5). Genetic diversity within Taiwanese and non-Taiwanese hspMaori was calculated in DnaSP4 (s6), as were the 95% confidence limits.

**Global migrations.** We used a population approach to determine pair-wise migration patterns between eight global *H. pylori* populations based on the multilocus haplotypes. Our raw data violates the assumption of no recombination within loci and we used the four-gamete criterion (s7) implemented in DnaSP4 (s6) to identify the blocks of DNA sequence that are most likely to be undergoing recombination between each pair of populations. These blocks were then discarded, and each of the remaining blocks of sequence was coded as a separate locus, resulting in pair-wise population data sets containing between 70 and 120 loci. This is justifiable because composite likelihoods tend to estimate the true posterior probability of a parameter when the number of loci is high (s8). We used Hey and Nielsen's model of isolation with migration (s9) to analyze these data. This models a single population split at time  $t$  into two populations which were subsequently isolated from each other. However, human populations undergoing genetic drift due to isolation probably continue to exchange a limited number of additional migrants over time. Hey and Nielsen's model accounts for this by introducing forward and reverse migration parameters ( $m1$  and  $m2$ ) that quantify the amount and direction of migration between both populations since the time of splitting. The estimation procedure is carried out using a Bayesian framework that requires user-defined prior input values, from which a heuristic MCMC exploration simultaneously generates posterior distributions for each of the model parameters, given the model and the sequence data. This enables the determination of not just the magnitude, but also the direction of the major migrations events that have occurred since the two populations split. This model does not assume that the two populations are at equilibrium for mutations, drift or migrations. These analyses were performed with the software IMA using 1,000,000 MCMC iterations, where genealogies were sampled every 100 iterations after a burn-in of 100,000 iterations. To stimulate mixing and convergence, each analysis implemented 100 geometrically-heated Metropolis-coupled chains, with 100 chain swapping attempts between iterations. Best mixing was found to occur when both heating ratios were set to 0.95. All estimates were taken after joint parameter maximization. The analysis was performed five times for each of the eight pairs of populations using differing input priors to obtain a range of parameter values around which the true parameter mean should be distributed. The number of migrants per generation ( $m$ ) was obtained by  $\theta m$ , which assumes clonal inheritance.

**Dating analyses.** Two important population splits – the split between Sahul and Asia; and within Sahul, the split between Australia and New Guinea – were estimated independently using IMA and ClonalFrame, which have different model assumptions: IMA: We obtained a population-based estimate for the TMRCA using the results of the IMA analyses described above. The values for the time splitting parameter ( $t$ ) from five repeated IMA analyses for each pair of populations (supplementary Table S2) were used

to construct a simple population tree (Fig. 2B). To obtain confidence limits on  $t$ , we used a Treefinder (*s5*) script (supplementary TreefinderScript1.tl) to generate 1,000 genealogies using a bootstrap procedure to sample the range of  $t$  values from each set of five IMA iterations.

**ClonalFrame:** To reduce computational time, a random sample of 70 globally representative isolates, drawn from all populations and subpopulations except hpEurope, was used for the reconstruction of the global phylogeny of *H. pylori* (Fig. 3A). hpEurope was excluded because this population resulted from admixture after secondary contact of two expanding ancestral populations (*s2*), which is not relevant for determining relationships between ancestral lineages. ClonalFrame was used to estimate the clonal genealogy 100 times using the same procedure as described for the Austronesian reconstruction above. This haplotype-based method not only estimates genealogy but also produces estimates of node height estimates ( $D$ ), which are directly proportional to the time to the most recent common ancestor (TMRCA) between two lineages. However, only node heights from topologically identical genealogies are directly comparable and we therefore discarded 13 genealogies of the 100 which did not conform to the consensus topology. The range of node height estimates in the remaining 87 genealogies was used for tree calibration.

**Tree calibration.** Both node height ( $D$ ) and the splitting parameter ( $t$ ) are related to the TMRCA via the equation  $D$  or  $t = 2\mu\text{TMRCA}$ , where  $\mu$  is the mutation rate. However, different lineages may not necessarily evolve at the same mutation rate, and applying the above formula to estimate the TMRCA without appropriate rate-smoothing may be misleading. Instead, we determined TMRCA by a calibration method. Six ranges of calibration points were used based on dates obtained from archaeological and molecular literature (Table S2). These were used as limits within which the range of node height/ $t$  values from multiple genealogies were calibrated using the global rate minimum deformation (GRMD) optimisation, implemented in Treefinder. The GRMD method is a rate-smoothing technique that minimises a cost function to maintain rates along different lineages that are as similar to each other as possible, given the imposed time boundaries. We felt that this method was appropriate given the near linear relationship of node height or  $t$  with calibration time (panel B, Figs 2 and 3). Ninety-five percent confidence limits for TMRCA estimates were also calculated in Treefinder.

Table S1. Population assignments of 517 unique haplotypes from <i>H. pylori</i> isolated in Asia or the Pacific											
Code	Country - Area	Ethnicity	Longitude	Latitude	hpSahul	hspEAsia	hspMaori	hpAsia2	hpEurope	hpNEAfrica	hpAfrica1
1	Japan - Hokkaido		141.40	43.10		9					
2	Japan - Honshu		139.75	35.68		10					
3	Japan - Okinawa		127.56	26.30		6	1				
4	Korea		127.46	35.88		14					
5	China - Hongkong		114.00	22.30		21			2		
6*	Taiwan	Kavalan	121.31	25.02		1					
7*	Taiwan	Atayal	121.31	25.02			4				
8*	Taiwan	Amis	121.31	25.02		9	7				
9*	Taiwan	Bunun	121.31	25.02		1	13				
10*	Taiwan	Rukai	121.31	25.02			3				
11*	Taiwan	Paiwan	121.31	25.02		3	15				
12*	Taiwan	Yami	121.31	25.02		2	1				
28	Taiwan <sup>1</sup>	Chinese	120.18	22.37		15					
13	Vietnam		106.66	10.75		20			2		
14	Singapore		103.81	1.34		9		1		1	
15	Malaysia		101.42	3.09		8		2			
16	Thailand		100.81	14.23				5	3		
17	Philippines		121.00	14.00			2	3	7		1
18	India - Ladakh		78.00	35.00		6		39	3		
19*	New Caledonia - Melanesians	Austronesian	68.38	48.80		2	6		5		
20*	New Caledonia - Polynesians from Wallis and Futuna <sup>2</sup>	Austronesian	176.12	-13.18			3		1		1
21	New Zealand	Austronesian	167.91	-41.56			15		1		2
22	Samoa/Tonga <sup>2</sup>	Austronesian	-171.95	-13.79			8		2		
23*	Australia - Alice Springs	Aboriginal	133.50	-23.40	3				4		2
24*	Australia - Jigalong	Aboriginal	120.78	-23.36	12				10		1
25*	Australia - Perth	Aboriginal	115.50	-31.60	11				45		2
26*	Australia - Torres Strait	Aboriginal	142.00	-10.00			2				
29	Australia - Alice Springs <sup>1</sup>	European	133.50	-23.40					9		1
30	Australia - Perth <sup>1</sup>	European	115.50	-31.60	1				74		
31	Australia - Busselton <sup>1</sup>	European	115.20	-33.39					13		1
27*	Papua New Guinea - Goroka	Papuan	145.23	-6.05	24				3		
	Total				51	136	80	50	184	1	11
*novel haplotypes from native inhabitants: 196; novel haplotypes from European Australians: 99; haplotypes described by Linz et al., 2007 (s2): 222											
<sup>1</sup> - Samples not shown on Fig. 1 because they represent control groups											
<sup>2</sup> - Longitude and Latitude values correspond to their origins, but Polynesians from Wallis/Futuna were resident in New Caledonia and Samoans/Tongans were resident in New Zealand.											



Table S2. Population parameters simulated using an isolation with migration model and two estimates of time to a most recent common ancestor (TMRCA)												
Population 1	Population 2	$\theta_1$	$\theta_2$	Migration rate/generation ( $m=\theta m$ )		$t$ (IMa)		Node height (CF)		TMRCA (CI 95)*		Reference
				m1	m2	Lower	Upper	Lower	Upper	IMa (kyr)	CF (kyr)	
Africa1	All non-African	0.1604	0.1665	2.0334	1.1889	0.1046	0.1638	0.1936	0.2443	56 - 115		<i>s2, s10, s11, s12</i>
Sahul	All Asian	0.5910	0.0606	1.5056	0.0001	0.0745	0.0866	0.1509	0.1917	30.8 - 37.0	31.8 - 33.0	
Australia	Papua	0.3865	0.0341	0.0000	0.2501	0.0659	0.0683	0.1143	0.1524	22.6 - 31.8	23.0 - 25.0	
Asia2	EastAsia	0.1538	0.2626	6.0251	2.5727	0.0700	0.0733	0.1363	0.1750	30 - 50		<i>s13</i>
East Asia	Amerind	0.3711	0.3576	1.6017	0.9060	0.0544	0.0685	0.1082	0.1414	19 - 23		<i>s14</i>
EastAsia	Taiwan-Maori	0.2985	0.1624	5.4391	1.8946	0.0400	0.0467	0.0958	0.1274	9.4 - 17		<i>s15</i>
Taiwan	Maori	0.2470	0.2657	1.5148	0.0000	0.0338	0.0479	0.0756	0.1075	4.9 - 5.0		<i>s16, s17</i>
Amerind North	Amerind South	0.3454	0.4396	6.6382	0.0001	0.0317	0.0451	0.0815	0.1100	13.98 - 14.22		<i>s18</i>

\* 95% confidence interval (CI) of the TMRCA for the two Sahul population splits are highlighted in yellow. Other values in this column are the referenced calibration times intervals used to infer the two TMRCA values.

The population parameter theta ( $\theta$ ) =  $2N_e\mu$  for clonal inheritance, where  $N_e$  is the effective population size and  $\mu$  is the mutation rate.



**Figure S1.** Distruct (*s19*) plot of the Bayesian assignment of 1068 *H. pylori* haplotypes to populations as determined by Structure V2.0 (no admixture model) . Each isolate is represented by a thin line that is color coded according to the population assignment. The isolates are grouped by population.

References For Supplementary Methods and Tables

- s1. M. Achtman *et al.*, *Mol. Microbiol.* **32**, 459 (1999).
- s2. B. Linz *et al.*, *Nature* **445**, 915 (2007).
- s3. D. Falush *et al.*, *Science* **299**, 1582 (2003).
- s4. X. Didelot, D. Falush, *Genetics* **175**, 1251 (2007).
- s5. G. Jobb, A. von Haeseler, K. Strimmer, *BMC. Evol. Biol.* **4**, 18 (2004).
- s6. J. Rozas, J. C. Sanchez-DelBarrio, X. Messeguer, R. Rozas, *Bioinformatics.* **19**, 2496 (2003).
- s7. R. R. Hudson, N. L. Kaplan, *Genetics* **111**, 147 (1985).
- s8. C. Wiuf, *Journal of Mathematical Biology* **53**, 821 (2006).
- s9. J. Hey, R. Nielsen, *Proc. Natl. Acad. Sci. U. S. A* **104**, 2785 (2007).
- s10. H. Liu, F. Prugnolle, A. Manica, F. Balloux, *Am. J. Hum. Genet.* **79**, 230 (2006).
- s11. M. Nei, A. K. Roychoudhury, *Am. J. Hum. Genet.* **26**, 421 (1974).
- s12. S. A. Tishkoff *et al.*, *Science* **271**, 1380 (1996).
- s13. K. O. Pope, J. E. Terrell, *J. Biogeography* **35**, 1 (2008).
- s14. N. J. Fagundes *et al.*, *Am. J. Hum. Genet.* **82**, 583 (2008).
- s15. J. A. Trejaut *et al.*, *PLoS. Biol.* **3**, e247 (2005).
- s16. M. Spriggs, *Antiquity* **63**, 587 (1989).
- s17. M. J. T. Spriggs, in *Prehistoric Mongoloid Dispersals*, T. Akazawa, E. Szathmary, Eds. (Oxford University Press, Oxford, 1996), pp. 322-346.
- s18. T. D. Dillehay *et al.*, *Science* **320**, 784 (2008).
- s19. N. A. Rosenberg, *Molecular Ecology Notes* **4**, 137 (2004).

**Article n°2: Evolutionary history of *Helicobacter pylori* sequences  
reflects past human migrations in Southeast Asia**

Breurec S, Guillard B, Hem S, Brisse S, Dieye FB, Huerre M,  
Oung C, Sreng Tan T, Raymond J, Thiberge JM, Vong S, Monchy  
D, Linz B.

PLoS ONE. 2011. 6: e22058.

## **Introduction**

La distribution fragmentée des peuples parlant l'une des cinq familles de langues majeures en Asie du sud-est continentale est le reflet d'importantes et complexes migrations. Les langues austro-asiatiques (essentiellement parlées en Inde, au Vietnam et au Cambodge), austronésiennes (cham du Vietnam) et hmong-mien (parlées essentiellement dans le sud de la Chine, le Laos et le nord du Vietnam) sont considérées comme les familles de langues les plus anciennes, tandis que les langues tai-kadai (parlées essentiellement en Thaïlande) et tibéto-birmanes (parlées en Chine, Inde, Tibet, Népal, Bhoutan, Birmanie et dans quelques régions du Pakistan, du Bangladesh, de la Thaïlande, du Laos et même du Vietnam) sont issues de migrations plus récentes. De plus, l'archéologie suggère un lien étroit entre l'Inde et l'Asie du sud-est continentale à partir du 3<sup>ème</sup> siècle jusqu'à la fin de la période Angkor (15<sup>ème</sup> siècle). Cependant, cette migration n'est pas confirmée par l'étude de l'ADNmt humain et de la linguistique. Qu'apporte l'étude de la diversité génétique de souches de *H. pylori* ?

## **Matériels et méthodes**

Un total de 66 patients d'ethnie khmer (36 hommes, âge médian 46.0 ans, intervalle 18-76 ans) présentant une infection prouvée à *H. pylori* (culture positive des biopsies gastriques prélevées lors d'une endoscopie gastrique) ont été inclus sur deux périodes, en 2004 et en 2007. Ces patients ont été prélevés au Centre Hospitalier Universitaire (CHU) Calmette de Phnom Penh (Cambodge) (n=37), un des plus grands hôpitaux du pays, et dans un centre de gastroentérologie privé (n=29). L'ADN de *H. pylori* a été extrait à partir de subcultures réalisées par repiquage d'une colonie de la primo-culture. L'analyse du polymorphisme des 7 gènes de ménage a été réalisée selon la méthodologie décrite précédemment. Les séquences générées ont été additionnées de séquences obtenues à partir de souches isolées de patients caucasiens vivant en France, mais également de séquences précédemment publiées de souches originaires d'Asie, d'Océanie et d'Europe.

La structure génétique des populations a été investiguée par la mesure de l'indice de diversification génétique (Fst) et l'analyse moléculaire de la variance (AMOVA) avec le programme ARLEQUIN (version 2.00) en utilisant un modèle « Kimura 2-parameter ». La significativité des Fst par paires de population a été calculée par permutation (10000) des individus. Les relations phylogénétiques entre les populations ont été investiguées par la méthode du Neighbour-Joining à partir de la matrice des Fst en utilisant le logiciel MEGA 4.

## Résultats et Discussion

L'analyse du polymorphisme génétique des 66 souches isolées au Cambodge a montré la coexistence de 2 populations génétiques présentes à peu près en proportion égale, hpEurope (n=34, 51.5 %) et hpEastAsia, sous-population hspEAsia (n=32, 48.5 %) (Logiciel STRUCTURE, modèle « no linkage »). L'Indochine était une ancienne colonie française (1863-1953), regroupant l'actuel Vietnam, l'actuel Laos et l'actuel Cambodge. La forte fréquence des souches hpEurope au Cambodge ainsi qu'en Thaïlande (36 %), pays qui n'a jamais été sous domination européenne, et leur rareté au Vietnam (9 %), suggèrent une introduction de souches d'origine indo-européenne beaucoup plus ancienne dans le temps que la colonisation française de l'Indochine.

La différenciation génétique des populations hpEurope a été quantifiée par l'indice  $F_{st}$ , en utilisant les séquences concaténées de populations issues d'Europe (173 souches), du Moyen-Orient (16 souches), d'Iran (125 souches), du Cambodge (34 souches), d'Inde (23 souches), de Malaisie (8 souches d'Indiens et 4 de Malais), de Thaïlande (6 souches de Thaïs) et des Philippines (6 souches). Une différenciation significative a été trouvée entre la population issue du Cambodge et celles issues d'Europe, excluant une introduction récente de souches hpEurope par les Français. L'arbre phylogénétique regroupait les populations hpEurope des Indiens, des Malais, des Thaïs et des Khmers dans un cluster distinct de celui regroupant les populations d'Europe et du Moyen-Orient, indiquant que ces souches sont issues de la même migration (Figure 1, p4 de l'article). Les souches des Indiens d'Inde et de Malaisie se groupaient ensemble, en accord avec la migration récente et massive d'Indiens en Malaisie pendant la colonisation britannique. En revanche, les isolats des Malais étaient génétiquement plus proches de ceux du Cambodge et de Thaïlande que de ceux des Indiens. Ces résultats sont en concordance parfaite avec l'analyse AMOVA. Ils indiquent l'introduction ancienne des souches hpEurope sur le sous-continent indien par la migration indo-aryenne (4000 à 10000 ans avant notre époque). Elle a été suivie par des migrations ultérieures de leurs descendants en Asie du sud-est, transportant des souches hpEurope dans leur estomac. Les souches des Malais ont pour origine une ancienne migration venue d'Inde et une migration plus récente. Ces données recourent les données archéologiques et confirment la résolution plus élevée de la diversité génétique de *H. pylori* pour retracer les anciennes migrations humaines par comparaison aux marqueurs génétiques humains traditionnels.

La langue vietnamienne (Vietnam) et la langue khmère (Cambodge) font parties de la famille des langues austro-asiatiques. La sous-population hspEAsia étant majoritaire au Vietnam, il n'est pas étonnant que cette sous-population soit retrouvée au Cambodge en forte

proportion (48 %). Récemment, de nombreuses études se sont intéressées à l'origine des peuples parlant des langues austro-asiatiques. Sur la base des données linguistiques, archéologiques et génétiques, plusieurs hypothèses ont été émises sur leur origine ancestrale. La différenciation génétique des populations hspEAsia a été quantifiée à l'aide de l'indice  $F_{st}$  en utilisant les séquences concaténées de souches du Cambodge (32 souches), Vietnam (20 souches), Thaïlande (18 souches), Malaisie (25 souches), Singapour (9 souches), Japon (24 souches), Corée du sud (10 souches), Taiwan (15 souches) et de différentes localités en Chine (93 souches). Un arbre phylogénétique a été généré par la méthode du Neighbour-Joining. Sa topologie (Figure 2, p5 de l'article) est en faveur d'une migration à partir du sud de la Chine des ancêtres des peuples parlant des langues austro-asiatiques au Cambodge et au Vietnam, hébergeant *H. pylori* dans leur estomac. Le peuplement a été daté à 4000 ans avant notre ère selon les données archéologiques. Toutefois, l'arbre généré manque de robustesse pour affirmer l'origine précise. Il montre également la migration récente des Chinois en Thaïlande, Malaisie et Singapour à partir des provinces du Guangdong et du Fujian (sud de la Chine), ce qui est en accord avec les données historiques.

La dernière population retrouvée en Asie du sud-est est la population hpAsia2, présente en proportion importante en Malaisie et en Thaïlande. La différenciation génétique des populations hspEAsia à l'aide de l'indice  $F_{st}$  a été quantifiée en utilisant les séquences concaténées de souches de Thaïlande (9 souches), de Malaisie (32 souches), du Bangladesh (3 souches), du nord de l'Inde (Ladakh) (39 souches) et des Philippines (3 souches). En se basant sur la topologie de l'arbre généré (Figure 3, p7 de l'article) et sur l'absence de la population hpAsia2 au Cambodge et au Vietnam, nous avons proposé l'hypothèse que les souches hpAsia2 avaient été introduites par deux migrations en Asie du sud-est : i) la migration des ancêtres des Thaïs, du Sud de la Chine vers l'actuelle Thaïlande au début du second millénaire après JC selon les données archéologiques, et ii) la migration récente d'Indiens en Malaisie.

La distribution des populations génétiques bactériennes semble fortement influencer l'incidence des CG puisque les pays avec une prédominance de souches hspEAsia présente une forte incidence de CG tandis que l'incidence est faible dans les pays présentant une forte proportion de souches hpEurope ou hpAsia2. L'expansion observée des souches hpEurope (souches dites « généralistes ») au détriment des souches hspEAsia (souches dites « spécialistes ») au sein de la population cambodgienne par un mécanisme de compétition et de subversion pourrait avoir un impact significatif sur l'incidence du CG au Cambodge dans le futur.

### **Conclusion**

Nous avons décrit trois anciennes migrations humaines (Figure 4, p7 de l'article):

- une migration d'Indiens vers l'Asie du sud-est (population génétique hpEurope), principalement entre 3<sup>ème</sup> et le 15<sup>ème</sup> siècle après JC,
- une migration de peuples parlant les ancêtres des langues austro-asiatiques au Vietnam et au Cambodge (sous-population génétique hspEAsia), avec une arrivée dans cette région il y a environ 4000 ans,
- une migration des ancêtres des Thaïs du sud de la Chine vers l'actuelle Thaïlande au début du second millénaire après JC (population génétique hpAsia2).

De plus, la diversité génétique de *H. pylori* en Asie du sud-est reflète depuis deux siècles la migration de Chinois en Thaïlande, Malaisie et Singapour (sous-population génétique hspEAsia) et celle d'Indiens en Malaisie (populations génétique hspAsia2 et hpEurope).



# Evolutionary History of *Helicobacter pylori* Sequences Reflect Past Human Migrations in Southeast Asia

Sebastien Breurec<sup>1,2\*</sup>, Bertrand Guillard<sup>2</sup>, Sopheak Hem<sup>2</sup>, Sylvain Brisse<sup>3</sup>, Fatou Bintou Dieye<sup>1</sup>, Michel Huerre<sup>4</sup>, Chakravuth Oung<sup>5</sup>, Josette Raymond<sup>6,7</sup>, Tek Sreng Tan<sup>8</sup>, Jean-Michel Thiberge<sup>3</sup>, Sirenda Vong<sup>9</sup>, Didier Monchy<sup>2,10</sup>, Bodo Linz<sup>11</sup>

**1** Unité de Biologie Médicale et Environnementale, Institut Pasteur, Dakar, Senegal, **2** Unité de Biologie Médicale, Institut Pasteur, Phnom Penh, Cambodia, **3** Plate-Forme Génotypage des Pathogènes et Santé Publique, Institut Pasteur, Paris, France, **4** Unité de Recherche et d'Expertises en Histotechnologie et Pathologie, Institut Pasteur, Paris, France, **5** Gastroenterology and Liver Unit, Calmette Hospital, Phnom Penh, Cambodia, **6** Unité Postulante Pathogénèse de *Helicobacter*, Institut Pasteur, Paris, France, **7** Université Paris Descartes, Faculté de Médecine, Paris, France, **8** Private Medical Center, Phnom Penh, Cambodia, **9** Unité d'Epidémiologie et de Santé Publique, Institut Pasteur, Phnom Penh, Cambodia, **10** Laboratoire de Biologie Médicale, Institut Pasteur, Bangui, République Centrafricaine, **11** Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania, United States of America

## Abstract

The human population history in Southeast Asia was shaped by numerous migrations and population expansions. Their reconstruction based on archaeological, linguistic or human genetic data is often hampered by the limited number of informative polymorphisms in classical human genetic markers, such as the hypervariable regions of the mitochondrial DNA. Here, we analyse housekeeping gene sequences of the human stomach bacterium *Helicobacter pylori* from various countries in Southeast Asia and we provide evidence that *H. pylori* accompanied at least three ancient human migrations into this area: i) a migration from India introducing hpEurope bacteria into Thailand, Cambodia and Malaysia; ii) a migration of the ancestors of Austro-Asiatic speaking people into Vietnam and Cambodia carrying hspEAsia bacteria; and iii) a migration of the ancestors of the Thai people from Southern China into Thailand carrying *H. pylori* of population hpAsia2. Moreover, the *H. pylori* sequences reflect iv) the migrations of Chinese to Thailand and Malaysia within the last 200 years spreading hspEAsia strains, and v) migrations of Indians to Malaysia within the last 200 years distributing both hpAsia2 and hpEurope bacteria. The distribution of the bacterial populations seems to strongly influence the incidence of gastric cancer as countries with predominantly hspEAsia isolates exhibit a high incidence of gastric cancer while the incidence is low in countries with a high proportion of hpAsia2 or hpEurope strains. In the future, the host range expansion of hpEurope strains among Asian populations, combined with human motility, may have a significant impact on gastric cancer incidence in Asia.

**Citation:** Breurec S, Guillard B, Hem S, Brisse S, Dieye FB, et al. (2011) Evolutionary History of *Helicobacter pylori* Sequences Reflect Past Human Migrations in Southeast Asia. PLoS ONE 6(7): e22058. doi:10.1371/journal.pone.0022058

**Editor:** J. Ross Fitzgerald, University of Edinburgh, United Kingdom

**Received:** April 1, 2011; **Accepted:** June 14, 2011; **Published:** July 19, 2011

**Copyright:** © 2011 Breurec et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This study was supported by grants from Institut Pasteur (ACIP *Helicobacter pylori*), the ERA-NET PathoGenoMics (project HELDIVNET, decision n° ANR-06-PATHO-007-01) and Institut de Veille Sanitaire (InVS, Saint-Maurice, France). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: sbreurec@pasteur.sn

## Introduction

The fragmented distribution of speakers of the five major language families in Southeast Asia is the result of extensive human migrations. Hmong Mien, Austro-Asiatic and Austronesian are considered the older language families in the region [1], whereas the presence of the Sino-Tibetan and Tai-Kadai language families can be attributed to relatively recent population expansions. Most fragmented is the distribution of Hmong-Mien speakers living in numerous small enclaves surrounded by Sino-Tibetan and Tai-Kadai speakers in Southern China, Laos and Northern Vietnam because of an extreme expansion of the Chinese subfamily of Sino-Tibetan (mostly during the Zhou dynasty 1100 to 221 BC) which distributed Chinese languages continuously over a large region from North to South China, pushing speakers of other languages further south and west. The Austro-Asiatic language family (with the examples of Vietnamese from Vietnam and Khmer from Cambodia) was previously

distributed from Vietnam in the east and South China in the north to the Malay Peninsula in the south and North India to the west [2] before massive expansions of Indo-European speakers in India and Tibeto-Burman speakers (a subgroup of Sino-Tibetan different from Chinese) from South China into Myanmar restricted Austro-Asiatic languages to numerous enclaves in this area. A subsequent expansion of Tai-Kadai speakers during the early second millennium AD from their homeland in South China into Thailand and Laos replaced Austro-Asiatic speakers in large parts of Southeast Asia that previously belonged to the Khmer empire [3,4,5]. Subsequently, Tai-Kadai is found from South China over Thailand to the Malay Peninsula and Myanmar.

In historic times, parts of Southeast Asia have repeatedly been ruled by colonial forces, but there has never been overall occupation [1,4]. The Han Chinese invaded North Vietnam (Tonkin) in the 1<sup>st</sup> century BC and stayed for nearly a millennium, after which Vietnamese dynasties from North Vietnam conquered central Vietnam (Annam) and South Vietnam (Cochin China).

The French occupied the same area (Tonkin, Annam, Cochin China) during a far shorter period (1863–1953), and added present day Cambodia and Laos to their colonial French Indochina. Both of these colonial episodes excluded Siam (Thailand), the only country in Southeast Asia never colonized by a European power.

Archaeology suggests an ancient close connection between India and the Thailand/Cambodia region through settlement [6,7,8,9], accompanied by an increasing exposure to Indian culture from about 300 BC. Early states-like societies from Southeast Asia called by the Sanskrit term “mandala” had in common the adoption of Indian forms of religion (Hinduism), the Sanskrit language and aspects of government (Funan mandala from 100 to 550 AD, Chenla mandala from 550 to 802 AD and Angkorian mandala from 802 to 1431 AD) [4]. However, the Indian influence in Southeast Asia was not supported by human mitochondrial DNA (mtDNA) data [10,11,12].

In previous studies, we have used housekeeping gene sequences of a bacterial parasite which infects the stomach of most humans, *Helicobacter pylori*, to elucidate the patterns of human prehistory. *H. pylori* accompanied modern humans during their migrations out of Africa ca. 60,000 years ago [13], and subsequent geographic separation plus founder effects have resulted in genetic populations of bacterial strains that are specific for large continental areas. In all, 7 bacterial genetic populations have been described [13,14,15,16,17,18]: hpEurope (isolated from Europe, the Middle East, India and Iran), hpNEAfrica (isolated in Northeast Africa), hpAfrica1 (isolated from countries in Western Africa and South Africa), hpAfrica2 (so far only isolated from South Africa), hpAsia2 (isolated from Northern India and among isolates from Bangladesh, Thailand and Malaysia), hpSahul (from Australian Aborigines and Papua New Guineans) and hpEastAsia with the subpopulations hspEAsia (from East Asians), hspMaori (from Taiwanese Aborigines, Melanesians and Polynesians) and hspAmerind (Native Americans). All these modern populations derived from six ancestral populations that were designated ancestral Europe1 (AE1), ancestral Europe2 (AE2), ancestral EastAsia, ancestral Africa1, ancestral Africa2 [14] and ancestral Sahul [16].

The specific geographic distribution and ethnic association of the *H. pylori* populations reflects numerous ancient and historic human migrations which established *H. pylori* sequences as a useful genetic marker to unravel debated topics in human population history. For example, the genetic variation in *H. pylori* has showed more discriminatory power in determining the ancient sources of human migrations in the Ladakh region of Northern India [19] and in the Pacific (Austronesian expansion) [16] than traditional human genetic markers such as the hypervariable region (HSV1) of mtDNA. Therefore, we analysed *H. pylori* sequences from Cambodia which borders Thailand to its west and northwest, Vietnam to its east and southeast and Laos to its north, to gain additional insights into the human population history in continental Southeast Asia.

## Materials and Methods

### Strains and ethics statement

Gastroduodenal endoscopy was performed at the Gastroenterology Department of the Calmette Hospital and at a private medical center in Phnom Penh, Cambodia with the permission of the Cambodian National Ethics Committee for Health Research (ethics certificate 017/03NECHR). Informed written consent was received from all participants.

A total of 66 *H. pylori* strains derived from 66 patients (36 (55%) males, median age 46.0 years; range 18–76 years) who suffered from upper abdominal pain were isolated in 2004 and in 2007.

Demographic data, the medical history and the presenting symptoms were prospectively collected by the physician. All the patients were of Khmer origin, and none had received proton pump inhibitors or antibiotics during the 4 weeks before endoscopy. Three biopsy samples were taken from the antrum and three from the fundus during upper gastrointestinal tract endoscopy. One biopsy from each site was cultured for *H. pylori* isolation, and the others were fixed and processed for histological analysis.

The Cambodian strains were supplemented by unpublished sequences of strains from French Caucasians (n=8), as well as sequences obtained from <http://pubmlst.org/helicobacter/> that were previously published by Falush *et al.* 2003 [14], Wirth *et al.* 2004 [15], Momynaliev *et al.* 2005 [20], Linz *et al.* 2007 [13], Devi *et al.* 2007 [17], Tay *et al.* 2009 [18], Liao *et al.* 2009 [21] and Moodley *et al.* 2009 [10]. Novel sequences experimentally obtained in this study were deposited in GenBank database under the following accession numbers HM362684 to HM362767.

### *H. pylori* isolates and genomic DNA

*H. pylori* culture was performed using Columbia agar plates with 10% (v/v) defibrinated horse blood and *H. pylori* selective antibiotic supplement (Oxoid, Basingstoke, UK) containing vancomycin (10 mg/L), cefsulodin (5 mg/L), trimethoprim (5 mg/L) and amphotericin B (5 mg/L). The plates were incubated for up to 10 days at 37°C under microaerophilic conditions (GENbag, Biomerieux). *H. pylori* was identified by colony and microscopic morphology and by positive urease, catalase, and oxidase tests. From primary growths, a single *H. pylori* colony from antrum or fundus was picked and subcultured in order to ensure that each strain consists of only a single genotype. Genomic DNA was extracted using a QIAmp™ kit (Qiagen, Courtaboeuf, France).

### Data analysis

PCR amplification and sequencing of *atpA*, *efp*, *mutY*, *pba*, *trpC*, *ureI*, and *yphC* were performed as previously described [13]. Strain population assignment was performed as described by Falush *et al.* [14] using the “no admixture model” of STRUCTURE [22]. The linkage model in STRUCTURE was used to estimate the proportion of nucleotides being derived from each ancestral population as described [13,14]. The estimated amount of ancestry from each population was plotted as a thin line for each isolate using DISTRUCT [23].

Pair-wise  $F_{ST}$  values as well as the analyses of molecular variance (AMOVA) were calculated in ARLEQUIN [24] as described before [25] using the Kimura 2-parameter model that was previously applied to *H. pylori* sequences [13,14,19,25]. The significance of the pair-wise  $F_{ST}$  values was estimated by running 10,000 permutations assuming no difference between the populations. Neighbor-joining trees from the pair-wise  $F_{ST}$  values were generated in MEGA v4 [26].

## Results and Discussion

### *H. pylori* from Khmer in Cambodia

*H. pylori* isolates were cultured from gastric biopsies obtained from 66 Khmer volunteers during gastroduodenal endoscopy at the Calmette Hospital (n=37) and at a private medical center (n=29) in Phnom Penh, Cambodia, in 2004 and 2007. The concatenated sequences of 7 housekeeping gene fragments (3406 base pairs, of which 838 were polymorphic) yielded 66 unique haplotypes that were compared to haplotypes from other countries in Asia and ~700 haplotypes from other sources including Europe

and Sahul. Bayesian clustering algorithms implemented in STRUCTURE (no admixture model) [18] assigned 34 (52%) new bacterial haplotypes to the *H. pylori* population hpEurope and 32 (48%) new haplotypes to hpEastAsia, subpopulation hspEAsia (Table 1), with no significant difference between 2004 and 2007 (data not shown). The large proportion of hpEurope strains is surprising because *H. pylori* from this population are known to be more characteristic of the Middle East, Europe and countries

colonized by Europeans [13,14], India [17] and central Asia including Iran [25]. Given the large geographical origin of the patients attending the study health facility, we believe that the sample may be representative of the country.

Strains of the hpEurope population were shown to be hybrids of two ancestral populations, AE1 from central Asia and AE2 from northeast Africa [13] while modern hpEastAsia strains are almost pure descendants of ancestral EastAsia. By using the linkage model

**Table 1.** Sources and population assignment of the analysed *H. pylori* strains.

Country/ethnic origin	Number of isolates assigned to					Reference
	hpEurope	hpAsia2	hspEAsia	hspMaori	hspAmerind	
Iran	125					27
Kazakhstan	6					13
Turkey	18					13
Estonia	10					14
Finland	10					14
Germany	23					14
Netherlands	5					13
UK	16					14
Russia	21					20
France	8					this study
Italy	9					13
Spain Spaniard	35					14
Spain Basque	44					13
Lebanon	5					13
Palestine	11					13
Thailand Thai	6	9				13
Thailand Chinese	1		18			13
India (Ladakh) Muslim	2	14				19
India (Ladakh) Buddhist		25	6			19
India (Andhra Pradesh)	23					17
Malaysia Indian	8	23	4			18
Malaysia Malay	4	9	2			18
Malaysia Chinese	1		25			18
Singapore Chinese			9			13
Taiwan Chinese			15			13
Cambodia	34		32			this study
Philippines	7	3				13
Bangladesh		3				13
Vietnam	2		20			13
Japan			24			13
China	5		93			21, 13
Korea			10			14
Taiwan Aboriginal				43		16
New Caledonia*				6		16
Wallis Polynesians				4		16
Samoa Polynesians				8		14
New Zealand Maori				15		14
North America					11	13
South America					7	13, 42

\*Melanesians.

doi:10.1371/journal.pone.0022058.t001

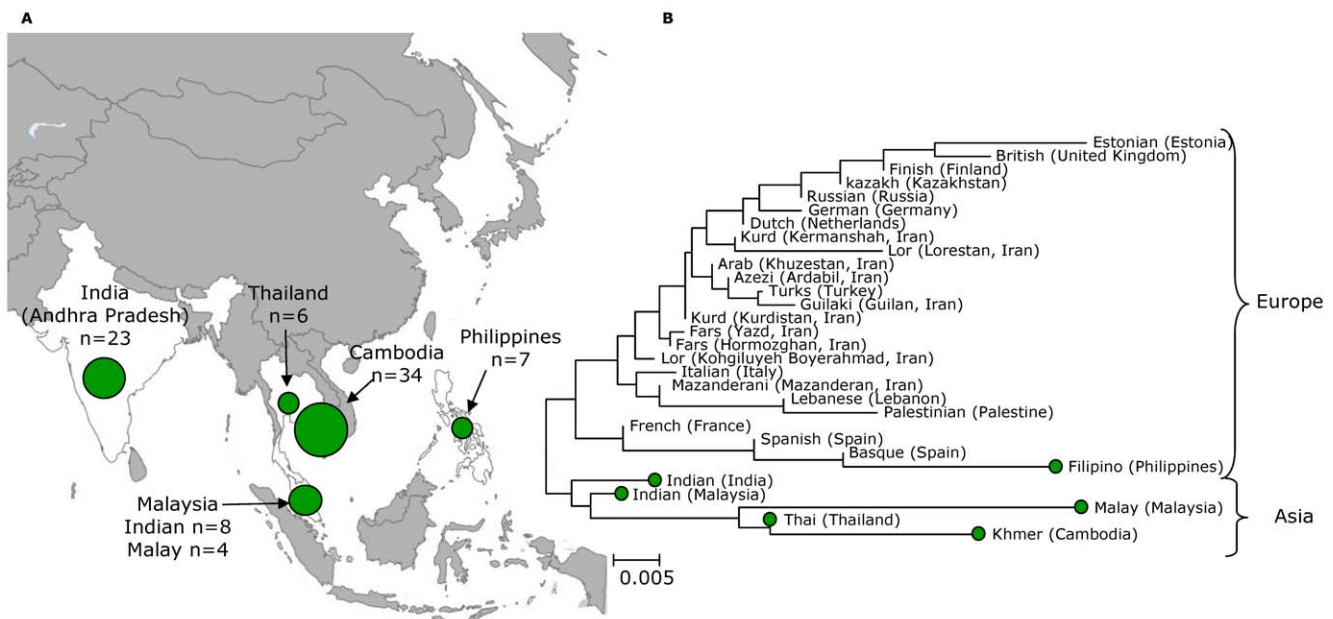
of STRUCTURE [22] to estimate the proportion of nucleotides derived from each of the previously identified ancestral populations [13,14,16], we identified isolates from Khmer that had acquired significant proportions (>20%) of foreign nucleotides from other ancestral populations. Four hspEAsia strains (12.5%) harboured a high proportion of AE2 while eight hpEurope strains (23%) contained a significant EastAsian ancestral component (Figure S1), indicating long time co-evolution of hpEurope and hspEAsia bacteria in the area. Introgressed nucleotides from other ancestral populations might change the level of differentiation between the *H. pylori* populations and thus distort the pair-wise  $F_{ST}$  values. We stripped the dataset from isolates with a proportion of imported nucleotides from other ancestral populations >20% which did not change what populations were significantly differentiated from each other. In addition, the topology of the neighbor-joining trees (Figure 1; Figure 2) was unaffected (not shown), and there were only minor differences in the length of a few branches. Therefore, all the strains were included in the subsequent analyses.

### An ancient migration from India introduced hpEurope strains to Southeast Asia

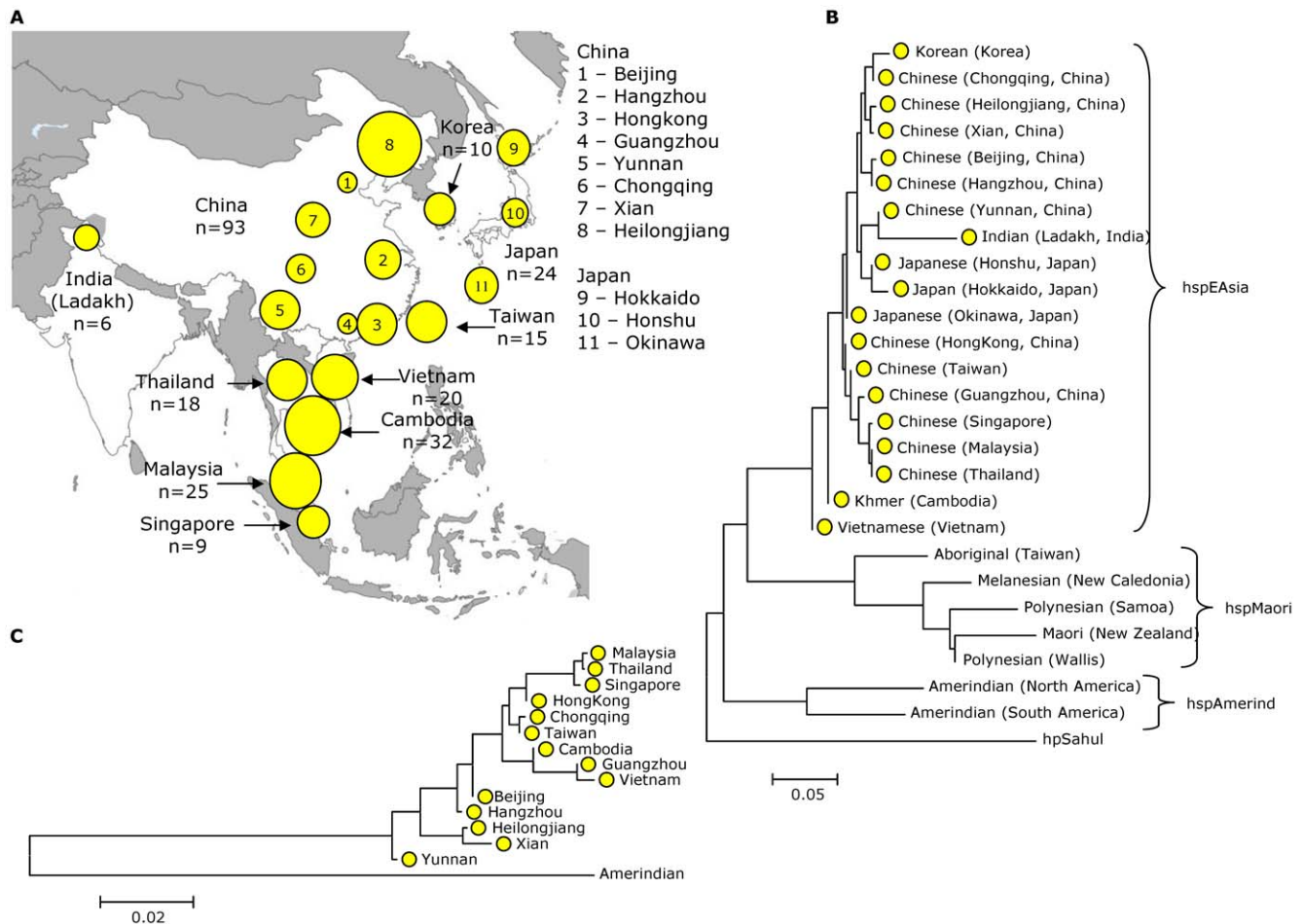
The high prevalence of hpEurope strains (52%) in Khmer population raises the question of the origin of these isolates. If modern introduction by French during the Indochina history were the source, hpEurope strains would be expected to be widespread in Vietnam and Cambodia and scarce in Thailand, because Vietnam and Cambodia were part of the French colonial empire for a short period (1887–1954) but the kingdom of Siam (Thailand) was never under European rule. However, the frequency of hpEurope strains among ethnic Thai was higher (37%) than among Vietnamese (9%) (Table 1) [13]. In order to investigate signatures of genetic differentiation, we calculated pairwise  $F_{ST}$  values in ARLEQUIN [19] using concatenated sequences of hpEurope strains from various countries in Europe (173 strains), from the Middle East (16 strains), from Iran (125 strains), from

Cambodia (34 strains), from India (23 strains), from Malaysia (8 strains from Indians and 4 from Malays), from Thailand (6 strains from Thai) and from the Philippines (7 strains) (Figure 1A). *H. pylori* haplotypes from the Philippines, that experienced over three centuries of Spanish colonial history (1565–1898), were significantly differentiated from the Khmer and Thai populations, but not from Spanish, and thus likely resulted from a recent introduction by Europeans. In contrast, the Khmer population was not significantly differentiated from the Thai population but was significantly differentiated from European populations including French population ( $p < 0.05$ ) (Table 2), rejecting the hypothesis of a recent introduction of hpEurope strains by the French during the Indochina history. These observations suggest that hpEurope bacteria in Southeast Asia might be a marker for an old human migration that predated the European colonial history.

A neighbor-joining tree based on these pairwise  $F_{ST}$  values (Figure 1B) joined the hpEurope haplotypes from the Indian, Thai, Khmer and Malay populations into a distinct cluster that was separated from haplotypes from Europe and the Middle East which indicated a common origin of these Asian hpEurope strains. Tay *et al.* [18] suggested a recent introduction of the hpEurope haplotypes by Indians into Malaysia within the last 200 years. Malaysian Indians are largely descended from people who migrated from southern India during the British colonization of Malaysia [27], and strains from modern Indians and Indians from Malaysia indeed clustered together, consistent with their origin (Figure 1B). However, strains from Malays were more closely related to those from Khmers and Thais than they were to Indian or Malaysian Indian strains, suggesting a common origin of these strains and arguing against an exclusively recent acquisition of Malaysian hpEurope strains from Indian immigrants, contrary to Tay *et al.*'s interpretation [27]. Moreover, people of Indian origin are not common in Cambodia or Thailand, a situation that contrasts with Malaysia where Indian ethnicity exceeds 7% of the general population. Strains from modern Indians and Malaysian Indians were located near the base of the branch leading to the



**Figure 1. The distribution and phylogenetic differentiation of European (hpEurope) *H. pylori* isolates in Southeast Asia.** (A) Map of sampling locations of hpEurope haplotypes in Southeast Asia. (B) The Neighbor-joining tree generated from pair-wise  $F_{ST}$  values indicates a common, non-European origin of the Southeast Asian hpEurope strains. doi:10.1371/journal.pone.0022058.g001



**Figure 2. hpEastAsia *H. pylori* strains in Southeast Asia.** (A) Map of sampling locations of hspEAsia haplotypes in Southeast Asia. The size of the chart indicates the number of hspEAsia haplotypes at each sampling location. (B) Neighbor-joining tree from pair-wise  $F_{ST}$  values of hpEastAsia haplotypes rooted with haplotypes of the population hpSahul. (C) Neighbor-joining tree of pair-wise  $F_{ST}$  values of the subpopulation hspEAsia. doi:10.1371/journal.pone.0022058.g002

Thai, Khmer and Malay haplotypes in the neighbor-joining tree (Figure 1B), suggesting the Indian subcontinent as the source of hpEurope bacteria in Thais, Khmers and Malays. Group assignments by AMOVA analyses for hpEurope strains provided strong statistical support of the tree topology (Table S1). Taken together, all these observations indicate an old introduction of hpEurope strains into the Indian subcontinent by Indo-Aryan migration (4000–10000 BP) as previously described [17,28]. This was followed by subsequent eastward migrations of their descendants into Southeast Asia, carrying hpEurope strains in their stomach, probably within the last 3000 years. The hpEurope strains in Malays likely originated from both migrations, the ancient migration and a more recent migration of Indians into Malaysia.

A study on the distribution of *H. pylori* virulence factor *cagA* among Vietnamese identified 84% of the strains harbouring the type II of the *cag*-right motif [29] which is characteristic for East Asian strains (hpEastAsia), ranging from 76% in Ho Chi Minh city in South Vietnam to 93% in Hanoi in North Vietnam. However, there was a remarkable difference in the frequency of *cag*-right motif of type I which is predominant in European (hpEurope) strains. While the type I motif was absent from North Vietnam, it was found in 8/49 (16%) of the samples from Ho Chi Minh city near the Mekong delta. Interestingly, prior to annexation by the

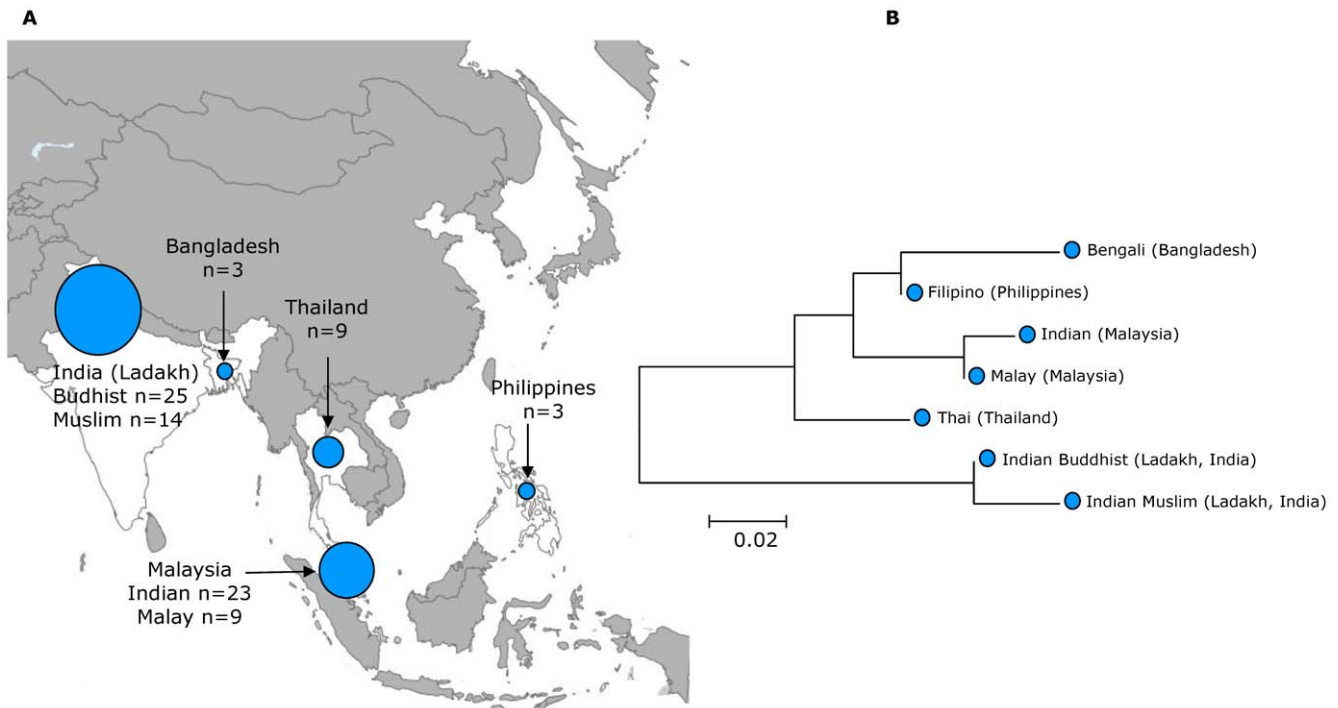
Vietnamese in the 17<sup>th</sup> century, this city was an important Khmer sea port known as Prey Nokor [4]. Thus, hpEurope strains also seem to be frequent among Vietnamese in the Mekong delta, and thus the Annamite mountain range that originates in the Tibetan and Yunnan regions of southwest China and forms Vietnam's border with Laos and Cambodia seem to have shaped an effective natural barrier for the containment of Indian influence in the Mekong basin, explaining the low prevalence of hpEurope strains elsewhere in Vietnam.

Our data are the first evidence of an important ancient genetic Indian influx this far south in Southeast Asia, except for some archaeological data. Recent excavations in peninsular Thailand have provided convincing evidence that there was a settlement there from the 3<sup>rd</sup> century BC of Indian artisans, probably of south Indian origin. Then, there was continuing Indian contact through trade and settlement throughout the period up to and including Angkor in Cambodia as well [6,7,8,9]. These data are in contrast to studies on the frequencies of human mtDNA haplotypes, which despite larger sample sizes and a larger number of nearby sampling locations, showed that the genetic makeup of South-East Asian populations from Cambodia, Laos and Vietnam was largely autochthonous [10,11,12]. An analysis of glucose-6-phosphate dehydrogenase (G6PD) deficiency alleles in Malaysian Malays [30] identified common Southeast Asian variants (52% of

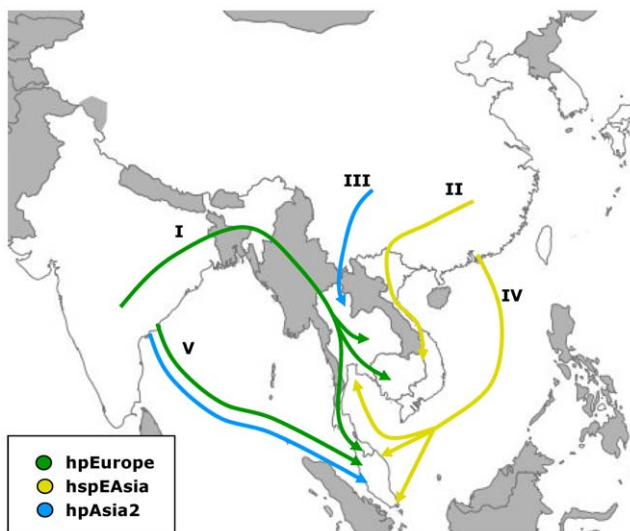
**Table 2.** Pair-wise  $F_{ST}$  values between *H. pylori* of the population hpEurope from Europe, the Middle East and Asia.

Country (No.)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
Iran (125)	1	-																				
Turkey (18)	2	0.003	-																			
Estonia (10)	3	<b>0.051</b>	<b>0.045</b>	-																		
Finland (10)	4	<b>0.043</b>	<b>0.023</b>	<b>0.013</b>	-																	
Germany (23)	5	<b>0.016</b>	<b>0.022</b>	<b>0.023</b>	<b>0.035</b>	-																
Netherlands (5)	6	0.002	0.003	0.030	0.028	-0.005	-															
UK (16)	7	<b>0.042</b>	<b>0.028</b>	<b>0.022</b>	0.014	<b>0.025</b>	0.020	-														
Kazakhstan (6)	8	0.013	-0.004	0.012	0.007	0.023	0.030	0.023	-													
Russia (21)	9	<b>0.015</b>	<b>0.025</b>	<b>0.015</b>	<b>0.030</b>	0.006	0.002	<b>0.027</b>	0.002	-												
Lebanon (5)	10	0.019	0.024	<b>0.075</b>	<b>0.061</b>	<b>0.052</b>	0.013	<b>0.047</b>	0.025	<b>0.042</b>	-											
Palestine (11)	11	<b>0.026</b>	<b>0.052</b>	<b>0.102</b>	<b>0.096</b>	<b>0.060</b>	<b>0.047</b>	<b>0.086</b>	<b>0.052</b>	<b>0.047</b>	-0.007	-										
France (8)	12	0.011	<b>0.027</b>	<b>0.032</b>	<b>0.029</b>	0.007	-0.015	0.022	0.012	0.002	0.025	<b>0.032</b>	-									
Italy (9)	13	<b>0.014</b>	<b>0.036</b>	<b>0.063</b>	<b>0.065</b>	0.019	0.006	<b>0.042</b>	<b>0.037</b>	<b>0.022</b>	0.017	<b>0.023</b>	0.004	-								
Spain Spainiard (35)	14	<b>0.026</b>	<b>0.050</b>	<b>0.067</b>	<b>0.062</b>	<b>0.026</b>	0.018	<b>0.064</b>	<b>0.043</b>	<b>0.026</b>	<b>0.035</b>	<b>0.030</b>	-0.003	0.015	-							
Spain Basque (44)	15	<b>0.040</b>	<b>0.058</b>	<b>0.074</b>	<b>0.083</b>	<b>0.040</b>	<b>0.037</b>	<b>0.075</b>	<b>0.060</b>	<b>0.041</b>	<b>0.044</b>	<b>0.041</b>	0.007	<b>0.032</b>	0.004	-						
Thailand (6)	16	<b>0.036</b>	<b>0.057</b>	<b>0.061</b>	<b>0.060</b>	<b>0.040</b>	0.014	<b>0.042</b>	0.031	<b>0.040</b>	<b>0.038</b>	<b>0.060</b>	-0.005	0.022	0.020	<b>0.045</b>	-					
India (22)	17	<b>0.031</b>	<b>0.045</b>	<b>0.057</b>	<b>0.059</b>	<b>0.026</b>	0.023	<b>0.049</b>	<b>0.039</b>	<b>0.037</b>	<b>0.044</b>	<b>0.048</b>	0.013	<b>0.026</b>	<b>0.030</b>	<b>0.049</b>	0.018	-				
Malaysia Indian (8)	18	<b>0.028</b>	<b>0.036</b>	<b>0.050</b>	<b>0.040</b>	0.020	0.027	<b>0.043</b>	0.024	<b>0.034</b>	0.049	<b>0.053</b>	0.011	0.030	<b>0.024</b>	<b>0.045</b>	0.006	0.013	-			
Malaysia Malay (4)	19	<b>0.076</b>	<b>0.067</b>	<b>0.085</b>	0.077	<b>0.079</b>	0.099	0.058	0.037	<b>0.093</b>	0.085	<b>0.128</b>	0.043	0.077	<b>0.080</b>	<b>0.088</b>	0.034	<b>0.074</b>	0.059	-		
Cambodia (34)	20	<b>0.065</b>	<b>0.076</b>	<b>0.092</b>	<b>0.079</b>	<b>0.069</b>	<b>0.066</b>	<b>0.070</b>	<b>0.077</b>	<b>0.068</b>	<b>0.075</b>	<b>0.095</b>	<b>0.035</b>	<b>0.063</b>	<b>0.060</b>	<b>0.078</b>	0.012	<b>0.049</b>	<b>0.051</b>	<b>0.061</b>	-	
Philippines (7)	21	<b>0.068</b>	<b>0.106</b>	<b>0.110</b>	<b>0.123</b>	<b>0.062</b>	<b>0.066</b>	<b>0.092</b>	<b>0.111</b>	<b>0.062</b>	<b>0.069</b>	<b>0.070</b>	0.030	<b>0.037</b>	0.016	0.019	<b>0.076</b>	<b>0.055</b>	<b>0.066</b>	<b>0.123</b>	<b>0.107</b>	-

Significant  $F_{ST}$ s (significance level = 0.05) are shown in bold.  
doi:10.1371/journal.pone.0022058.t002



**Figure 3. The distribution hpAsia2 haplotypes in Southeast Asia.** (A) Sampling locations of hpAsia2 haplotypes in Southeast Asia. (B) A neighbor-joining tree constructed from pair-wise  $F_{ST}$  values of hpAsia2 haplotypes indicates a common ancestral origin of hpAsia2 strains from Thailand, Bangladesh, Malaysia and the Philippines. doi:10.1371/journal.pone.0022058.g003



**Figure 4. Human migrations in Southeast Asia as proposed from *H. pylori* sequences.** I) An ancient migration from India distributed hpEurope bacteria in Southeast Asia. II) An ancient migration of Austro-Asiatic speakers from China carrying bacteria of the population hspEAsia. III) A migration of Tai-Kadai speakers introduced hpAsia2 bacteria in Thailand. IV) Recent migrations of Chinese from the Guangdong and Fujian provinces spread hspEAsia bacteria in Malaysia and Thailand within the last 200 years. V) Recent migrations of Indians to Malaysia brought both hpEurope and hpAsia2 bacteria to Malaysia. doi:10.1371/journal.pone.0022058.g004

the total variants) that also suggested a shared ancestral origin with the Cambodians, Laotians and Thais. Interestingly, a “Mediterranean” variant that accounts for 27% of the disease alleles among Malays [30] which is also present at low frequency in Thailand [31] and among Mon from Myanmar [32], is the most frequent variant among Indian caste groups [33]. However, this variant was not found among Khmer from Cambodia [34], and hence the “Mediterranean” G6PD deficiency allele does probably not reflect the ancient Indian genetic influx in Southeast Asia. Thus, our analysis and previous studies [16,19,25,35] demonstrate that *H. pylori* genetic diversity has more discriminatory power than traditional human genetic markers in distinguishing the sources of relatively recent human migrations.

#### Asian *H. pylori* in Southeast Asia

Vietnamese (Vietnam) and Khmer (Cambodia) are related languages in the sub-family Mon-Khmer of the Austro-Asiatic language family [36]. Since strains of the population hpEastAsia, subpopulation hspEAsia, were previously described as the predominant *H. pylori* in Vietnam [13], we anticipated Khmers also to carry *H. pylori* of this population which was indeed the case. Recent attention has focussed on the question of localising the Austro-Asiatic homeland, and interdisciplinary research sought evidence from linguistics, genetics, and archaeology [37,38]. Here, we analyzed pairwise  $F_{ST}$  values using concatenated sequences of hspEAsia strains from Cambodia (32 strains), Vietnam (20 strains), Thailand (18 strains), Malaysia (25 strains), Singapore (9 strains), Japan (24 strains), Korea (10 strains), Taiwan (15 strains) and various geographic locations in China (93 strains) (Figure 2A). For comparison, we added isolates of the hspMaori population (76 strains) from native Taiwanese, Melanesians, Samoans and New Zealand Maoris, as well as isolates of the hspAmerind population

(18 strains) from North and South America. The tree (Figure 2B) displayed three distinct clusters that corresponded to the three subpopulations hspEAsia (found in East Asians), hspMaori (Pacific islanders) and hspAmerind (Native Americans) in agreement with AMOVA analyses (Table S2). Within hspMaori, the tree reflects the trajectory of the Austronesian expansion that started from Taiwan and dispersed one of several hspMaori clades along with one of several subgroups of the Austronesian language family into Melanesia and Polynesia [16]. Although our data are not conclusive on the source of the Austro-Asiatic expansion, the tree topology of the subcluster hspEAsia (Figure 2C) that was supported by AMOVA analyses (Table S3) is consistent with the hypothesis that ancestors of the Austro-Asiatic people migrated from southern China into Southeast Asia, introducing hspEAsia bacteria into Vietnam and Cambodia. This language family might have been spread together with rice agriculture as part of a Neolithic human diaspora from the Yangzi and Yellow River Basins in China into Southeast Asia. The settlement of Southeast Asia has been dated from about 2000 BC [39,40].

The origin of the hspEAsia strains from Malaysia, Thailand and Singapore is different as those were isolated from patients with Chinese origin or ancestry [13,18] and thus reflect recent migrations within the last 200 years. Accordingly, they clustered with recent isolates from China (Figure 2B), particularly from Guangzhou and Hongkong (historically both Guangdong province), in perfect agreement with the historical origin of Malayan Chinese and Thai Chinese in China as the most of them came from Guangdong and the neighboring province Fujian. Immigrants from the same provinces made up the majority of the today's Taiwan Chinese which is also reflected in the tree.

The remaining *H. pylori* strains isolated from Malaysia and Thailand were assigned to hpAsia2. If an ancient migration from India was the source, hpAsia2 strains would be expected to be widespread in Cambodia. However, this genetic population was absent in isolates from Khmer people. Then, we calculated pairwise  $F_{ST}$  values between pairs of labelled populations from Thailand (9 strains), from Malaysia (32 strains), from Bangladesh (3 strains), from North India (Ladakh) (39 strains) and from the Philippines (3 strains), and generated a neighbor-joining tree (Figure 3A). As expected, isolates from Buddhists and Muslims from Ladakh in North India clustered together. However, due to substantial introgression of nucleotides from East Asian *H. pylori* [13,19], these isolates are strongly differentiated from other hpAsia2 populations. hpAsia2 strains from Thailand, Bangladesh, Malaysia and the Philippines clustered together in the neighbor-joining tree indicating a common ancestral origin, which was supported by the AMOVA analyses (Table S4). Based on the tree topology and the absence of hpAsia2 strains in Vietnam and Cambodia (Figure 3B), we propose that two migrations introduced hpAsia2 strains into Southeast Asia, a first migration of the ancestors of the Thai people during the early second millennium AD from southern China into Thailand [3,4,5], and a recent migration of Indians to Malaysia (see above), carrying the bacteria into a pre-existing Malay population with low *H. pylori* carriage, in agreement with Tay et al [18].

### Strain competition and subversion, host range expansion

The absence of Western Asian lineages in human mtDNA from Southeast Asia [10,11,12] indicates that this ancient migration from India alone does not explain such a high frequency of hpEurope strains. Host range expansion has been described in South-America with the displacement of hspAmer-

ind strains by hpEurope strains due to strain competition or strain subversion by transformation, integrating DNA from other strains [14,41,42,43]. Inter-strain recombination which has been identified as the major driving force behind allelic diversity in *H. pylori* is critically dependent on the frequent occurrence of mixed infections, which seem to be common in developing countries [44,45,46]. The re-shuffling of the genetic material generates organisms that can inhabit a wide array of niches (generalist strains), and the fittest strains, e.g. of the population hpEurope, will eventually outcompete the specialist strains, e.g. of the population hspAmerind, that lack the necessary genetic diversity to efficiently colonize a wide host spectrum (specialist strains) [42].

The low prevalence of hspEAsia strains among ethnic Thai (0 out of 14 strains) [13,47] and Malays (2 out of 15 strains) [18], despite early Chinese and Khmer influences [4,5,27], indicates specialist strains with a lower ability to adapt to a wide range of human hosts. In contrast, the observed host range expansion of hpEurope strains in Southeast Asia, as well as their spread among South American Amerinds and mestizos [42], indicates that these are generalist strains with a broad host spectrum. Subversion of hpEastAsia strains by transformation with DNA from hpEurope strains eventually changes those into hpEurope strains thereby further broadening the host range. The high frequency of hpAsia2 strains in Malays (9 out of 15 strains) suggests strains with a higher ability than hspEAsia strains to adapt to a wide range of human hosts and/or stronger interactions between Malays and Malaysian Indians than between Malays and Malaysian Chinese.

### *H. pylori* populations and the incidence rate of gastric cancer

Gastric carcinoma (GC), the fourth most common cancer worldwide is the second leading cause of cancer-related deaths [48]. The highest age standardized incidences (ASR) have been described in Asia but regional variations exist [49] that do not match the distribution of infection prevalence rates except for Malaysia [50]. Even if the clinical outcome of *H. pylori* infection is a complex process, the regional variations of GC incidence within Asia seem to be closely related to the distribution of the *H. pylori* genetic populations. In countries where almost all the strains are assigned to hspEAsia (Japan, China, Korea, and Vietnam) [14], the incidence of GC is high (ASR 18.9 to 41.4/100 000). In contrast, incidence is low (ASR 3.5 to 5.2/100 000) in countries with a high proportion of hpAsia2 or hpEurope strains (India and Thailand) [13,17,18]. Cambodia that displays a mixture of hpEurope and hspEAsia strains is classified among countries with intermediate risk of GC (ASR 9.8/100 000) [49]. The genetic background might be a marker of virulence factors directly involved in clinical outcome. Further studies are needed to investigate *H. pylori* virulence factors. In the future, human mobility combined with the host range expansion of hpEurope strains may accelerate the genetic admixture of *H. pylori* populations, and thus may have a significant impact on GC incidence in Asia.

In conclusion, Southeast Asia was probably free of *H. pylori* before major human migrations. These movements included (Figure 4) i) an ancient migration from India introducing hpEurope bacteria into Thailand, Cambodia and Malaysia; ii) an ancient migration of the ancestors of Austro-Asiatic people from China into Vietnam and Cambodia carrying hspEAsia bacteria; iii) an ancient migration of the ancestors of the Thai people into Thailand carrying *H. pylori* of population hpAsia2; iv) a recent migration of Chinese from the Guangdong and Fujian provinces into Southeast Asia spreading hspEAsia strains; and v) a



recent migration of Indians to Malaysia carrying both hpAsia2 and hpEurope bacteria.

## Supporting Information

**Figure S1 DISTRUCT plot of the proportions of ancestral nucleotides in *H. pylori* isolates from India, Thailand, Cambodia, Vietnam and China according to the ethnic group or the religion, as determined by Structure V2.0 (linkage model).** A vertical line for each isolate indicates the estimated amount of ancestry from each ancestral population as five coloured segments. Vertical black lines separate the individuals into (sub)-populations, as determined by the no-admixture model in STRUCTURE V2.0. (TIF)

**Table S1 AMOVA analyses for hpEurope isolates.** (DOC)

**Table S2 AMOVA analyses for hpEastAsia isolates.** (DOC)

**Table S3 AMOVA analyses for hpEastAsia.** (DOC)

## References

- LeBar F, Hickey G, Musgrave J (1964) Ethnic groups of mainland Southeast Asia. New Haven: Human Relations Area Files Press. 288 p.
- Kumar V, Reddy AN, Babu JP, Rao TN, Langstieh BT, et al. (2007) Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol Biol* 7: 47.
- Eberhard W (1977) A history of China. 4th edition. Berkeley: University of California Press. 382 p.
- Higham C (1989) The archaeology of mainland Southeast Asia: from 10,000 B.C. to the fall of Angkor. Cambridge: Cambridge University Press. 404 p.
- Lertrit P, Poolsuwan S, Thosarat R, Sanpachudayan T, Boonyarit H, et al. (2008) Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. *Am J Phys Anthropol* 137: 425–440.
- Bellina B (2002) Le port protohistorique de Khao Sam Kaeo en Thaïlande péninsulaire: lieu privilégié pour l'étude des premières interactions indiennes et sud-est asiatiques. *Bulletin de l'École française d'Extrême-Orient* 89: 329–343.
- Bellina B (2007) Cultural Exchange Between India and Southeast Asia. Production and Distribution of Hard Stone Ornaments (6 c. BC–6 c. AD). Paris: Maison des Sciences de l'Homme. 126 p.
- Bellina-Pryce B, Silapanth P (2006) Weaving cultural identities on trans-Asiatic networks: Upper Thai-Malay peninsula – an early socio-political landscape. *Bulletin de l'École française d'Extrême Orient* 93: 257–294.
- Bellina B, Silapanth P (2006) Khao Sam Kaeo and the Upper Thai Peninsula: understanding the mechanism of early trans-asiatic trade and cultural exchange. In: Bacus EA, Glover IC, Pigott VC, eds. *Uncovering Southeast Asia's Past*. Singapore: National University Press. pp 379–392.
- Soares P, Trejaut JA, Loo JH, Hill C, Mormina M, et al. (2008) Climate change and postglacial human dispersals in southeast Asia. *Mol Biol Evol* 25: 1209–1218.
- Hill C, Soares P, Mormina M, Macaulay V, Clarke D, et al. (2007) A mitochondrial stratigraphy for island southeast Asia. *Am J Hum Genet* 80: 29–43.
- Hill C, Soares P, Mormina M, Macaulay V, Meehan W, et al. (2006) Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol Biol Evol* 23: 2480–2491.
- Linz B, Balloux F, Moodley Y, Manica A, Liu H, et al. (2007) An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445: 915–918.
- Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, et al. (2003) Traces of human migrations in *Helicobacter pylori* populations. *Science* 299: 1582–1585.
- Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, et al. (1999) Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol* 32: 459–470.
- Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, et al. (2009) The peopling of the Pacific from a bacterial perspective. *Science* 323: 527–530.
- Devi SM, Ahmed I, Francalacci P, Hussain MA, Akhter Y, et al. (2007) Ancestral European roots of *Helicobacter pylori* in India. *BMC Genomics* 8: 184.
- Tay CY, Mitchell H, Dong Q, Goh KL, Dawes IW, et al. (2009) Population structure of *Helicobacter pylori* among ethnic groups in Malaysia: recent acquisition of the bacterium by the Malay population. *BMC Microbiol* 9: 126.
- Wirth T, Wang X, Linz B, Novick RP, Lum JK, et al. (2004) Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. *Proc Natl Acad Sci U S A* 101: 4746–4751.
- Momynaliev KT, Chelysheva VV, Akopian TA, Selezneva OV, Linz B, et al. (2005) Population identification of *Helicobacter pylori* isolates from Russia. *Genetika* 41: 1434–1437.
- Liao YL, Guo G, Mao XH, Xie QH, Zhang WJ, et al. (2009) Core genome haplotype diversity and vacA allelic heterogeneity of Chinese *Helicobacter pylori* strains. *Curr Microbiol* 59: 123–129.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
- Rosenberg N (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137–138.
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1: 47–50.
- Latifi-Navid S, Ghorashi SA, Siavoshi F, Linz B, Massarrat S, et al. (2010) Ethnic and geographic differentiation of *Helicobacter pylori* within Iran. *PLoS One* 5: e9645.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
- Vlieland CA (1934) The Population of the Malay Peninsula: A Study in Human Migration. *Geogr Rev* 24: 61–78.
- Ahmed N, Dobrindt U, Hacker J, Hasnain SE (2008) Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention. *Nat Rev Microbiol* 6: 387–394.
- Uchida T, Nguyen LT, Takayama A, Okimoto T, Kodama M, et al. (2009) Analysis of virulence factors of *Helicobacter pylori* isolated from a Vietnamese population. *BMC Microbiol* 9: 175.
- Ainon O, Yu YH, Amir Muhriz AL, Boo NY, Cheong SK, et al. (2003) Glucose-6-phosphate dehydrogenase (G6PD) variants in Malaysian Malays. *Hum Mutat* 21: 101.
- Laosombat V, Sattayasevana B, Janejindamai W, Viprakasit V, Shirakawa T, et al. (2005) Molecular heterogeneity of glucose-6-phosphate dehydrogenase (G6PD) variants in the south of Thailand and identification of a novel variant (G6PD Songklanagarind). *Blood Cells Mol Dis* 34: 191–196.
- Nuchprayon I, Louicharoen C, Charoenvej W (2008) Glucose-6-phosphate dehydrogenase mutations in Mon and Burmese of southern Myanmar. *J Hum Genet* 53: 48–54.
- Mohanty D, Mukherjee MB, Colah RB (2004) Glucose-6-phosphate dehydrogenase deficiency in India. *Indian J Pediatr* 71: 525–529.
- Matsuoka H, Nguon C, Kanbe T, Jalloh A, Sato H, et al. (2005) Glucose-6-phosphate dehydrogenase (G6PD) mutations in Cambodia: G6PD Viangchan (871G>A) is the most common variant in the Cambodian population. *J Hum Genet* 50: 468–472.
- Moodley Y, Linz B (2009) *Helicobacter pylori* Sequences Reflect Past Human Migrations. *Genome Dyn* 6: 62–74.
- Lewis P (2009) *Ethnologue: Languages of the World*. Dallas: SIL International. 1248 p.

**Table S4 AMOVA analyses for hpAsia2.** (DOC)

## Acknowledgments

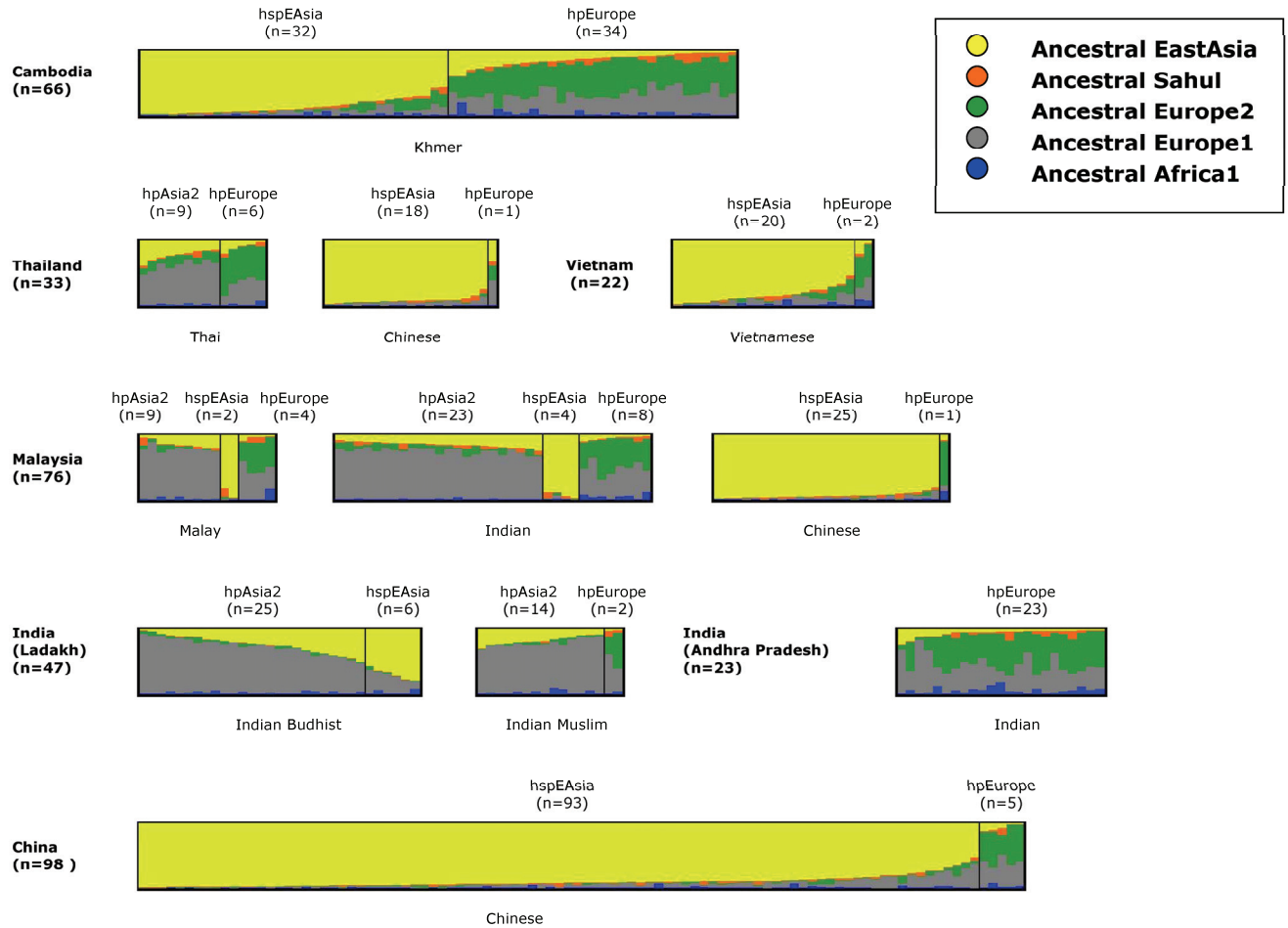
We thank Pr Charles Higham (Department of Anthropology, University of Otago, New Zealand), Dr Maru Mormina (Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, United Kingdom) and Benoit Garin (Institut Pasteur de Madagascar) for helpful discussions, and all the clinicians involved in the conduct of this study. We would also like to acknowledge the different groups that provided the sequences included in this study, particularly S. Manjulata Devi, Irshad Ahmed, Paolo Francalacci, M. Abid Hussain, Yusuf Akhter, Ayesha Alvi, Leonardo A. Sechi, Francis Mégraud and Niyaz Ahmed for sequences from Indian strains.

## Author Contributions

Conceived and designed the experiments: S. Breurec DM BG JR MH BL. Performed the experiments: S. Breurec BG SH S. Brisse FBD MH CO TST JMT SV DM. Analyzed the data: S. Breurec S. Brisse FBD JMT BL. Contributed reagents/materials/analysis tools: S. Breurec BG SH S. Brisse FBD MH CO TST JMT SV DM BL. Wrote the paper: S. Breurec BL. Found funding: S. Breurec JR.

37. Sagart L, Blench R, Sanchez-Mazas A (2005) The peopling of East Asia: putting together archaeology, linguistics and genetics. London: Routledge Curzon. 360 p.
38. Jin L, Seielstad M, Xiao C (2001) Genetic, Linguistic and Archaeological Perspectives on Human Diversity in Southeast Asia. River Edge, New Jersey: World Scientific Publishing Co. 172 p.
39. Higham C (2001) Prehistory, language and human biology: is there a consensus in East and Southeast Asia. In: Jin L, Seielstad M, Xiao C, eds. Genetic, linguistic and archaeological perspectives on human diversity in Southeast Asia. New Jersey: World Scientific. pp 3–16.
40. Higham C (2003) Languages and farming dispersals: Austroasiatic languages and rice cultivation. In: Bellwood P, Renfrew C, eds. Examining the farming/language dispersal hypothesis. Cambridge: McDonald Institute for Archaeological Research. pp 223–232.
41. Ghose C, Perez-Perez GI, Dominguez-Bello MG, Pride DT, Bravi CM, et al. (2002) East Asian genotypes of *Helicobacter pylori* strains in Amerindians provide evidence for its ancient human carriage. Proc Natl Acad Sci U S A 99: 15107–15111.
42. Dominguez-Bello MG, Perez ME, Bortolini MC, Salzano FM, Pericchi LR, et al. (2008) Amerindian *Helicobacter pylori* strains go extinct, as european strains expand their host range. PLoS One 3: e3307.
43. Yamaoka Y, Orito E, Mizokami M, Gutierrez O, Saitou N, et al. (2002) *Helicobacter pylori* in North and South America before Columbus. FEBS Lett 517: 180–184.
44. Ghose C, Perez-Perez GI, van Doorn IJ, Dominguez-Bello MG, Blaser MJ (2005) High frequency of gastric colonization with multiple *Helicobacter pylori* strains in Venezuelan subjects. J Clin Microbiol 43: 2635–2641.
45. Schwarz S, Morelli G, Kusecek B, Manica A, Balloux F, et al. (2008) Horizontal versus familial transmission of *Helicobacter pylori*. PLoS Pathog 4: e1000180.
46. Morales-Espinosa R, Castillo-Rojas G, Gonzalez-Valencia G, Ponce de Leon S, Cravioto A, et al. (1999) Colonization of Mexican patients by multiple *Helicobacter pylori* strains with different *vacA* and *cagA* genotypes. J Clin Microbiol 37: 3001–3004.
47. Vilaichone RK, Mahachai V, Tumwasorn S, Wu JY, Graham DY, et al. (2004) Molecular epidemiology and outcome of *Helicobacter pylori* infection in Thailand: a cultural cross roads. Helicobacter 9: 453–459.
48. Parkin DM (2004) International variation. Oncogene 23: 6329–6340.
49. Ferlay J, Shin H, Bray F, Forman D, Mathers C, et al. (2008) GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet]. Lyon, France: International Agency for Research on Cancer; 2010. Available: <http://globocan.iarc.fr>. Accessed 2011 June 7.
50. Fock KM, Ang TL (2010) Epidemiology of *Helicobacter pylori* infection and gastric cancer in Asia. J Gastroenterol Hepatol 25: 479–486.

## Supporting Information Legends section



**Figure S1. DISTRUCT plot of the proportions of ancestral nucleotides in *H. pylori* isolates from India, Thailand, Cambodia, Vietnam and China according to the ethnic group or the religion, as determined by Structure V2.0 (linkage model). A vertical line for each isolate indicates the estimated amount of ancestry from each ancestral population as five coloured segments. Vertical black lines separate the individuals into (sub)-populations, as determined by the no-admixture model in STRUCTURE V2.0.**

**Table S1. AMOVA analyses for hpEurope isolates**

<b>Source</b>	<b>Source assigned to group</b>						
Estonia	1	1	1	1	1	1	1
Finland	1	1	1	1	1	1	1
UK	1	1	1	1	1	1	1
Germany	1	1	1	1	1	1	1
Netherlands	1	1	1	1	1	1	1
Russia	1	1	1	1	1	1	1
Kazakhstan	1	1	1	1	1	1	1
Iran	1	1	1	1	1	1	1
Turkey	1	1	1	1	1	1	1
Italy	1	1	1	1	1	1	1
Lebanon	1	1	1	1	1	1	1
Palestine	1	1	1	1	1	1	1
France	2	2	2	2	2	2	4
Spaniard	2	2	2	2	2	2	2
Basque	2	2	2	2	2	2	2
Philippines	2	3	2	2	2	2	2
Indian	3	3	3	3	3	3	3
Malaysia	3	3	3	3	3	3	3
Indian							
Malaysia	3	3	3	2	4	3	4
Malay							
Thailand	3	3	3	2	4	4	4
Cambodia	3	3	2	2	4	4	4
<b><math>F_{ST}</math></b>	0.04841	0.04701	0.04529	0.04539	<b>0.04948</b>	0.04923	0.04884

The highest  $F_{ST}$  value (in bold) indicates the most likely group assignment supporting the Neighbor-joining tree in Fig. 1.

**Table S2. AMOVA analyses for hpEastAsia isolates**

<i>Source</i>	<b>Source assigned to group number</b>					
Amerindian North	1	1	1	1	1	1
Amerindian South	1	1	1	1	1	1
Taiwanese Aboriginal	2	1	1	1	1	2
Melanesian Malanesia	2	2	2	2	2	2
Polynesian Wallis	2	2	2	2	2	2
Polynesian Samoa	2	2	2	2	2	2
New Zealand Maori	2	2	2	2	2	2
Korean Korea	3	3	3	3	3	3
Japanese Hoshu	3	3	3	3	3	3
Japanese Hokkaido	3	3	3	3	3	3
Japanese Okinawa	3	3	3	3	3	3
Indian Ladakh	3	3	3	3	3	3
Chinese Beijing	3	3	3	3	3	3
Chinese Heilongjiang	3	3	3	3	3	3
Chinese Hangzhou	3	3	3	3	3	3
Chinese Xian	3	3	3	3	3	3
Chinese Yunnan	3	3	3	3	3	3
Chinese Chongqing	3	3	3	3	3	3
Chinese HongKong	3	4	4	4	4	3
Chinese Guangzhou	3	4	4	4	4	3
Chinese Taiwan	3	4	4	4	4	3
Chinese Thailand	3	4	5	5	4	3
Chinese Malaysia	3	4	5	5	4	3
Chinese Siangapore	3	4	5	5	4	3
Vietnamese Vietnam	3	3	3	3	4	2
Khmer Cambodia	3	3	3	3	4	2
<b><i>F<sub>ST</sub></i></b>	<b>0.23006</b>	<b>0.17866</b>	<b>0.17514</b>	<b>0.17514</b>	<b>0.16868</b>	<b>0.16715</b>

**Table S3. AMOVA analyses for hpEastAsia**

<i>Source</i>	<b>Source assigned to group number</b>					
Chinese Beijing	1	1	1	1	1	1
Chinese Heilongjiang	1	1	1	1	1	1
Chinese Hangzhou	1	1	1	1	1	1
Chinese Xian	1	1	1	1	1	1
Chinese Yunnan	1	1	1	3	1	1
Chinese Chongqing	2	2	2	2	2	2
Chinese HongKong	2	2	2	2	2	2
Chinese Guangzhou	2	2	2	2	3	2
Chinese Taiwan	2	2	2	2	2	2
Chinese Thailand	2	2	2	2	2	3
Chinese Malaysia	2	2	2	2	2	3
Chinese Siangapore	2	2	2	2	2	3
Vietnamese	2	1	3	3	3	2
Khmer Cambodia	2	1	3	3	3	2
<b><i>F<sub>ST</sub></i></b>	<b>0.03597</b>	0.03247	0.03226	0.03104	0.03011	0.03264

**Table S4. AMOVA analyses for hpAsia2**

<b>Source</b>	<b>Source assigned to group number</b>					
Indian Buddhist, Ladakh	1	1	2	1	1	1
Indian Muslim, Ladakh	1	1	1	1	1	1
Thai Thailand	2	2	2	2	2	2
Malay Malaysia	2	2	2	2	2	3
Indian Malaysia	2	2	1	1	1	3
Philippino Phillipines	2	3	2	2	3	3
Bengali Bangladesh	2	3	2	2	3	3
<b><i>F<sub>ST</sub></i></b>	<b>0.18332</b>	0.17059	0.12167	0.14854	0.14345	0.17073

**Article n°3: Expansion of European *vacA* and *cagA* alleles to East-Asian *Helicobacter pylori* strains in Cambodia**

**Breurec S, Guillard B, Hem S, Brisse S, Dieye FB, Huerre M, Monchy D, Oung C, Sreng Tan T, Thiberge JM, Vong S, Linz L, Raymond J.**

Infection, Genetics and Evolution 2011 (article sous presse).

## Contexte

Les incidences de CG les plus élevées dans le monde sont observées en Asie, mais des variations existent sur ce continent avec une diminution de l'incidence vers les latitudes plus au sud. Cette distribution ne s'explique pas par des différences de prévalence de l'infection à *H. pylori*. En partant de l'hypothèse que la distribution géographique des facteurs de virulence associés à *H. pylori* pourrait expliquer au moins partiellement celle des incidences du CG en Asie, nous avons étudié les facteurs de virulence CagA et VacA chez 66 souches isolées de patients d'ethnie khmer (Cambodge), préalablement assignées à deux (sous-) populations génétiques, hspEAsia (n=32, 48.5 %) et hpEurope (n=34, 51.5 %) (Article n°3).

## Matériels et méthodes

Les polymorphismes du gène *vacA* et le statut *cagA* des souches ont été étudiés par PCR. La région 3' du gène *cagA* a été séquencée avec des amorces dessinées spécialement pour cette étude et les motifs EPIYA ont été identifiés.

Nous avons également séquencé par la technique du « primer walking » la totalité du gène *cagA* chez 20 souches choisies au hasard parmi les 66 souches assignées aux (sous-) populations hspEAsia et hpEurope. L'arbre phylogénétique a été obtenu en utilisant le logiciel MEGA par la méthode du Neighbour-Joining avec le modèle « Kimura-2-paramètres ». L'assignement du gène *cagA* à une population génétique a été réalisé avec le logiciel STRUCTURE, modèle « no linkage », et la proportion des nucléotides dérivant de chaque population ancestrale a été estimée avec le modèle « linkage » de STRUCTURE.

## Résultats et discussion

Le gène *cagA* était présent chez 92.4 % (61/66) des souches et était significativement plus fréquent chez les souches hspEAsia (100 %) que chez les souches hpEurope (85.3 %) ( $p < 0.05$ ). Les combinaisons EPIYA prépondérantes étaient de type ABC (n=30, 45.5 %) et ABD (n=23, 34.8 %). Le typage du gène *vacA* a montré que 63 (95.5 %) souches étaient de type s1, 32 (48.5 %) étaient de type m1 et 62 (93.9 %) étaient de type i1. Le faible nombre de cas de CG (n=5) ne nous a pas permis d'étudier les facteurs de risque associés. Des études épidémiologiques, chez l'animal et *in vitro* suggèrent que les souches hspEAsia possédant un « East Asian CagA » et un allèle *vacA* de type m1 semblent plus virulentes que les autres. La faible fréquence des souches possédant les caractéristiques génétiques précédentes au Cambodge (10 sur 66, 15.2 %) comparée à celles retrouvées dans les pays d'Asie de l'est,



région de forte incidence de CG, pourrait expliquer l'incidence plus faible au Cambodge, et par extension en Asie du sud-est, Vietnam excepté.

Les souches hpEurope présentent en général un « Western CagA » et un allèle *vacA* de type s1a tandis que les souches hspEAsia possèdent un « East Asian CagA » et un allèle de type s1c dans la grande majorité des cas. La forte prévalence de l'allèle *vacA* s1a (31.4 %) et du « Western CagA » (28.1 %) chez les souches hspEAsia indique une diffusion d'allèles *cagA* et *vacA* à partir des souches « européennes » (hpEurope) vers celles d'Asie de l'est (hspEAsia) par recombinaison. L'inverse était rarement observé puisque une seule souche hpEurope présentait un « East Asian CagA » et que seules 4 souches hpEurope (11.8 %) présentaient un allèle *vacA* s1c. Cette expansion est confirmée par l'analyse de la structure génétique du gène *cagA*. Les haplotypes *cagA* des 20 souches hpEurope et hspEAsia investiguées ont été comparés à 35 haplotypes *cagA* d'une collection de souches regroupant l'ensemble des (sous-) populations génétiques possédant un *cag* PAI (hpAfrica1, hpNEAfrica, hpEurope, hpAsia2, hspEAsia et hspMaori). L'arbre phylogénétique généré à partir des séquences *cagA* montrait 2 clusters correspondant au « Western CagA » et à l'« East Asian CagA », avec 100 % d'assignement du motif EPIYA-C au « Western CagA » et du motif EPIYA-D à l'« East Asian CagA », ce qui était en parfait accord avec les résultats obtenus à partir du modèle « no linkage » du logiciel STRUCTURE (Figure 1, p17 de l'article). Nous avons alors étudié la proportion des nucléotides dérivant des populations ancestrales européennes et d'Asie de l'est (Logiciel STRUCTURE, « modèle linkage »). Sur les 20 souches étudiées, 2 présentaient un gène *cagA* mosaïque, résultat de recombinaisons importantes entre ces 2 populations ancestrales. Ces deux souches étant assignées à la sous-population hspEAsia, ces recombinaisons indiquent la subversion d'un « East Asian (hspEAsia) CagA » par transformation avec de l'ADN d'un « Western (hpEurope) CagA ».

### **Conclusion**

Le Cambodge est actuellement classé parmi les pays avec un risque intermédiaire de CG. La faible prévalence des souches hspEAsia possédant un « East Asian CagA » et un allèle *vacA* m1 pourrait l'expliquer en partie. CagA et VacA ayant des effets antagonistes, la diffusion d'allèles *cagA* et *vacA* des souches « européennes » (hpEurope) vers celles d'Asie de l'est (hspEAsia) pourrait entraîner la rupture de l'équilibre entre les effets biologiques de ces deux protéines et provoquer une dérégulation ayant des conséquences sur l'évolution de la maladie. L'incidence du CG pourrait augmenter dans les années futures au Cambodge.



Contents lists available at SciVerse ScienceDirect

## Infection, Genetics and Evolution

journal homepage: [www.elsevier.com/locate/meegid](http://www.elsevier.com/locate/meegid)Expansion of European *vacA* and *cagA* alleles to East-Asian *Helicobacter pylori* strains in Cambodia

Sebastien Breurec<sup>a,b,\*</sup>, Bertrand Guillard<sup>b</sup>, Sopheak Hem<sup>b</sup>, Konstantinos S. Papadakos<sup>c</sup>, Sylvain Brisse<sup>d</sup>, Michel Huerre<sup>e</sup>, Didier Monchy<sup>b,f</sup>, Chakravuth Oung<sup>g</sup>, Dionyssios N. Sgouras<sup>c</sup>, Tek Sreng Tan<sup>h</sup>, Jean-Michel Thiberge<sup>d</sup>, Sirenda Vong<sup>i</sup>, Josette Raymond<sup>j</sup>, Bodo Linz<sup>k</sup>

<sup>a</sup>Unité de Bactériologie Médicale et Environnementale, Institut Pasteur, 36 Avenue Pasteur, BP 220, Dakar, Senegal

<sup>b</sup>Unité de Biologie Médicale, Institut Pasteur, 5 Boulevard Monivong, BP 983, Phnom Penh, Cambodia

<sup>c</sup>Laboratory of Medical Microbiology, Hellenic Pasteur Institute, 127 Vas. Sofias Avenue, Athens 11521, Greece

<sup>d</sup>Plate-forme Génomique des Pathogènes et Santé Publique, Institut Pasteur, 25-28 rue du Docteur Roux, 75724 Paris Cedex 15, France

<sup>e</sup>Unité de Recherche et d'Expertises en Histotechnologie et Pathologie, Institut Pasteur, Paris, 25-28 rue du Docteur Roux, 75724 Paris Cedex 15, France

<sup>f</sup>Laboratoire de Biologie médicale, Institut Pasteur, BP923, Bangui, République Centrafricaine, France

<sup>g</sup>Gastroenterology and Liver Unit, Calmette Hospital, 3 boulevard Monivong, Phnom Penh, Cambodia

<sup>h</sup>Private Medical Center, No. 29 street 288, Boeng Kengkang 1, Chamkamon, Phnom Penh, Cambodia

<sup>i</sup>Unité d'Epidémiologie et de Santé Publique, Institut Pasteur, 5 Boulevard Monivong, BP 983, Cambodia

<sup>j</sup>Unité Postulante Pathogénèse de *Helicobacter*, Institut Pasteur, 25-28 rue du Docteur Roux, 75724 Paris Cedex 15, France

<sup>k</sup>Department of Biochemistry and Molecular Biology, The Pennsylvania State University, 312 Wartik Lab, University Park, PA 16802, USA

## ARTICLE INFO

## Article history:

Received 4 May 2011

Received in revised form 4 August 2011

Accepted 7 August 2011

Available online xxx

## Keywords:

*H. pylori*

*cagA*

*vacA*

Recombination

Genetic population

Cambodia

## ABSTRACT

*Helicobacter pylori* infection is associated with gastric cancer (GC). The highest incidence rates have been described in Asia, but regional variations exist that do not match the distribution of infection prevalence rates. The aim of the study was to examine the possible contribution of *H. pylori* virulence factors to geographic differences in the incidence of GC across East and Southeast Asia. We studied 66 isolates from Cambodian patients that had previously been assigned to two genetic populations based on sequences of seven housekeeping genes, namely hpEurope ( $n = 34$ , 51.5%) and hpEastAsia, subpopulation hspEAsia ( $n = 32$ , 48.5%). These strains were characterized with respect to *vacA* polymorphism and *cagA* status by PCR, and the CagA C-terminal region was sequenced. We also sequenced the complete *cagA* gene from 10 hpEurope and 10 hspEAsia strains chosen at random. The *cagA* gene was present in 92.4% of the 66 isolates and was mainly of Western type ( $n = 36$ , 59.0%). hspEAsia strains carrying East-Asian CagA and the m1-type *vacA* allele (15.2%) were less frequent among the 66 Cambodian isolates than reported in East Asian countries, a finding that might partly explain the intermediate incidence of GC in Cambodia, and by extension, in Southeast Asia (except for Vietnam). The observed high prevalence of s1a alleles (34.4%) and Western CagA (28.1%) among hspEAsia strains indicates frequent introgression of European *vacA* and *cagA* alleles into East Asian *H. pylori* strains. This expansion might have severe consequences for individual disease outcome.

© 2011 Elsevier B.V. All rights reserved.

## 1. Introduction

*Helicobacter pylori* infects the gastric mucosa of more than half the world population. Although all infected individuals develop *H. pylori*-associated gastritis, only a small percentage of colonized

individuals develop severe gastroduodenal diseases such as gastric and duodenal ulcer (10%), or gastric adenocarcinoma and MALT lymphoma (1–4%) (Hatakeyama, 2009). Outcome of the infection is dependent on host, environmental and *H. pylori* virulence factors.

*H. pylori* isolates have been subdivided into distinct (sub-) populations on the basis of sequence differences in seven core genes, with specific geographic distributions that reflect ancient human migrations (Falush et al., 2003b; Linz et al., 2007; Moodley et al., 2009; Tay et al., 2009): hpEurope (present in Europe, the Middle East, India and Iran), hpNEAfrica (present in Northeast Africa), hpAfrica1 (present in Western Africa and South Africa), hpAfrica2

Abbreviations: *cag* PAI, *cag* pathogenicity island; EPIYA, Glu-Pro-Ile-Tyr-Ala; GC, Gastric carcinoma; SH2, cytoplasmic Src homology 2; SHP-2, Src homology 2 phosphatase; NFAT, nuclear factor of activated T cells.

\* Corresponding author at: Unité de Bactériologie Médicale et Environnementale, Institut Pasteur, 36 Avenue Pasteur, BP 220, Dakar, Senegal. Tel.: +221 33 839 92 30; fax: +221 33 822 70 52.

E-mail address: [sbreurec@pasteur.sn](mailto:sbreurec@pasteur.sn) (S. Breurec).

1567-1348/\$ - see front matter © 2011 Elsevier B.V. All rights reserved.

doi:10.1016/j.meegid.2011.08.007

Please cite this article in press as: Breurec, S., et al. Expansion of European *vacA* and *cagA* alleles to East-Asian *Helicobacter pylori* strains in Cambodia. Infect. Genet. Evol. (2011), doi:10.1016/j.meegid.2011.08.007

(present in South Africa), hpAsia2 (present in Northern India, Bangladesh, Thailand and Malaysia), hpSahul (from Australian Aboriginals and Papua New Guineans) and hpEastAsia with the subpopulations hspEAsia (from East Asians), hspMaori (from Taiwanese Aboriginals, Melanesians and Polynesians) and hspAmerind (Native Americans).

The CagA protein, one of the most extensively studied *H. pylori* virulence factors, is delivered into gastric epithelial cells by the type IV secretion system, which is encoded by genes present in the *cag* pathogenicity island (*cag* PAI) (Hatakeyama, 2009). Translocated CagA undergoes tyrosine phosphorylation by multiple Src family kinases on specific tyrosine residues within repeating Glu-Pro-Ile-Tyr-Ala (EPIYA) motifs. EPIYA segments are designated EPIYA-A, -B -C and -D according to the amino acid sequence that surrounds the EPIYA motif (Hatakeyama, 2009). The EPIYA-A and -B motifs are found in almost all *H. pylori* strains worldwide. In addition, the EPIYA-C segment is found in “Western CagA” in hpAfrica1, hpEurope, hpNEAfrica strains and some hpAsia2 strains from Southeast Asia, while the EPIYA-D segment is characteristic for “East-Asian CagA” in hpEastAsia strains and in some hpAsia2 strains from Ladakh in Northern India. The only available to date *cagA* sequence of a hpSahul strain possessed the EPIYA-A and EPIYA-B motifs but neither EPIYA-C nor EPIYA-D. In addition, none of the hpAfrica2 strains possessed the *cag* PAI (Olbermann et al., 2010). Once phosphorylated, CagA binds to a cytoplasmic Src homology 2 (SH2) domain of the Src homology 2 phosphatase (SHP-2) and deregulates SHP-2. CagA also binds to and inhibits PAR1b/MARK2 polarity-regulating kinase, which leads to disruption of apical junctions and epithelial-basolateral polarity (Isomoto et al., 2010). Together, these CagA activities may contribute to neoplastic transformation of gastric epithelial cells.

A second major virulence factor is the cytotoxin VacA, a protein that can promote cell death through apoptosis, suppress the local immune system, disrupt endocytic trafficking in host cells, and induce vacuolisation of human epithelial cells *in vitro* (Isomoto et al., 2010). Virtually all strains carry a functional *vacA* gene but not all encoded VacA proteins induce vacuolisation, thus pointing to *vacA* genetic variability. Three major polymorphic VacA protein regions have been identified: the signal region (s1 and s2), the intermediate region (i1 and i2) and the middle region (m1 and m2). s1/i1/m1-or-m2 VacA isotypes generally induce high levels of vacuolisation, whereas s1-or-s2/i2/m1-or-m2 strains induce little if any vacuolisation (Rhead et al., 2007).

Gastric carcinoma (GC), the fourth most common cancer worldwide, is the second cause of cancer-related deaths (Parkin, 2004). Age-standardized incidence rates (ASR) are highest in East Asia, but they decline at increasingly southerly latitudes (Ferlay et al., 2010). Variations of ASRs do not match variations of infection prevalence rates, except in Malays (Fock and Ang, 2010), indicating that other factors are also involved. Mongolia, Japan, Korea, China and Vietnam are classified as countries with a high risk of GC (ASR 18.9 to 41.4/100 000), while Cambodia (South East Asia) is classified as a country with an intermediate risk of GC (ASR 9.8/100 000), along with Myanmar (ASR 11.0/100 000) and Singapore (ASR 10.9/100 000). The risk is low in Bangladesh, Laos, Thailand and India (ASR 3.5 to 5.2/100 000).

Cambodia borders Thailand to the west and northwest and Vietnam to the east and southeast, and is therefore situated between regions with low and high risks of GC. Cambodia's population is mainly of Khmer origin and speaks the Khmer language related to the Vietnamese language (Vietnam) in the sub-family Mon-Khmer of the Austro-Asiatic language family (Lewis, 2009). Here we characterized *H. pylori* clinical isolates from Cambodia with respect to their *cagA* status, the diversity of the *cagA* gene 3' region, and *vacA* s, m, and i region polymorphisms, in order to examine the possible contribution of these factors to geographic differences in

the incidence of GC across East and South-East Asia. We also examined the *cagA* population genetic structure of 20 randomly selected strains.

## 2. Materials and methods

### 2.1. Patients

We studied isolates from patients presenting with upper abdominal pain to the Gastroenterology Department of Calmette Hospital ( $n = 37$ ) or a private medical center ( $n = 29$ ), both located in Phnom Penh, Cambodia, in 2004 and 2007. All the patients were of Cambodian origin, and none had received proton pump inhibitors or antibiotics within 4 weeks before endoscopy. Three biopsy samples each were taken from the antrum and the fundus during upper gastrointestinal tract endoscopy. One biopsy from each site was cultured for *H. pylori* isolation, and the four others were fixed and processed for histological analysis. All endoscopes and accessory devices were decontaminated after every endoscopy according to local written procedures. Written informed consent was obtained from all the patients and the study protocol was approved by the Cambodian national medical ethics committee for Health Research under ethics certificate 017/03NECHR.

### 2.2. Histology

Gastric biopsies were fixed then embedded and studied after usual staining. The classification of gastritis and other lesions was performed according to the criteria of updated Sydney system (Fenoglio-Preiser et al., 1999).

### 2.3. *H. pylori* isolation, genomic DNA extraction, and determination of *cagA* status and *cagA* and *vacA* polymorphism

*H. pylori* culture was performed using Wilkins Chalgren agar plates with 7% (v/v) defibrinated horse blood and *H. pylori* selective antibiotic supplement (Oxoid, Basingstoke, UK) as previously described. DNA was extracted from a single colony grown from an antrum or fundus biopsy using a QIAmp™ kit (Qiagen, Courtabouef, France). Polymorphism of *vacA* regions s, m, and i, and *cagA* status, were determined as described elsewhere (Breurec et al., *in press*; Panayotopoulou et al., 2007; Panayotopoulou et al., 2010).

To amplify the variable 3' region of the *cagA* gene, we designed a new forward primer after initial alignment of Genbank *cagA* sequences of *H. pylori* isolates from Asian patients: primer 2551*cagA*-F (5'- GGAACCCTAGTCGGTAAAYGGG -3', where Y = T or C) was designed to recognize positions 582473 to 582491 of the *H. pylori* 26695 genome. The reverse primer 3189*cagA*-R (5'- GCGTATGTGGCTGTTAGTAGCG -3') was previously described by Uchida et al (Uchida et al., 2009). PCR amplification was carried out in 50- $\mu$ l reaction mixtures containing 2.5  $\mu$ l of 10  $\times$  PCR buffer (Eurobio), 1.5 U of Taq (Eurobluetaq, Eurobio), 200  $\mu$ M deoxynucleoside triphosphates, 0.5  $\mu$ M each primer, and 50 ng of microbial genomic DNA. The PCR conditions included an initial denaturation step at 94 °C for 5 min, followed by 35 cycles at 94 °C for 30 s, 60 °C for 45 s, and 72 °C for 45 s, and then an extension step at 72 °C for 5 min. PCR products were analyzed on 1.5% agarose gel stained with ethidium bromide. Amplified DNA fragments were purified with Microspin S-400 HR columns (Amersham Pharmacia Biotech, Inc., Uppsala, Sweden), and direct sequencing was performed with the ABI PRISM BigDye Terminator v3.0 cycle sequencing kit (PE Applied Biosystems, Foster City, Calif.) on an ABI 3700 DNA sequencer (PE Applied Biosystems). The number and type of EPIYA segments were inferred as previously described (Panayotopoulou et al., 2007).

#### 2.4. Sequencing and analysis of the complete *cagA* gene

The full-length *cagA* gene was amplified by PCR as previously described (Yamazaki et al., 2005). PCR products were analyzed on 1% agarose gel stained with ethidium bromide. Amplified DNA fragments were purified with Microspin S-400 HR columns (Amersham Pharmacia Biotech, Inc., Uppsala, Sweden), and direct sequencing was performed with the ABI PRISM BigDye Terminator v3.0 cycle sequencing kit (PE Applied Biosystems, Foster City, Calif.) on an ABI 3700 DNA sequencer (PE Applied Biosystems). The full *cagA* gene was sequenced by primer-walking using the primers listed in Table 1.

Full-length sequences of the *cagA* gene were aligned using the ClustalW program (Larkin et al., 2007). The neighbor-joining tree was obtained with MEGA v4 (Tamura et al., 2007), using a distance matrix calculated with the Kimura 2 parameter model of sequence evolution. *cagA* population assignment was based on the approach described by Falush et al (Falush et al., 2003b) using the “no admixture” model of STRUCTURE (Falush et al., 2003a). The linkage model in STRUCTURE was used to estimate the proportion of nucleotides derived from each ancestral population (Falush et al., 2003b; Linz et al., 2007).

#### 2.5. GenBank accession numbers

The nucleotide sequences have been deposited in Genbank under accession numbers JF798688–JF798748.

#### 2.6. Statistical analysis

Microsoft Access 2003 was used for data entry, and Stata Version 10 (StataCorp, College Station, TX, USA) for statistical analysis. A Fisher's exact test was used to identify determinants of ulcerated lesions and GC. *P*-values <0.05 were considered to denote a significant association.

### 3. Results and discussion

#### 3.1. *H. pylori* isolates from Cambodians, Cambodia

We studied 66 *H. pylori* isolates recovered from gastric biopsies of 66 different patients (36 males, 30 females, mean 44.7 years,

median 46.0 years, range 18–76 years). Based on endoscopic findings, 60 patients had gastritis only, five had ulcerated lesions, and one signs of neoplasia. GC was histologically confirmed in the latter patient. The number of cases was too small to identify risk factors for severe gastroduodenal disease (data not shown).

The genetic background of the 66 *H. pylori* isolates studied here has been described elsewhere based on sequences of seven house-keeping genes (Breurec et al., 2011a). These isolates belonged to the genetic populations hpEurope (*n* = 34, 51.5%), and hpEastAsia, subpopulation hspEAsia (*n* = 32, 48.5%).

As *H. pylori* is a bacterial genetic marker of human migrations, we provided evidence that at least three ancient human migrations introduced distinct genetic populations carrying specific virulence factors into Southeast Asia: (i) a migration of Austro-Asiatic speaking people about 4000 years ago from China into Vietnam and Cambodia carrying hspEAsia bacteria, (ii) a migration from India introducing hpEurope bacteria into Thailand, Cambodia and Malaysia within the last 2000 years. (iii) a migration of the ancestors of the Thai people from Southern China into Thailand carrying *H. pylori* of population hpAsia2 during the early second millennium AD (Breurec et al., 2011a). The distribution of *H. pylori* populations in East and Southeast Asia is closely related to the distribution of gastric cancer (Falush et al., 2003b; Ferlay et al., 2010; Linz et al., 2007; Tay et al., 2009). Countries where hspEAsia predominates (Japan, China, Korea, and Vietnam) exhibit a high incidence of GC, while the incidence is low in countries where hpAsia2 predominates, such as Thailand. Cambodia, which displays a mixture of hpEurope and hspEAsia strains, is classified as a country with an intermediate risk of GC. Then, the genetic population might be a marker for virulence factors directly involved in the outcome of *H. pylori* infection.

#### 3.2. *cagA* status, EPIYA diversity, and gastric cancer incidence

The *cagA* gene was present in 61 (92.4%) of the 66 isolates, and was significantly more frequent in hspEAsia strains (100%) than in hpEurope strains (85.3%), as previously described (Olbermann et al., 2010). The empty site was successfully amplified in the 5 *cag* PAI-negative strains with primers derived from the *cag* PAI flanking regions. The EPIYA diversity of the 61 *cagA* genes was analysed, revealing the presence of all four EPIYA motifs, namely

**Table 1**  
List of oligonucleotides primers used for the full *cagA* sequencing.

Primer	Primer sequence (5'–3')	Corresponding DNA sequence <sup>a</sup>	Reference
<i>cagA</i> L2 (+)	AAGGAGAAACAATGACTAACGAAACTATTG	–11–19	40
<i>cagA</i> L2 (–)	TCCTTTAAGATTTTGGAAACACCTTTTG	+4–3536	40
<i>cagA</i> A1-F (+)	CTGCCAARCAATCTTTTGCAGG	499–520	This study
	R = A/G		
<i>cagA</i> A1-R (–)	TCTTTGASATCRGAMGATTG	691–672	This study
	S = C/G, R = A/G, M = C/A		
<i>cagA</i> 2-F (+)	CCGTGGTTATAAARACCAACAAGG	1015–1039	This study
	R = A/G		
<i>cagA</i> 4-R (–)	TGCTTCYAAATGRGAAACGCCATT	1604–1581	This study
	Y = T/C, R = A/G		
<i>cagA</i> 3-F (+)	TYACTGAGTTTGGTAATGGGGA	1387–1408	This study
	Y = T/C		
<i>cagA</i> 12-R (–)	GGAACCTAGTCGGTAAYGGTT	2572–2552	This study
	Y = T/C		
<i>cagA</i> 4-F (+)	GCRAAAKCTCAAGCTAACAGC	1947–1964	This study
	K = G/T		
3189 <i>cagA</i> -R (–)	GCGTGTGTGGCTGTTAATAGC	3209–3189	This study
2551 <i>cagA</i> -F (+)	GGAACCTAGTCGGTAAYGGG	2550–2571	This study
	Y = T/C		
<i>cagA</i> A11-F (+)	GTARRGTAGGGCTTTCAAGG	2878–2894	This study
	R = A/G		

<sup>a</sup> Nucleotide positions in the HP0547 (*cagA*) gene of *H. pylori* 26695 (GenBank accession no. AE000569).

EPIYA-A (EPIYAQVVK), EPIYA-B (EPIYAQVAK), EPIYA-C (EPIYA-TIDD/E), and EPIYA-D (EPIYATIDF). The predominant EPIYA combinations in our strains were ABC ( $n = 30$ , 45.5%) and ABD ( $n = 23$ , 34.8%). None of the strains with an EPIYA-D segment exhibited duplication, in keeping with a previous report (Argent et al., 2008a). The remaining strains were AB ( $n = 1$ ), ABBD ( $n = 1$ ), BC ( $n = 2$ ), and ABCC ( $n = 4$ ) (Table 2).

CagA protein EPIYA segment diversity has been described as a major determinant of *H. pylori* pathogenicity. SHP-2 specifically binds to the tyrosine-phosphorylated EPIYA-C or EPIYA-D segment of CagA, thus stimulating SHP-2 activity. This type of deregulation has been implicated in a variety of human malignancies (Hatakeyama, 2009). In contrast to the EPIYA-C segment, the sequence flanking the tyrosine phosphorylation site of the EPIYA-D segment perfectly matches the consensus high-affinity binding sequence for the SH2 domains of SHP2. As a result, the EPIYA-D segment (East-Asian CagA) interacts more strongly with SHP-2 than does the EPIYA-C segment (Western CagA), and is more closely associated with precancerous lesions and GC (Argent et al., 2008a). Furthermore, East-Asian CagA has been described as more carcinogenic than Western CagA in epidemiological studies (Jones et al., 2009), and also in transgenic mice, although malignancies occurred in these animals' non gastric organs (Miura et al., 2009).

The clinical outcome of *H. pylori* infection depends on host, bacterial and environmental factors, and geographic differences in the incidence of GC cannot be explained solely by the distribution of Western and East-Asian CagA. In East and South East Asia, where human genetic diversity is relatively limited, the incidence of GC clearly matches the prevalence rates of East-Asian CagA (Yamaoka, 2010), albeit with some inconsistencies. As the genetic structure of the modern human populations of Vietnam and Cambodia are similar (Hill et al., 2007), these countries are suitable for studying the influence of the CagA type on the incidence of GC. The GC ASR in

Cambodia is half that in Vietnam (9.8 versus 18.9 per 100,000), where 97% of isolates harbor the East-Asian CagA (Uchida et al., 2009). The ASR is low in Thailand and Malaysia compared with East Asian countries, but the large Chinese communities in these countries appear to have a significant independent risk factor for GC (Goh et al., 2007; Vilaichone et al., 2004). It has not previously been possible to establish the role of the East-Asian CagA as a determinant of GC, owing to the high frequency of this CagA type among patients of Chinese ethnicity and the low incidence of GC in other ethnic groups (Schmidt et al., 2010; Vilaichone et al., 2004). The overall incidence of GC is high in Japan (ASR 31.1 per 100,000), but a far lower rate is observed in Okinawa, a group of islands in southern Japan, with a strong American influence. This low incidence might be explained by the higher frequency of Western CagA (42%) in *H. pylori* strains in Okinawa (Satomi et al., 2006) than in other parts of Japan (Yamaoka, 2010). In Korea, where East-Asian CagA *H. pylori* strains predominate (89%), as in Vietnam (Jones et al., 2009), the incidence of GC is twice that observed in Vietnam, showing that the East-Asian CagA is not the only major determinant of GC (Jones et al., 2009). All these data suggest that endemic circulation of *H. pylori* strains harboring the East-Asian CagA contributes substantially to the high incidence of GC in East Asian countries. Thus, the low prevalence of East-Asian CagA in Cambodia and, by extension, in Southeast Asia (except for Vietnam), might partly explain the lower incidence of GC than in other East Asian countries.

### 3.3. *vacA* gene diversity and gastric cancer incidence rates

*vacA* typing showed that 63 (95.5%) of the 66 strains were s1, 32 (48.5%) were m1 and 62 (93.9%) were i1 (Table 2). All s1/m1 strains were type i1, whereas all s2/m2 strains were i2. The s2/m2/i2 strains were all *cagA*-negative. Most strains possessed presumed high vacuolation isotypes, namely s1/i1/m1 ( $n = 32$ , 48.5%) or s1/i1/m2 ( $n = 30$ , 45.4%). The s1 allele has been identified as a determinant of GC in Western countries (Yamaoka, 2010), and i1-type strains are strongly associated with gastric adenocarcinoma in Iranian patients (Rhead et al., 2007). As the majority of our strains harbored s1 and i1 alleles, similar to other countries in Southeast and East Asia (Yamaoka, 2010), these alleles seem not appear to be major risk factors for GC in Cambodia.

The decreasing gradient of m1-type strains observed from South Korea (95%) to Vietnam (37%) has been shown to match the GC incidence gradient (Yamaoka, 2010). In Vietnam, the GC incidence is higher in the North (Hanoi; ASR 27.0 per 100,000 males) than in the South (Ho Chi Minh, 18.7 per 100,000 males), matching the frequency distribution of m1-type strains (50% in Hanoi vs 25% in Ho Chi Minh) (Uchida et al., 2009). As most strains in this region are assigned to the hspEAsia subpopulation, the m region of *vacA* might be a virulence marker in this subpopulation. Indeed, this correlation is not found in other countries in South East Asia (Chomvarin et al., 2008; Linpisarn et al., 2007; Tan et al., 2005) which display a high frequency of hpAsia2 and hpEurope strains (Linz et al., 2007; Tay et al., 2009). In Cambodia, the frequency of m1-type strains among hspEAsia strains (59.4%) is similar to that reported in Northern Vietnam, suggesting that those hspEAsia strains are less virulent than strains from Japan and Korea.

### 3.4. Expansion of European *vacA* and *cagA* alleles among hspEAsia strains

Interstrain recombination is critically dependent on the frequent infection with multiple strains, which is common in developing countries (Ghose et al., 2005). In addition, recombination events during mixed infection have been identified as a major driving force behind allelic diversity in *H. pylori*. The phylogeny of the

**Table 2**  
*cagA* status, diversity of the EPIYA motifs, and the *vacA* s, m, and i region polymorphisms in hpEurope and hspEAsia strains from Cambodia.

	hpEurope ( $n = 34$ )	hspEAsia ( $n = 32$ )	Total ( $n = 66$ )
<b><i>cagA</i></b>			
Present	29 (85.3)	32 (100.0)	61 (92.4)
Absent	5 (14.7)	0	5 (7.6)
<b>EPIYA motif</b>			
AB	1 (2.9)	0	1 (1.5)
BC	2 (5.9)	0	2 (3.0)
ABC	23 (67.7)	7 (21.9)	30(45.5)
ABCC	2 (5.9)	2 (6.2)	4 (6.1)
ABBD	0	1 (3.1)	1 (1.5)
ABD	1 (2.9)	22 (68.8)	23 (34.8)
<b><i>vacA</i> s</b>			
s1	31 (91.2)	32 (100.0)	63 (95.5)
	s1a	11 (34.4)	38 (57.6)
	s1c	21 (65.6)	25 (37.9)
s2	3 (8.8)	0	3 (4.5)
<b><i>vacA</i> m</b>			
m1	13 (38.2)	19 (59.4)	32 (48.5)
m2	21 (61.8)	13 (40.6)	34 (51.5)
<b><i>vacA</i> i</b>			
i1	30 (88.2)	32 (100.0)	62 (93.9)
i2	4 (11.8)	0	4 (6.1)
<b><i>vacA</i> smi</b>			
s1a/i1/m1	13 (38.2)	8 (25.0)	21 (31.8)
s1a/i1/m2	13 (38.2)	3 (9.4)	16 (24.2)
s1a/i2/m2	1 (3.0)	0	1 (1.5)
s1c/i1/m1	0	11 (34.4)	11 (16.7)
s1c/i1/m2	4 (11.8)	10 (31.2)	14 (21.2)
s2/i2/m2	3 (8.8)	0	3 (4.6)

*cag* PAI largely reflects that of *H. pylori*'s housekeeping genes, with only a limited number of genes being under diversifying selection (Olbermann et al., 2010). Accordingly, hpEurope and hspEAsia strains are expected to carry a Western and an East-Asian CagA, respectively. Although we did not find CagA sequences containing both segment C and segment D, we observed that recombination involving the *cagA* gene is not a rare event, as one out of 34 (2.9%) Cambodian hpEurope strains possessed a CagA with an East Asian EPIYA motif (ABD) while 9/32 (28.1%) hspEAsia strains contained a Western CagA with the EPIYA motifs ABC or ABCC (Table 2).

In order to study the population genetic structure of the *cagA* gene in *H. pylori* isolates from Cambodian patients, we randomly selected 10 hpEurope and 10 hspEAsia *cagA*-positive strains for complete *cagA* sequencing (Table 3). Five (4 hspEAsia, 1 hpEurope) of the 10 East-Asian *cagA* genes in Cambodian strains had the 18 bp deletion in the *cagA* gene that was previously described for Vietnamese strains (Truong et al., 2009), while the remaining 5 hspEAsia isolates possessed the 39 bp deletion common in other East Asian strains (Uchida et al., 2009). Both deletions are located ca. 300 bp upstream of the EPIYA-A segment. Further studies are needed to determine the oncogenic influence of these subtypes.

The 20 unique *cagA* haplotypes were compared to 35 unique haplotypes of *cagA* sequences from a globally representative and well characterized strain collection from all bacterial genetic populations possessing the *cagPAI* (hpAfrica1, hpNEAfrica, hpEurope, hpSahul, hpAsia2, hspEAsia, hspMaori) (Olbermann et al., 2010). A Neighbor-joining tree constructed from the *cagA* sequences displayed two strain clusters that corresponded to the Western and East-Asian CagA (Fig. 1), with 100% assignment of EPIYA-C motif to the Western CagA (hpAfrica1, hpNEAfrica, hpEurope, some hpAsia2 strains) and of the EPIYA-D motif to East Asian CagA (hspEAsia, hspMaori, some hpAsia2 strains). We used the linkage model of Structure (Falush et al., 2003a) to estimate the proportion of nucleotides derived from European and East Asian ancestral populations in the individual *cagA* sequences. While 18/20 *cagA* genes from Cambodian strains have derived 100% of their nucleotides from either European or East Asian ancestral sources, two strains dis-

played extensive recombination within the *cagA* gene (Table 3). One strain was of East Asian ancestry up to nucleotide 91, and of European ancestry thereafter, whereas the second strain was of European ancestry up to nucleotide 147, of East Asian ancestry from nucleotide 148 to nucleotide 370, followed by European ancestry from nucleotide 371 to nucleotide 1159, and of East Asian ancestry thereafter. Based on housekeeping gene sequences, both strains were assigned to hspEAsia, suggesting subversion of an East-Asian *cagA* gene by transformation with DNA from a Western (hpEurope) *cagA*. These findings, together with the relatively high proportion (28.1%) of Western CagA in hspEAsia strains, in contrast to only 1/34 EastAsian CagA in hpEurope strains, provides evidence for the spread of Western CagA among East Asian strains in Cambodia. This is of major importance as CagA interacts, through individual domains, with a large number of host intracellular proteins such as SH2 proteins (SHP-2, Csk, Grb2, Crk, PI3K, SHP-1, Ras-GAP1 and Grb7) and protein kinases (Src, Abl, MARK2/PAR1b kinase family) (Selbach et al., 2009). Thus, the recombinant CagA protein probably reflects functional adaptation or modulation in response to host protein polymorphism (Olbermann et al., 2010).

Fine typing of *vacA* s1 alleles revealed the presence of both s1a (38 strains, 57.6%) and s1c (25 strains, 37.9%) alleles among the Cambodian strains. The majority of the Cambodian hpEurope strains (27/34, 79.4%) possessed the s1a allele (Table 2) which is found frequently among European (hpEurope) strains (Van Doorn et al., 1999), while the remaining strains were s1c (4/34, 11.8%) or s2 (3/34, 8.8%). The s1c allele was previously shown to be associated with East Asian *H. pylori* strains (Yamaoka et al., 2002) as 449/474 (94.7%) strains from East Asia were described as s1c and only 4.7% as s1a. However, the rate of s1c alleles among East Asian (hspEAsia) strains from Cambodia was much lower (21/32 strains, 65.6%) while the other 34.4% (11/32 strains) were typed as s1a. The observed high prevalence of s1a alleles among Cambodian hspEAsia strains indicates frequent introgression of European *vacA* gene alleles into East Asian *H. pylori* strains, which is similar to the expansion of European *cagA* genes.

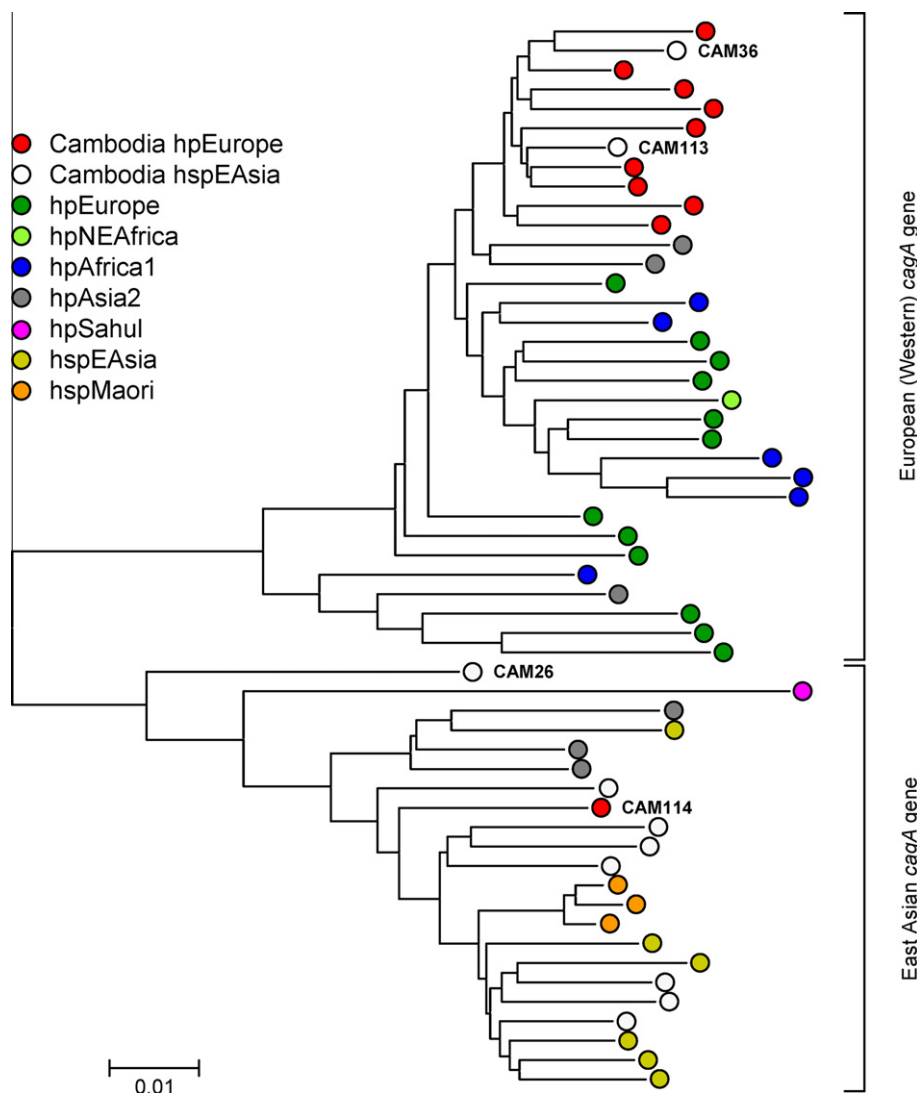
The simultaneous expansion of both European *cagA* and *vacA* alleles in seven of the 32 hspEAsia strains (21.9%) might be associ-

**Table 3**  
Genetic features of the 20 strains investigated by complete *cagA* sequencing.

Strains	Structure no linkage model		Structure linkage model	EPIYA motif	Size of <i>cagA</i> gene (bp)	Deletion <sup>a</sup> (bp)	<i>vacA</i> type
	(sub) population	CagA type					
CAM22	hpEurope	Western	European ancestry	ABC	3540	None	s1a/i1/m1
CAM23	hpEurope	Western	European ancestry	ABC	3543	None	s1a/i1/m1
CAM40	hpEurope	Western	European ancestry	ABC	3555	None	s1a/i1/m1
CAM27	hpEurope	Western	European ancestry	ABC	3543	None	s1a/i1/m1
CAM46	hpEurope	Western	European ancestry	ABC	3543	None	s1a/i1/m1
CAM25	hpEurope	Western	European ancestry	ABCC	3642	None	s1a/i1/m1
CAM114	hpEurope	East Asian	East Asian ancestry	ABD	3552	18	s1a/i1/m1
CAM17	hpEurope	Western	European ancestry	ABC	3480	None	s1a/i1/m2
CAM14	hpEurope	Western	European ancestry	ABC	3543	None	s1c/i1/m2
CAM15	hpEurope	Western	European ancestry	ABC	3537	None	s1c/i1/m2
CAM113	hspEAsia	Western	European ancestry	ABC	3549	None	s1a/i1/m1
CAM36	hspEAsia	Western	nt <sup>b</sup> 1 to 91 East Asian ancestry nt 92 to 3558 European ancestry	ABC	3558	None	s1a/i1/m1
CAM16	hspEAsia	East Asian	East Asian ancestry	ABD	3546	18	s1a/i1/m2
CAM29	hspEAsia	East Asian	East Asian ancestry	ABD	3492	39	s1c/i1/m1
CAM47	hspEAsia	East Asian	East Asian ancestry	ABD	3528	39	s1c/i1/m1
CAM32	hspEAsia	East Asian	East Asian ancestry	ABD	3531	18	s1c/i1/m1
CAM26	hspEAsia	East Asian	nt 1 to 147 European ancestry nt 148 to 37 East Asian ancestry nt 371 to 1159 European ancestry nt 1160 to 3537 East Asian ancestry	ABD	3537	18	s1c/i1/m1
CAM112	hspEAsia	East Asian	East Asian ancestry	ABD	3549	18	s1c/i1/m2
CAM33	hspEAsia	East Asian	East Asian ancestry	ABD	3531	39	s1c/i1/m2
CAM37	hspEAsia	East Asian	East Asian ancestry	ABD	3531	39	s1c/i1/m2

<sup>a</sup> Deletion located ca. 300 bp upstream of the EPIYA-A segment.

<sup>b</sup> nt, nucleotide.



**Fig. 1.** Neighbor-joining (NJ) tree of 55 *cagA* sequences color-coded according to population assignment of *H. pylori* housekeeping genes by Structure into the populations hpEurope, hpNEAfrica, hpAfrica1, hpAsia2, hpSahul and hpEastAsia (subpopulations hspMaori and hspEAsia). The NJ tree displays the phylogenetic relationship of 20 *cagA* sequences from Cambodia and 35 *cagA* sequences that are representative for the worldwide *cagA* diversity. Strains that showed introgression of European *cagA* alleles into East Asian strains (CAM26, CAM36, CAM113) or import of East Asian *cagA* into a hpEurope strain (CAM114) are indicated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ated with fitness advantage in the human host population and/or reflect selection for a stable equilibrium of the effects of CagA and VacA on epithelial cells. Indeed, VacA and CagA compete for multiple cell signalling pathways (Argent et al., 2008b; Tegtmeyer et al., 2009; Yokoyama et al., 2005) and deregulation of any of the pathways by either CagA or VacA is likely to contribute to the various disease outcome. Thus, changes in the genomic composition of *H. pylori* strains by transformation of DNA from unrelated strains may disrupt the balancing effects of VacA and CagA activities on epithelial cells which might result in severe consequences for host response and GC risk. Therefore, the observed range expansion of European *vacA* and *cagA* alleles among hspEAsia strains may have a significant impact on the future incidence of GC in Cambodia.

#### 4. Conclusions

These findings are particularly important, especially given the difficulty of carrying out such studies in countries with inadequate healthcare systems. Even though our findings may not be repre-

sentative of the overall situation, it should be noted that the study took place in one of the main tertiary hospitals in Cambodia with a large geographical recruitment of patients. Our isolates were characterized by a lower frequency of hspEAsia strains carrying the East-Asian CagA and the m1-type *vacA* allele (10 out of 66, 15.2%) than reported in East Asian countries, which may partly explain the low incidence of GC in Cambodia and, by extension, in Southeast Asia. We observed the range expansion of European *vacA* and *cagA* alleles (hpEurope strains) among hspEAsia strains which may have a significant impact on the future incidence of GC in Cambodia. Further investigations are needed to confirm our findings based on large cohort of patients in order to take account simultaneously the different factors potentially involved in gastric carcinogenesis: *H. pylori* infection rate and host, environmental and *H. pylori* virulence factors.

#### Acknowledgements

We thank all the patients and clinicians who participated in this study. We thank also Fatou Bintou Dieye (Institut Pasteur, Dakar,

Senegal) for her technical help. This work was supported by grants from the Pasteur Institute (ACIP 2003 and 2006) and the ERA-NET PathoGenoMics (Project HELDIVNET).

## References

- Argent, R.H., Hale, J.L., El-Omar, E.M., Atherton, J.C., 2008a. Differences in *Helicobacter pylori* CagA tyrosine phosphorylation motif patterns between western and East Asian strains, and influences on interleukin-8 secretion. *J. Med. Microbiol.* 57, 1062–1067.
- Argent, R.H., Thomas, R.J., Letley, D.P., Rittig, M.G., Hardie, K.R., Atherton, J.C., 2008b. Functional association between the *Helicobacter pylori* virulence factors VacA and CagA. *J. Med. Microbiol.* 57, 145–150.
- Breurec, S., Guillard, B., Hem, S., Brisse, S., Dieye, F.B., Huerre, M., Oung, C., Tan, T., Thiberge, J., Raymond, J., Vong, S., Monchy, D.B.L., 2011a. Evolutionary history of *Helicobacter pylori* sequences reflect past human migrations in Southeast Asia. *PLoS One* 6, e22058.
- Breurec, S., Michel, R., Seck, A., Brisse, S., Côme, D., Dieye, F.B., Garin, B., Huerre, M., Mbengue, M., Fall, C., Sgouras, D.N., Thiberge, J.M., Dia, D., Raymond, J., in press. Clinical Relevance of *cagA* and *vacA* Gene Polymorphisms in *Helicobacter pylori* Isolates from Senegalese Patients. *Clin. Microbiol. Infect.*
- Chomvarin, C., Namwat, W., Chaicumpar, K., Mairiang, P., Sangchan, A., Sripan, B., Tor-Udom, S., Vilaichone, R.K., 2008. Prevalence of *Helicobacter pylori vacA, cagA, cagE, iceA* and *babA2* genotypes in Thai dyspeptic patients. *Int. J. Infect. Dis.* 12, 30–36.
- Falush, D., Stephens, M., Pritchard, J.K., 2003a. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587.
- Falush, D., Wirth, T., Linz, B., Pritchard, J.K., Stephens, M., Kidd, M., Blaser, M.J., Graham, D.Y., Vacher, S., Perez-Perez, G.I., Yamaoka, Y., Megraud, F., Otto, K., Reichard, U., Katzowitsch, E., Wang, X., Achtman, M., Suerbaum, S., 2003b. Traces of human migrations in *Helicobacter pylori* populations. *Science* 299, 1582–1585.
- Fenoglio-Preiser, C.M., Noffsinger, A.E., Lantz, P.E., 1999. *Gastrointestinal Pathology: an Atlas and Text*, second ed. Lippincott-Raven, Philadelphia.
- Ferlay, J., Shin, H., Bray, F., Forman, D., Mathers, C., Parkin, D., 2010. GLOBOCAN 2008, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10. International Agency for Research on Cancer, Lyon, France.
- Fock, K.M., Ang, T.L., 2010. Epidemiology of *Helicobacter pylori* infection and gastric cancer in Asia. *J. Gastroenterol. Hepatol.* 25, 479–486.
- Ghose, C., Perez-Perez, G.I., van Doorn, L.J., Dominguez-Bello, M.G., Blaser, M.J., 2005. High frequency of gastric colonization with multiple *Helicobacter pylori* strains in Venezuelan subjects. *J. Clin. Microbiol.* 43, 2635–2641.
- Goh, K.L., Cheah, P.L., Md, N., Quek, K.F., Parasakthi, N., 2007. Ethnicity and *H. pylori* as risk factors for gastric cancer in Malaysia: a prospective case control study. *Am. J. Gastroenterol.* 102, 40–45.
- Hatakeyama, M., 2009. *Helicobacter pylori* and gastric carcinogenesis. *J. Gastroenterol.* 44, 239–248.
- Hill, C., Soares, P., Mormina, M., Macaulay, V., Clarke, D., Blumbach, P.B., Vizuete-Forster, M., Forster, P., Bulbeck, D., Oppenheimer, S., Richards, M., 2007. A mitochondrial stratigraphy for island southeast Asia. *Am. J. Hum. Genet.* 80, 29–43.
- Isomoto, H., Moss, J., Hirayama, T., 2010. Pleiotropic actions of *Helicobacter pylori* vacuolating cytotoxin, VacA. *Tohoku J. Exp. Med.* 220, 3–14.
- Jones, K.R., Joo, Y.M., Jang, S., Yoo, Y.J., Lee, H.S., Chung, I.S., Olsen, C.H., Whitmire, J.M., Merrell, D.S., Cha, J.H., 2009. Polymorphism in the CagA EPIYA motif impacts development of gastric cancer. *J. Clin. Microbiol.* 47, 959–968.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Lewis, P., 2009. *Ethnologue: Languages of the World*, 16 ed. SIL International, Dallas.
- Linpisarn, S., Suwan, W., Lertprasertsuk, N., Koosirirat, C., Steger, H.F., Prommuangyong, K., Phornphutkul, K., 2007. *Helicobacter pylori cagA, vacA* and *iceA* genotypes in northern Thai patients with gastric disease. *Southeast Asian J. Trop. Med. Public Health* 38, 356–362.
- Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S.W., Yamaoka, Y., Graham, D.Y., Perez-Trallero, E., Wadstrom, T., Suerbaum, S., Achtman, M., 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445, 915–918.
- Miura, M., Ohnishi, N., Tanaka, S., Yanagiya, K., Hatakeyama, M., 2009. Differential oncogenic potential of geographically distinct *Helicobacter pylori* CagA isoforms in mice. *Int. J. Cancer.* 125, 2497–2504.
- Moodley, Y., Linz, B., Yamaoka, Y., Windsor, H.M., Breurec, S., Wu, J.Y., Maady, A., Bernhoff, S., Thiberge, J.M., Phuanukoonnon, S., Jobb, G., Siba, P., Graham, D.Y., Marshall, Achtman, M., . The peopling of the Pacific from a bacterial perspective. *Science* 323, 527–530.
- Olberrmann, P., Josenhans, C., Moodley, Y., Uhr, M., Stamer, C., Vauterin, M., Suerbaum, S., Achtman, M., Linz, B., 2010. A global overview of the genetic and functional diversity in the *Helicobacter pylori cag* pathogenicity island. *PLoS Genet.* 6, e1001069.
- Panayotopoulou, E.G., Sgouras, D.N., Papadakos, K., Kalliaropoulos, A., Papatheodoridis, G., Mentis, A.F., Archimandritis, A.J., 2007. Strategy to characterize the number and type of repeating EPIYA phosphorylation motifs in the carboxyl terminus of CagA protein in *Helicobacter pylori* clinical isolates. *J. Clin. Microbiol.* 45, 488–495.
- Panayotopoulou, E.G., Sgouras, D.N., Papadakos, K.S., Petraki, K., Breurec, S., Michopoulos, S., Mantzaris, G., Papatheodoridis, G., Mentis, A., Archimandritis, A., 2010. CagA and VacA polymorphisms are associated with distinct pathological features in *H. pylori*-infected adults with peptic and non-peptic ulcer disease. *J. Clin. Microbiol.* 48, 2237–2239.
- Parkin, D.M., 2004. International variation. *Oncogene* 23, 6329–6340.
- Rhead, J.L., Letley, D.P., Mohammadi, M., Hussein, N., Mohagheghi, M.A., Eshagh Hosseini, M., Atherton, J.C., 2007. A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology* 133, 926–936.
- Satomi, S., Yamakawa, A., Matsunaga, S., Masaki, R., Inagaki, T., Okuda, T., Suto, H., Ito, Y., Yamazaki, Y., Kuriyama, M., Keida, Y., Kutsumi, H., Azuma, T., 2006. Relationship between the diversity of the *cagA* gene of *Helicobacter pylori* and gastric cancer in Okinawa. *Jpn. J. Gastroenterol.* 41, 668–673.
- Schmidt, H.M., Andres, S., Nilsson, C., Kovach, Z., Kaakoush, N.O., Engstrand, L., Goh, K.L., Fock, K.M., Forman, D., Mitchell, H., 2010. The *cag* PAI is intact and functional but HP0521 varies significantly in *Helicobacter pylori* isolates from Malaysia and Singapore. *Eur. J. Clin. Microbiol. Infect. Dis.* 29, 439–451.
- Selbach, M., Paul, F.E., Brandt, S., Guye, P., Daumke, O., Backert, S., Dehio, C., Mann, M., 2009. Host cell interactome of tyrosine-phosphorylated bacterial proteins. *Cell Host Microbe* 5, 397–403.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599.
- Tan, H.J., Rizal, A.M., Rosmadi, M.Y., Goh, K.L., 2005. Distribution of *Helicobacter pylori cagA, cagE* and *vacA* in different ethnic groups in Kuala Lumpur, Malaysia. *J. Gastroenterol. Hepatol.* 20, 589–594.
- Tay, C.Y., Mitchell, H., Dong, Q., Goh, K.L., Dawes, I.W., Lan, R., 2009. Population structure of *Helicobacter pylori* among ethnic groups in Malaysia: recent acquisition of the bacterium by the Malay population. *BMC Microbiol.* 9, 126.
- Tegtmeier, N., Zabler, D., Schmidt, D., Hartig, R., Brandt, S., Backert, S., 2009. Importance of EGF receptor, HER2/Neu and Erk1/2 kinase signalling for host cell elongation and scattering induced by the *Helicobacter pylori* CagA protein: antagonistic effects of the vacuolating cytotoxin VacA. *Cell Microbiol.* 11, 488–505.
- Truong, B.X., Mai, V.T., Tanaka, H., Ly, T., Thong, T.M., Hai, H.H., Van Long, D., Furumatsu, K., Yoshida, M., Kutsumi, H., Azuma, T., 2009. Diverse characteristics of the CagA gene of *Helicobacter pylori* strains collected from patients from southern vietnam with gastric cancer and peptic ulcer. *J. Clin. Microbiol.* 47, 4021–4028.
- Uchida, T., Nguyen, L.T., Takayama, A., Okimoto, T., Kodama, M., Murakami, K., Matsuhisa, T., Trinh, T.D., Ta, L., Ho, D.Q., Hoang, H.H., Kishida, T., Fujioka, T., Moriyama, M., Yamaoka, Y., 2009. Analysis of virulence factors of *Helicobacter pylori* isolated from a Vietnamese population. *BMC Microbiol.* 9, 175.
- Van Doorn, L.J., Figueiredo, C., Megraud, F., Pena, S., Midolo, P., Queiroz, D.M., Carneiro, F., Vanderborgh, B., Pegado, M.D., Sanna, R., De Boer, W., Schneeberger, P.M., Correa, P., Ng, E.K., Atherton, J., Blaser, M.J., Quint, W.G., 1999. Geographic distribution of *vacA* allelic types of *Helicobacter pylori*. *Gastroenterology* 116, 823–830.
- Vilaichone, R.K., Mahachai, V., Tumwasorn, S., Wu, J.Y., Graham, D.Y., Yamaoka, Y., 2004. Molecular epidemiology and outcome of *Helicobacter pylori* infection in Thailand: a cultural cross roads. *Helicobacter* 9, 453–459.
- Yamaoka, Y., 2010. Mechanisms of disease: *Helicobacter pylori* virulence factors. *Nat. Rev. Gastroenterol. Hepatol.* 7, 629–641.
- Yamaoka, Y., Orito, E., Mizokami, M., Gutierrez, O., Saitou, N., Kodama, T., Osato, M.S., Kim, J.G., Ramirez, F.C., Mahachai, V., Graham, D.Y., 2002. *Helicobacter pylori* in North and South America before Columbus. *FEBS Lett.* 517, 180–184.
- 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. *J. Clin. Microbiol.* 43, 3906–3916.
- Yokoyama, K., Higashi, H., Ishikawa, S., Fujii, Y., Kondo, S., Kato, H., Azuma, T., Wada, A., Hirayama, T., Aburatani, H., Hatakeyama, M., 2005. Functional antagonism between *Helicobacter pylori* CagA and vacuolating toxin VacA in control of the NFAT signaling pathway in gastric epithelial cells. *Proc. Natl. Acad. Sci. USA* 102, 9661–9666.



**Article n°4: Clinical relevance of *cagA* and *vacA* gene polymorphisms in *Helicobacter pylori* isolates from Senegalese patients**

Breurec S, Michel R, Seck A, Brisse S, Côme D, Dieye FB, Garin B, Huerre M, Mbengue M, Fall C, Sgouras DN, Thiberge JM, Dia D, Raymond J.

Clinical Microbiology and Infection 2011 (article sous presse).

## **Contexte**

Très peu de données concernant les facteurs de virulence de *H. pylori* existent en Afrique. La seule étude incluant un nombre suffisant de CG concernait des souches d'Afrique du Sud. L'objet de ce travail était l'étude de l'association potentielle entre les polymorphismes des facteurs de virulence *cagA* et *vacA* et la pathologie gastro-duodénale chez des patients d'origine sénégalaise.

## **Matériels et méthodes**

Entre janvier 2007 et décembre 2008, 187 patients présentant des symptômes évocateurs d'une pathologie digestive haute consultant au CHU de l'Hôpital Le Dantec à Dakar, ont été inclus. Des biopsies gastriques ont été prélevées lors d'une endoscopie gastrique à partir desquelles ont été réalisés des examens histologiques et bactériologiques. Cent dix-sept patients (âge moyen 44.3 ans, intervalle interquartile 31 à 56.6 ans, intervalle 14 à 93 ans) présentaient une culture positive à *H. pylori*. L'ADN génomique a été extrait à partir d'un mélange de colonies de l'antrum et du fundus de façon à tenir compte des infections mixtes. Le polymorphisme du gène *vacA* et le statut *cagA* ont été étudiés par PCR, la région 3' du gène *cagA* a été séquencée et les motifs EPIYA ont été identifiés. La polyclonalité des souches isolées chez un même patient a été recherchée en utilisant une approche se basant sur l'hétérogénéité des gènes *cagA* et *vacA*. Les patients infectés par des souches possédant plusieurs allèles *vacA* différents ont été considérés comme présentant une infection mixte. Pour ceux présentant un allèle *vacA* unique mais des ambiguïtés après séquençage de l'extrémité 3' du gène *cagA*, les biopsies gastriques ont été décongelées et 10 colonies de l'antrum et du fundus ont été choisies pour subculture, puis extraction de l'ADN. L'existence de relations clonales entre isolats a été recherchée par RAPD. Une analyse univariée puis une analyse multivariée au moyen du modèle de régression logistique ont été utilisées pour identifier les facteurs de risque de CG et de lésions ulcérées parmi les patients présentant une infection à *H. pylori*. Les facteurs associés en analyse univariée avec une valeur  $p < 0.2$  ont été retenus dans l'analyse multivariée. La performance du modèle a été évaluée par le test d'adéquation selon Hosmer-Lemeshow et sa valeur discriminante a été appréciée par la surface sous la courbe ROC, la sensibilité et la spécificité. Une valeur  $p < 0.05$  a été considérée comme une différence significative.

## Résultats

Quatorze pour cent (17/117) des patients présentaient une infection mixte, parmi lesquels 5 patients ont été inclus dans la population d'étude finale puisque les 2 clones qu'ils hébergeaient possédaient un allèle *vacA* et une extrémité 3' du gène *cagA* identique (Figure 1, p3 de l'article). Au total, 105 patients ont été inclus pour lesquels le diagnostic endoscopique était le suivant: une gastrite seule chez 29 patients, des lésions ulcéreuses chez 61 patients et une suspicion de CG chez 15 patients. Toutes les suspicions de CG ont été histologiquement confirmées. Le gène *cagA* était présent chez 73.3 % (77/105) des isolats. Les protéines CagA contenaient zéro (3.7 %), un (93.9 %) et deux (2.4 %) segments EPIYA-C et toutes étaient de type « Western CagA ». La plupart des souches possédaient un allèle *vacA* décrit comme cytotoxique (*s1/i1/m1* (57.1 %) ou *s1/i1/m2* (21.9 %)). L'analyse univariée ne montrait pas d'association significative entre les lésions ulcérées et le statut *cagA*, le génotype *vacA* et le nombre de motifs EPIYA-C ( $p > 0.2$ ). En revanche, elle montrait une association significative entre la survenue d'un CG et un âge avancé ( $p = 0.001$ ), la présence de deux motifs EPIYA-C ( $p = 0.03$ ) et l'allèle *vacA* de type *s1* ( $p = 0.002$ ). Les deux derniers facteurs n'ont pu être inclus dans l'analyse multivariée puisque toutes les souches associées au CG sont de type *vacA s1* et que les 2 souches contenant 2 motifs EPIYA-C sont présentes parmi les CG. Le résultat de l'analyse univariée a donc été considéré pour ces 2 facteurs. Après analyse multivariée, un âge avancé ( $p < 0.001$ ) et la présence du gène *cagA* ( $p = 0.03$ ) étaient significativement associés à la survenue d'un CG. Le test de Hosmer-Lemeshow montre une bonne adéquation entre les cas de CG estimés et observés ( $p = 0,37$ ). L'aire sous la courbe de 0.87 prouve le bon pouvoir discriminant de notre modèle, la sensibilité était de 80 % et la spécificité de 83.3 %.

## Conclusion

Les données recueillies sont particulièrement importantes en raison de la difficulté à mener de telles études dans des pays avec un réseau sanitaire de qualité insuffisante. L'absence de biopsies gastriques systématiques explique la surreprésentation des lésions ulcérées et des CG dans notre population d'étude. Bien que nos résultats puissent ne pas être représentatifs de la situation globale au Sénégal, l'étude a été mise en place dans un des plus grands hôpitaux du pays (CHU Le Dantec). Les infections mixtes sont fréquentes chez les patients sénégalais, mettant en évidence l'importance des recombinaisons avec de l'ADN exogène qui ont été identifiées comme la principale force générant de la diversité génétique chez *H. pylori*. La conséquence peut être une modification de la pathogénicité de la souche. Malgré le faible nombre de cas, le CG était associé à un âge avancé, à la présence du gène

*cagA*, à la présence de deux motifs EPIYA-C et à l'allèle *vacA* de type s1, indiquant que la relation est robuste. La faible proportion de souches avec plusieurs motifs EPIYA-C par rapport aux autres continents (11.5 à 30 %) pourrait contribuer à la faible incidence du CG au Sénégal. Des études ultérieures au Sénégal en incluant un plus grand nombre de souches associées à un CG, mais également dans d'autres pays africains, seront nécessaires pour confirmer cette hypothèse.

# Clinical relevance of *cagA* and *vacA* gene polymorphisms in *Helicobacter pylori* isolates from Senegalese patients

S. Breurec<sup>1</sup>, R. Michel<sup>2</sup>, A. Seck<sup>1</sup>, S. Brisse<sup>3</sup>, D. Côme<sup>4</sup>, F. B. Dieye<sup>1</sup>, B. Garin<sup>1</sup>, M. Huerre<sup>4</sup>, M. Mbengue<sup>5</sup>, C. Fall<sup>1</sup>, D. N. Sgouras<sup>6</sup>, J. M. Thiberge<sup>3</sup>, D. Dia<sup>5</sup> and J. Raymond<sup>7,8</sup>

1) Institut Pasteur, Unité de Biologie Médicale et Environnementale, 2) Institut Pasteur, Unité d'Epidémiologie, Dakar, Senegal, 3) Institut Pasteur, Plateforme Génomique des Pathogènes et Santé Publique, 4) Institut Pasteur, Unité de Recherche et d'Expertises en Histotechnologie et Pathologie, Paris, France, 5) Centre Hospitalier Le Dantec, Département de Gastro-entérologie, Dakar, Senegal, 6) Hellenic Pasteur Institute, Laboratory of Medical Microbiology, Athens, Greece, 7) Institut Pasteur, Unité Postulante Pathogénèse de *Helicobacter* and 8) Université Paris Descartes, Faculté de Médecine, Paris, France

## Abstract

The molecular epidemiology of *Helicobacter pylori* in Africa is poorly documented. From January 2007 to December 2008, we investigated 187 patients with gastric symptoms in one of the main tertiary hospitals in Dakar, Senegal. One hundred and seventeen patients were culture-positive for *H. pylori*. Polymorphisms in *vacA* and *cagA* status were investigated by PCR; the 3'-region of *cagA* was sequenced, and EPIYA motifs were identified. Bacterial heterogeneity within individuals was extensively assessed by using an approach based on *vacA* and *cagA* heterogeneity. Fourteen per cent of *H. pylori*-positive patients displayed evidence of mixed infection, which may affect disease outcome. Patients with multiple *vacA* alleles were excluded from subsequent analyses. Among the final study population of 105 patients, 29 had gastritis only, 61 had ulcerated lesions, and 15 had suspicion of neoplasia based on endoscopic findings. All cases of suspected neoplasia were histologically confirmed as gastric cancer (GC). The *cagA* gene was present in 73.3% of isolates. CagA proteins contained zero (3.7%), one (93.9%) or two (2.4%) EPIYA-C segments, and all were western CagA. Most of the isolates possessed presumed high-vacuolization isotypes (slilm1 (57.1%) or slilm2 (21.9%)). Despite the small number of cases, GC was associated with *cagA* ( $p$  0.03), two EPIYA-C segments in the C-terminal region of CagA ( $p$  0.03), and the s1 *vacA* allele ( $p$  0.002). Multiple EPIYA-C segments were less frequent than reported in other countries, possibly contributing to the low incidence of GC in Senegal.

**Keywords:** Africa, *cagA*, EPIYA, *H. pylori*, mixed infections, Senegal, *vacA*

**Original Submission:** 2 June 2010; **Revised Submission:** 20 February 2011; **Accepted:** 7 March 2011

Editor: F. Megraud

*Clin Microbiol Infect*

**Corresponding author:** S. Breurec, Unité de Biologie Médicale et Environnementale, Institut Pasteur, 36 avenue Pasteur, BP 220, Dakar, Senegal  
**E-mail:** sbreurec@pasteur.sn

## Introduction

*Helicobacter pylori* is associated with severe gastroduodenal disorders, including peptic ulcer disease [1], distal gastric adenocarcinoma, and gastric mucosa-associated lymphoid tissue lymphoma [2]. Although about half of the world population is infected, only a small percentage of colonized individuals develop such severe disorders, indicating that

outcome is influenced by *H. pylori* strain virulence, the host response, or environmental factors.

A major *H. pylori* virulence determinant is the *cag* pathogenicity island (*cag* PAI) [3], a 40-kb region originally acquired horizontally and inserted into the glutamate racemase gene. Genes in the *cag* PAI encode a type IV secretion system through which CagA is translocated into gastric epithelial cells. Once inside the host cell, CagA disrupts signalling pathways through phosphorylation-dependent and phosphorylation-independent mechanisms, leading to abnormal proliferation and motility of gastric epithelial cells and to cytoskeletal changes [3]. CagA undergoes tyrosine phosphorylation by multiple members of the Src family of kinases on specific tyrosine residues within repeating Glu-Pro-Ile-Tyr-Ala (EPIYA) motifs. The EPIYA segments are defined as

EPIYA-A, EPIYA-B, EPIYA-C, and EPIYA-D, according to the amino acid sequence that surrounds the EPIYA motif. Once phosphorylated, CagA binds to a cytoplasmic Src homology 2 domain of Src homology 2 phosphatase (SHP-2) and deregulates SHP-2. CagA also binds to and inhibits PAR1b/MARK2 polarity-regulating kinase, disrupting apical junctions and epithelial-basolateral polarity [4,5]. Together, these CagA activities may contribute to the transformation of gastric epithelial cells. A second major virulence factor is the VacA cytotoxin, a protein with multiple effects on host cells [6]. In contrast to *cagA*, *vacA* is present in almost all *H. pylori* strains. Three major polymorphic regions have been identified: the signal region (s1 and s2), the intermediate region (i1 and i2), and the middle region (m1 and m2) [6,7]. The mosaic combination of s-region, m-region and i-region alleles in *vacA* determines production of the cytotoxin, as well as its host cell specificity and the degree of cytotoxicity [6–8].

As compared with the situation in industrialized countries, few data are available on the importance of *H. pylori* virulence factors for clinical outcome in Africa [9–12]. Here, we examined the potential association of *cagA* and *vacA* virulence factor polymorphisms with *H. pylori*-related diseases and histopathological lesions in symptomatic patients in Senegal.

## Materials and Methods

### Patients

Patients who suffered from upper abdominal pain were investigated at the Gastroenterology Department of Le Dantec Hospital (Dakar, Senegal) between January 2007 and December 2008. Demographic data, medical history and presenting symptoms were prospectively collected by the physician. All patients were of Senegalese origin, and none

had received proton-pump inhibitors or antibiotics during the 4 weeks before endoscopy. Patients were excluded from the study if they were culture-negative for *H. pylori*. Three biopsy samples were taken from the antrum and three from the fundus during upper gastrointestinal tract endoscopy. One biopsy sample from each site was cultured for *H. pylori* isolation, and the four others were fixed and processed for histological analysis. All endoscopes and accessory devices were decontaminated after every endoscopy, according to local written procedures. On the basis of endoscopic findings, patients were classified as having gastritis only, ulcerated lesion, or suspicion of neoplasia.

### Histology

Histopathological diagnoses were based on the updated Sydney System [13], and the Vienna classification for dysplasia [14]. Written informed consent was obtained from all patients, and the study protocol was approved by the Senegalese national ethics committee.

### *H. pylori* isolation, genomic DNA extraction, and determination of *cagA* status and *cagA/vacA* polymorphisms

*H. pylori* culture was performed by the use of Wilkins Chalgren agar plates with 7% (v/v) defibrinated horse blood and *H. pylori* selective antibiotic supplement (Oxoid, Basingstoke, UK), as previously described [15]. Genomic DNA was extracted from a mixture of colonies from antrum and fundus samples with a QIAmp kit (Qiagen, Courtaboeuf, France). Polymorphisms in the s-region, m-region and i-region of *vacA* and *cagA* status were determined as described elsewhere, using the primers reported in Table 1 [6,7,15,16]. The 3'-region of *cagA*, encoding the EPIYA segments, was sequenced, and the number and type of EPIYA segments were inferred as previously described [15]. Nucleotide

TABLE 1. List of oligonucleotide primers used

Gene	Region amplified	Primer	Primer sequence (5'–3')	Size (bp)	Reference
<i>vacA</i>	s1	vA1F	GAAATACAACAAACACACCCGC	259	[6]
		vA1R	GGCTTGTGGAGCCCCCAG		
	s2	vA1F	GAAATACAACAAACACACCCGC	286	[6]
		vA1R	GGCTTGTGGAGCCCCCAG		
	m1	vA3F	GGTCAAATGCGGTCATGG	290	[6]
		vA3R	CATCAGTATTTCCGACCACA		
	m2	vA4F	CCAGGAAACATTGCCGGCAAA	352	[6]
vA4R		CATAACTAGCGCCTTGAC			
i1	VacF1-i	GTTGGGATTGGGGGAATGCCG	426	[7]	
	C1R-i	TTAATTTAACGCTGTTTGAAG			
i2	VacF1-i	GTTGGGATTGGGGGAATGCCG	432	[7]	
	C2R-i	GATCAACGCTCTGATTTGA			
<i>cagA</i>	EPIYA	cagA2530S	GTTAARAATRGTTGTRAAAYGG	Variable	[15]
			R = A or G and Y = T or C		
	cagA (5'-region)	cagA3000AS	TTTAGCTTCTGATACCGC	183	[16]
		cagA-F	TTGACCAACAACCAACAACCGAAG		
Empty site	cagA-R	CTTCCCTTAATTGCGAGATTCC	360	[16]	
	2	ACATTTTGGCTAAATAAACGCTG			
	25	TCATGCGAGCGCGGATGTG			

sequences were deposited in GenBank under accession numbers HM362686–HM362767.

### Mixed infections

Patients with multiple *vacA* alleles were considered to have mixed infection. For patients with a single *vacA* allele and ambiguities after sequencing of the EPIYA PCR product, gastric biopsy samples were defrosted, and ten independent colonies were randomly selected from each primary culture (antrum and fundus) and subcultured. After DNA extraction, random amplification of polymorphic DNA (RAPD) PCR was used to assess the clonal relationship between isolates from the same patient, as previously described [17]. Only patients with identical *vacA* alleles and EPIYA combinations were subsequently analysed.

### Statistical analysis

Microsoft Access 2003 was used for data entry, and Stata Version 10 (StataCorp, College Station, TX, USA) for statistical analysis. Univariate analysis and logistic regression were used to identify determinants of ulcerated lesions and gastric cancer (GC) among *H. pylori*-positive patients. In both cases, patients with ulcerated lesions or GC were first compared with patients with any other lesion, and then with patients with gastritis only on the basis of endoscopic findings.

In univariate analysis, the chi-square test (or Fisher's exact test when appropriate) and Student's *t*-test were used to compare categorical and continuous variables, respectively. Factors associated with ulcerated lesion or GC in univariate analysis with *p*-values  $\leq 0.2$  were retained for multivariate analysis. A backward stepwise selection procedure was used to build the models. Interactions were then tested and the models were assessed with the Hosmer and Lemeshow goodness-of-fit test. The accuracy of the final models was assessed from the area under the receiver operating characteristic curve, sensitivity, and specificity, and *p*-values  $< 0.05$  were considered to denote significant associations in all analyses.

## Results

### Patients

Among the 187 patients included (Fig. 1), 117 (54.7% male) were culture-positive for *H. pylori*. The median age was 40.0 years (mean, 44.3 years; interquartile range (IQR) 31.0–56.6 years; range, 14–93 years). The predominant ethnic groups were Wolof (35.2%), Fulani (17.1%), and Serer (15.2%), in keeping with the ethnic distribution in Senegal. The most frequent presenting symptoms were weight loss

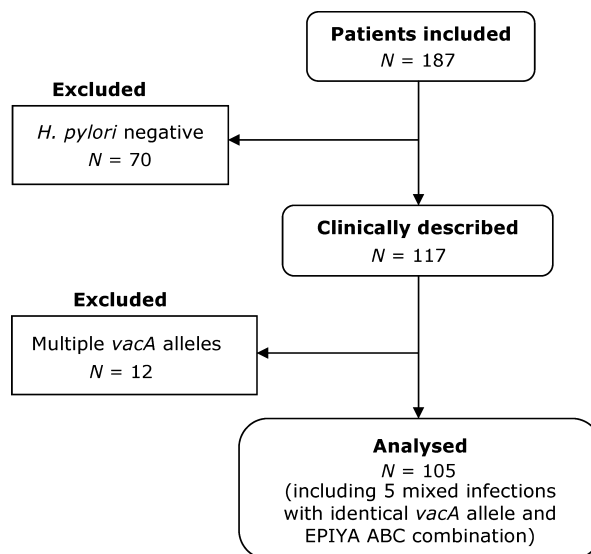


FIG. 1. Flow chart.

(*n* = 76, 65.0%), dyspepsia (*n* = 68, 58.1%), heartburn (*n* = 53, 45.3%), regurgitation (*n* = 50, 42.7%), anorexia (*n* = 49, 41.9%), nausea (*n* = 43, 36.8%) and dysphagia (*n* = 34, 29.1%).

### *vacA* typing and diversity

Multiple *vacA* alleles were detected in 12 cases. These patients were excluded from subsequent analyses. The results of *vacA* typing among the remaining 105 patients are shown in Table 2. All s1m1 strains were type i1, whereas all s2m2 strains were type i2. Most of the strains had presumed high-vacuolization isotypes (s1i1m1 (*n* = 60, 57.1%) or s1i1m2 (*n* = 23, 21.9%)).

### Endoscopic and histological findings

On the basis of endoscopic findings, 29 of the 105 patients had gastritis only (12 males; median age, 38.5 years; IQR 32.0–54.4), 61 had ulcerated lesion (38 males; median age, 40.0 years; IQR 31.0–50.9), and 15 had suspicion of neoplasia (eight males; median age, 59.0 years; IQR 55.6–67.1). All cases of suspected neoplasia were histologically confirmed as GC. Histological evaluations and the correlation with endoscopic findings are shown in Table 3.

### *cagA* status

The *cagA* gene was successfully amplified in 77 (73.3%) of the 105 *H. pylori* isolates. The absence of *cag* PAI was confirmed by empty site PCR in all 28 *cagA*-negative strains.

### Mixed infection

Five patients had possible mixed infection, as one or more nucleotide positions showed ambiguities after sequencing of

**TABLE 2.** Risk factors for gastric cancer

	Lesion		Univariate analysis			Multivariate analysis		
	Gastric cancer	Other	OR	95% CI	p-value	OR	95% CI	p-value
Age (years), mean (SD)	58.8 (13.4)	41.8 (15.8)	1.06	1.03–1.11	0.001	1.10	1.04–1.16	<0.001
Gender, n (%)								
Male	8 (13.8)	50 (86.2)	1	–	0.87			
Female	7 (14.9)	40 (85.1)	1.1	0.36–3.27				
<i>cagA</i> , n (%)								
Absent	1 (3.6)	27 (96.4)	1	–	0.09	1	–	0.03
Present	14 (18.2)	63 (81.8)	6.0	0.75–47.9		17.7	1.3–241	
Two EPIYA-C segments, n (%)								
No	12 (16.0)	63 (84.0)	1	–	0.03			
Yes	2 (100)	0 (0)	5.1	2.8–9.7				
<i>vacA</i> s, n (%)								
s1	15 (17.6)	70 (72.9)	–	–	0.002			
s2	0 (0)	20 (100)	–	–				
<i>vacA</i> m, n (%)								
m1	12 (20.0)	48 (80.0)	1	–	0.07			
m2	3 (6.8)	41 (93.2)	0.29	0.08–1.1				
<i>vacA</i> i, n (%)								
i1	14 (17.1)	68 (82.9)	1	–	0.15			
i2	1 (4.3)	22 (95.7)	0.22	0.03–1.8				
<i>vacA</i> si, n (%)								
Others	1 (4.3)	22 (95.7)	1	–	0.16			
s1i1	14 (17.1)	68 (82.9)	4.5	0.56–36.4				
<i>vacA</i> smi, n (%)								
Others	3 (6.7)	42 (93.3)	1	–	0.05	1	–	0.30
s1i1m1	12 (20.0)	48 (80.0)	3.5	0.92–13.2		2.2	0.5–10.3	

SD&lt; standard deviation.

the EPIYA PCR product. ES PCR gave negative results in all of these patients, indicating the absence of *cag* PAI-negative strains in the mixture. Two different RAPD banding patterns among the ten isolates from each patient were observed in all five patients, indicating the presence of two different clones. Seventeen (14.5%) of the 117 *H. pylori*-positive patients (including the 12 patients with multiple *vacA* alleles) had evidence of mixed infection. As all isolates from individual patients yielded EPIYA PCR products of identical size and an identical *vacA* allele, one representative isolate from each RAPD banding pattern was analysed for EPIYA diversity (ten in total).

#### EPIYA diversity

The EPIYA diversity of 82 *H. pylori* clones was analysed, revealing the presence of EPIYA-A, EPIYA-B and EPIYA-C segments. The four segments most frequently detected within each EPIYA segment are shown in Table 4. All of the strains shared the TIDD motif following EPIYA in segment C, except for one strain with a TIED motif. QVNK and QVAK occupied this position in all segments A and B, respectively. The EPIYA combinations were ABC ( $n = 75$ , 91.5%), AB ( $n = 3$ ), ABCC ( $n = 2$ ), AABC ( $n = 1$ ), and BC ( $n = 1$ ).

All gastroduodenal disorders were associated with a high frequency of strains possessing an EPIYA ABC combination: 80.0% (12 of 15) in GC, 90.7% (39 of 43) in ulcerated lesions, and 100.0% (19 of 19) in the remaining patients. The

**TABLE 3.** Correlation between the endoscopic and the histological findings (no. (%))

Histological finding	Gastritis	Ulcerated lesion	Suspicion of neoplasia	Total
Gastritis without any other lesion	1 (3.5)	5 (8.2)	0	6 (5.7)
ACG as the most severe lesion	4 (13.8)	5 (8.2)	0	9 (8.6)
IM as the most severe lesion	24 (82.7)	49 (80.3)	0	73 (69.5)
LGD as the most severe lesion	0	2 (3.3)	0	2 (1.9)
GC	0	0	15 (100)	15 (14.3)
Total	29 (100)	61 (100)	15 (100)	105 (100)

ACG, atrophic chronic gastritis; GC, gastric cancer; IM, intestinal metaplasia; LGD, low-grade dysplasia.

only two strains with an EPIYA ABCC combination were recovered from patients with GC.

#### Univariate analysis of factors associated with *H. pylori*-related diseases and histological types

Among the 105 patients analysed, no correlation was found between histological findings and either the *vacA* isotype or *cagA* status ( $p > 0.2$ ). Univariate analysis showed no correlation between ulcerated lesions and *cagA* status, EPIYA diversity, or *vacA* isotype ( $p > 0.2$ ), whether the comparator group was 'other lesions' or 'gastritis only'. As a result, multivariate analysis was not used to identify determinants of ulcerated lesion.



**TABLE 4.** The four most frequent EPIYA motifs among 82 *cagA*-positive isolates from 77 Senegalese patients. Bold type indicates the polymorphic sites

	n (%)
<b>EPIYA-A motif</b>	
KELNEKFKNFNNNNGLKNGKDKGPEEPIYAQVNKKKTGQVA	44 (53.6)
KELNEKFKNFNNNNN-GLKNSTEPIYAQVNKKKTGQVA	9 (11.0)
KELNEKFKNFNNNNN-GLKNGGPEPIYAQVNKKKTGQVA	7 (8.5)
KELNEKFKNFNNNNN-GLKNSGEPIYAQVNKKKTGQVA	5 (6.1)
<b>EPIYA-B motif</b>	
TGQVASPEEPIYAQVAKKVTQKIDQLNQAASFGGVGQAG	23 (28.0)
TGQVASPEEPIYAQVAKKV <b>SA</b> KIDQLNQAASFGGVGQAG	14 (17.1)
TGQVASPEEPIYAQVAKKVT <b>K</b> KIDQLNQAASFGGVGQAG	13 (15.9)
TGQVASPEEPIYAQVAKKV <b>NA</b> KINQLNQAASFGGVGQAG	13 (15.9)
<b>EPIYA-C motif</b>	
FP-LKRHDKVDDLKSVGRSVSPEPIYATIDDLGGS	25 (30.5)
FP-LKRHDKVEDLSKVGRSVSPEPIYATIDDLGGS	15 (18.3)
FP-LKRHDKVDDLKSVGRSVSPEPIYATIDDLGGP	8 (9.8)
FP-LKKHDKVEDLSKVGRSVSPEPIYATIDDLGGS	7 (8.5)

Factors associated with GC were older age, s1-type *vacA*, and two EPIYA-C segments (Table 2). However, the latter two factors could not be included in the multivariate analysis, owing to sampling zeroes. The univariate p-values of these two factors are therefore considered.

**Multivariate analysis of factors associated with GC**

The two models used to identify determinants of GC were similar. Only *cagA* status and older age were independently associated with GC. No statistically significant interaction was found. As this model showed a poor fit (p 0.04), smi allele status was added to the final model, which allowed us to include s1 allele status (Table 2). This latter model fitted well (p 0.37): the area under the receiver operating characteristic curve was 0.87, the sensitivity was 80.0%, and the specificity was 83.3%.

A third model was used to compare the 17 patients with GC or dysplasia with the other patients. This model was similar to the previous one, but had lower sensitivity and specificity (76.4% and 78.4%, respectively).

**Discussion**

On the basis of age-standardized incidence rates, Senegal is classified as a country with a low risk of GC (data available at <http://globocan.iarc.fr/>). Seven different genetic populations of *H. pylori* have been described worldwide, and their distribution is closely related to the distribution of the incidence of GC [18]. In Africa, where most strains are assigned to hpAfrica1, the incidence of GC is low, although the incidence is probably underestimated, owing to limited access to healthcare facilities and short life-expectancies of populations. The genetic background might be a marker of virulence factors directly involved in clinical outcome. Seventy-

nine of our 105 isolates have been randomly selected for multilocus sequence analysis with a previously described method [19], and all were assigned to hpAfrica1 (*S. Breurec*, unpublished data).

The *cagA* gene was present in 73.3% of our isolates, a proportion similar to that reported elsewhere in Africa [9,10]. As in many western studies [20], *cagA*-positive strains were significantly associated with GC in Senegal. To the best of our knowledge, such an association has never previously been described in an African population, probably owing to insufficient data for individual countries in previous studies. CagA protein EPIYA segment diversity has been suggested to be a major determinant of *H. pylori* pathogenesis. As strains with multiple EPIYA-C segments were recovered only from patients with GC, we were unable to measure the strength of the association between these two factors in multivariate analysis. However, in univariate analysis, the presence of two EPIYA-C segments was associated with an increased risk of GC. SHP-2 specifically binds to the tyrosine phosphorylated EPIYA-C segment of CagA, resulting in the stimulation of SHP-2 phosphatase activity [21]; this deregulation is involved in a variety of human malignancies [22]. Strains with a higher number of EPIYA-C segments exhibit a stronger propensity to interact with SHP-2, and are more closely associated with precancerous lesions and GC [23,24].

Although most of the patients had intestinal metaplasia as their most severe lesion, the small proportion of strains harbouring at least two EPIYA-C segments (2.4%) relative to other continents (11.5–30%) is one of our most important findings [15,23–27]. In the only other study of isolates from African patients (South Africa), 27% of *H. pylori* isolates contained two or more EPIYA-C segments, but this discrepancy could be attributable to the selection criteria, as 45% of patients in the South African study had GC [12]. Thus, the low prevalence of strains with multiple EPIYA-C segments might contribute to the low incidence of GC in Senegal and, by extension, in Africa, although further studies are needed to evaluate their prevalence in other African countries.

Here, we investigated *vacA* region i for the first time in strains from an African population. The *vacA* s1 genotype was linked to an increased risk of GC in univariate analysis, as already suggested in a meta-analysis of African studies [11]. Its presence in all the strains associated with GC prevented us from measuring the strength of the association in multivariate analysis. Type m1, which was also found to be a risk factor for GC in the meta-analysis, was not associated with an enhanced risk of GC in our study, possibly owing to the small sample size. i1-type strains have been strongly associated with GC in Iranians [7], but not in East Asian or

Southeast Asian populations [28]. The lack of association between the iI type and GC in Senegalese patients highlights the difficulty of extrapolating results from one geographical area to another.

The main bias in most studies of *H. pylori* mixed infection is that bacterial heterogeneity within individuals is not extensively assessed. We tried to overcome this weakness by using an approach based on *H. pylori* isolation from gastric biopsy samples obtained with the sweeping method, followed by molecular investigations [15]. Fourteen per cent of our patients had evidence of mixed infection. This, combined with data from South Africa [29], suggests that mixed infections are common in African populations. Knowledge of the prevalence of mixed infections in a given population is of major importance. Indeed, recombination during mixed infection has been identified as a major driving force behind allelic diversity in *H. pylori*, which may affect individual disease outcomes.

The data reported here are particularly important, given the difficulty of carrying out such studies in hospitals in countries with inadequate healthcare systems. Consequently, most data on *H. pylori* and GC in Africa are from South Africa. The lack of systematic gastric biopsies during upper gastrointestinal tract endoscopy explains the over-representation of ulcerated lesions and GC in our study population. Even though our findings may not be representative of the overall situation, it should be noted that the study took place in one of the main tertiary hospitals in Senegal. Despite the small number of GC cases studied, we found that increasing age and the presence of *cagA*, two EPIYA-C segments and the sI *vacA* allele were associated with GC, suggesting that the relationship is robust. Our isolates were characterized by a lower prevalence of multiple EPIYA-C segments than reported in other countries, which may partly explain the low incidence of GC in Senegal.

## Acknowledgements

We thank S. Bergère (Institut Pasteur, Unité de Recherche et d'Expertises en Histotechnologie et Pathologie, Paris, France) for her technical assistance, and all the clinicians involved in this study. We also thank C. Bouchiat (Institut Pasteur, Dakar, Senegal) for helpful discussions, and I. Eronini and D. Young for editorial assistance.

## Transparency Declaration

This study was supported by a grant from Institut Pasteur (ACIP *Helicobacter pylori*) and was coordinated by the Institut

Pasteur in Dakar. All of the authors declare that they have no conflicts of interest.

## References

- Covacci A, Censini S, Bugnoli M *et al.* Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. *Proc Natl Acad Sci USA* 1993; 90: 5791–5795.
- Blaser MJ, Perez-Perez GI, Kleanthous H *et al.* Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res* 1995; 55: 2111–2115.
- Hatakeyama M. *Helicobacter pylori* and gastric carcinogenesis. *J Gastroenterol* 2009; 44: 239–248.
- Saadat I, Higashi H, Obuse C *et al.* *Helicobacter pylori* CagA targets PARI/MARK kinase to disrupt epithelial cell polarity. *Nature* 2007; 447: 330–333.
- Zeaiter Z, Cohen D, Musch A *et al.* Analysis of detergent-resistant membranes of *Helicobacter pylori* infected gastric adenocarcinoma cells reveals a role for MARK2/Par1b in CagA-mediated disruption of cellular polarity. *Cell Microbiol* 2008; 10: 781–794.
- Atherton JC, Cao P, Peek RM *et al.* Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific *vacA* types with cytotoxin production and peptic ulceration. *J Biol Chem* 1995; 270: 17771–17777.
- Rhead JL, Letley DP, Mohammadi M *et al.* A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology* 2007; 133: 926–936.
- Cover TL, Blaser MJ. Purification and characterization of the vacuolating toxin from *Helicobacter pylori*. *J Biol Chem* 1992; 267: 10570–10575.
- Sanz-Pelaez O, Santana-Rodriguez E, Maroto AA *et al.* *Helicobacter pylori* and *cagA* seroprevalence in sub-Saharan immigrants recently arrived to Gran Canaria (Spain). *Scand J Infect Dis* 2008; 40: 756–758.
- Smith SI, Kirsch C, Oyedeji KS *et al.* Prevalence of *Helicobacter pylori vacA*, *cagA* and *iceA* genotypes in Nigerian patients with duodenal ulcer disease. *J Med Microbiol* 2002; 51: 851–854.
- Sugimoto M, Yamaoka Y. The association of *vacA* genotype and *Helicobacter pylori*-related disease in Latin American and African populations. *Clin Microbiol Infect* 2009; 15: 835–842.
- Argent RH, Kidd M, Owen RJ *et al.* Determinants and consequences of different levels of CagA phosphorylation for clinical isolates of *Helicobacter pylori*. *Gastroenterology* 2004; 127: 514–523.
- Dixon MF, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. *Am J Surg Pathol* 1996; 20: 1161–1181.
- Schlemper RJ, Riddell RH, Kato Y *et al.* The Vienna classification of gastrointestinal epithelial neoplasia. *Gut* 2000; 47: 251–5.
- Panayotopoulou EG, Sgouras DN, Papadakos K *et al.* Strategy to characterize the number and type of repeating EPIYA phosphorylation motifs in the carboxyl terminus of CagA protein in *Helicobacter pylori* clinical isolates. *J Clin Microbiol* 2007; 45: 488–495.
- Akopyants N, Clifton SW, Kersulyte D *et al.* Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. *Mol Microbiol* 1998; 28: 37–53.
- Taylor NS, Fox JG, Akopyants NS *et al.* Long-term colonization with single and multiple strains of *Helicobacter pylori* assessed by DNA fingerprinting. *J Clin Microbiol* 1995; 33: 918–923.
- Yamaoka Y, Kato M, Asaka M. Geographic differences in gastric cancer incidence can be explained by differences between *Helicobacter pylori* strains. *Intern Med* 2008; 47: 1077–1083.

19. Moodley Y, Linz B, Yamaoka Y et al. The peopling of the Pacific from a bacterial perspective. *Science* 2009; 323: 527–530.
20. Huang JQ, Zheng GF, Sumanac K et al. Meta-analysis of the relationship between *cagA* seropositivity and gastric cancer. *Gastroenterology* 2003; 125: 1636–1644.
21. Higashi H, Tsutsumi R, Fujita A et al. Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci USA* 2002; 99: 14428–14433.
22. Tartaglia M, Niemeyer CM, Fragale A et al. Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat Genet* 2003; 34: 148–150.
23. Xia Y, Yamaoka Y, Zhu Q et al. A comprehensive sequence and disease correlation analyses for the C-terminal region of CagA protein of *Helicobacter pylori*. *PLoS ONE* 2009; 4: e7736.
24. Argent RH, Hale JL, El-Omar EM, Atherton JC. Differences in *Helicobacter pylori* CagA tyrosine phosphorylation motif patterns between western and East Asian strains, and influences on interleukin-8 secretion. *J Med Microbiol* 2008; 57: 1062–1067.
25. Reyes-Leon A, Atherton JC, Argent RH et al. Heterogeneity in the activity of Mexican *Helicobacter pylori* strains in gastric epithelial cells and its association with diversity in the *cagA* gene. *Infect Immun* 2007; 75: 3445–3454.
26. Devi SM, Ahmed I, Khan AA et al. Genomes of *Helicobacter pylori* from native Peruvians suggest admixture of ancestral and modern lineages and reveal a western type *cag*-pathogenicity island. *BMC Genomics* 2006; 7: 191.
27. Basso D, Zambon CF, Letley DP et al. Clinical relevance of *Helicobacter pylori cagA* and *vacA* gene polymorphisms. *Gastroenterology* 2008; 135: 91–99.
28. Ogiwara H, Graham DY, Yamaoka Y. *vacA* i-region subtyping. *Gastroenterology* 2008; 134: 1267.
29. Schwarz S, Morelli G, Kusecek B et al. Horizontal versus familial transmission of *Helicobacter pylori*. *PLoS Pathog* 2008; 4: e1000180.

## 2. DISCUSSION

### 2.1. Migrations humaines à travers un marqueur bactérien : *H. pylori*

#### 2.1.1. *H. pylori*, le fidèle compagnon de l'homme dans ses migrations

La diversité génétique de l'ADN humain moderne provient de l'accumulation d'évènements génétiques transmis par nos ancêtres. Cette diversité, connue sous le nom de polymorphismes de l'ADN, peut être analysée et utilisée pour rechercher notre histoire génétique, complétant par conséquent les approches plus traditionnelles de l'archéologie, de la paléontologie, de l'anthropologie et de la linguistique.

Deux marqueurs génétiques ont joué un rôle majeur dans l'histoire de cette science : le chromosome Y spécifique du sexe masculin, et l'ADNmt, apporté spécifiquement par la mère dans les mitochondries de ses ovules. L'hérédité uniparentale, l'absence de recombinaison sur la totalité de l'ADNmt et la plus grande partie du chromosome Y (Giles *et al.* 1980; Jobling and Tyler-Smith 2003) sont des avantages inhérents permettant de décrire différents haplotypes (que l'on peut regrouper d'après leurs relations phylogénétiques en haplogroupes) et d'estimer la chronologie des événements grâce à des arbres phylogénétiques en utilisant des taux de mutation calibré. Ces caractéristiques rendent l'interprétation des histoires génétiques des hommes moins complexe que celles déduites de l'analyse des gènes présents sur les autosomes. Ces deux marqueurs génétiques ont apporté une contribution majeure à la compréhension de l'histoire des populations. Ainsi, elle soutient le modèle qui propose que tous les hommes modernes puissent être raccordés à une population ancestrale unique de taille limitée située en Afrique, et qui se serait répandue à travers les autres continents. Ce modèle est également supporté par l'étude des autosomes (Underhill and Kivisild 2007).

L'ADNmt est une molécule circulaire de 16569 pb. Son polymorphisme est défini par diverses mutations dans la séquence ADN, notamment dans la région hypervariable 1 (HVSI) et la région hypervariable 2 (HVSII) de la région contrôle mais aussi dans la région codante. Connu pour muter cinq à dix fois plus vite que l'ADN nucléaire (Brown *et al.* 1979), l'ADNmt a été pendant longtemps le marqueur génétique de référence. Cependant, bien qu'un effort conséquent ait été consenti depuis quelques années pour étudier la totalité de l'ADNmt (Torroni *et al.* 2006), les régions les plus étudiées ont été les régions HVSI et HVSII. Leur longueur étant de quelques centaines de paires de bases, le polymorphisme nucléotidique limité ne permettait pas de retracer avec précision les migrations humaines. De plus, l'ADNmt se caractérise par un haut niveau d'homoplasie (apparitions indépendantes de

mutations similaires chez des groupes n'étant pas nécessairement apparentés) liée à des zones privilégiées de mutation (hot-spot), qui va brouiller le signal phylogénétique (Torroni *et al.* 2006). Enfin, les relations génétiques entre les populations humaines déduites sur la base de la diversité de l'ADNmt ne montrent pas toujours une forte correspondance avec le groupe linguistique, ethnique et géographique, contrairement à l'étude du chromosome Y (Poloni, Semino *et al.* 1997). Ces différences ont été attribuées à des facteurs culturels, démographiques et sociaux spécifiquement masculins, tels que la polygamie, l'exogamie patrilocale (tendance pour une femme à se déplacer dans la maison natale de son mari), un plus fort taux de migration des femmes que des hommes (à petite distance). A titre d'exemple, dans plusieurs tribus du Sinaï, une faible diversité génétique du chromosome Y a été mise en évidence par rapport à celle de l'ADNmt. Il s'est avéré que les tribus étudiées étaient polygames, en parfait accord avec ces observations (Salem *et al.* 1996). En Afrique du Sud, les « coloured » désigne un groupe ethnique issu principalement de mélanges interraciaux entre Européens, Khoisans, Indonésiens, Indiens, Malgaches, Mozambicains et Mauriciens. L'analyse du chromosome Y et de l'ADNmt a permis de démontrer que la dynamique de fondation de cette population a été très différente entre hommes et femmes : l'ascendance des hommes est à 38 % d'origine européenne et 45 % d'origine africaine alors que celle des femmes est majoritairement d'origine khoisane (79 %) et très faiblement d'origine européenne (5 %) (Quintana-Murci *et al.* 2010), reflétant principalement le métissage entre des hommes européens ou africains avec des femmes aborigènes d'origine khoisane il y a environ 350 ans. A cette époque, l'union entre des européennes et des africains n'était socialement pas tolérée.

La grande taille du chromosome Y (approximativement 60 millions paires de base) et le faible niveau d'homoplasie (Jobling and Tyler-Smith 2003) permet de mieux retracer l'histoire démographique et la structure génétique de différentes populations à travers le temps (Underhill and Kivisild 2007). Cependant, la plupart des études ont étudié une petite portion du chromosome (<1 %) d'où un manque de résolution. De plus, certains mécanismes brouillent le signal phylogénétique: i) de nombreuses mutations semblent plus liées à un mécanisme de conversion génique plutôt qu'à un polymorphisme nucléotidique (Adams *et al.* 2006), ii) la sélection pourrait jouer un rôle important dans l'évolution du chromosome Y et serait alors responsable d'une réduction de la diversité génétique (Pannell and Pujol 2009), iii) de larges délétions et insertions chromosomiques surviennent fréquemment (Repping *et al.* 2006).

Devant ces inconvénients, d'autres alternatives ont été développées, utilisant des commensaux et des pathogènes humains possédant une structure phyléogéographique. Cependant, la transmission horizontale fréquente de ces micro-organismes entre individus non apparentés est un facteur limitant pour retracer les migrations humaines. Ce problème apparaît principalement avec les virus, mais également avec les mycobactéries. Une étude de la diversité génétique de 875 souches de *Mycobacterium tuberculosis* originaires d'Afrique, d'Amérique, d'Asie et d'Europe montre un assignement à 6 clusters génétiques majeurs, associés de façon imparfaite à de larges régions géographiques (Gagneux *et al.* 2006). Cependant, la topologie de l'arbre phylogénétique n'indique pas une origine ancestrale africaine de l'ensemble des souches, ce qui va à l'encontre d'une coévolution entre l'homme et la bactérie. De plus, *M. tuberculosis* possède un polymorphisme nucléotidique limité, responsable d'un manque de résolution. Ce polymorphisme nucléotidique limité est encore plus prononcé chez *Mycobacterium leprae*. Monot *et al.* ont recherché des mutations connues sous le nom de polymorphismes d'un seul nucléotide (SNPs) qui sont des substitutions de simples nucléotides à des endroits précis du génome. Seules trois régions du génome contenaient des SNPs informatifs dont la combinaison générait 4 sous-types après analyse de 175 isolats collectés dans le monde (Monot *et al.* 2005). Le type I était prédominant en Asie centrale, le type II en Afrique de l'est, le type III en Europe, en Afrique du Nord et en Amérique, le type IV en Afrique de l'ouest et les Caraïbes. Une coévolution entre l'homme et la bactérie est fortement suspectée sur la base de la distribution de ces sous-types. Une origine africaine ou proche-orientale de la bactérie est privilégiée. Les hommes auraient ensuite emporté la maladie dans leurs pérégrinations, vers l'Asie et vers l'Europe, ce dernier continent étant particulièrement touché au Moyen Age. La colonisation européenne la propage ensuite en Afrique de l'ouest. Avec le commerce des esclaves africains, elle se répand au Brésil et aux Antilles. Cependant, ce marqueur manque de résolution pour retracer les détails des migrations humaines.

Le virus JC appartient à la famille des Papovaviridae humains et a été proposé comme marqueur des migrations humaines : i) l'infection est acquise en général pendant l'enfance avec une transmission principalement des parents à l'enfant, ii) au moins 14 sous-types viraux ont été décrits dont certains sont associés à des populations humaines spécifiques (Yogo *et al.* 2004). Cependant, une analyse récente de l'évolution génomique a démontré que le virus JC n'a pas coévolué avec l'homme mais s'est plutôt développé parallèlement à certaines populations humaines (Shackelton *et al.* 2006), une limite également observée chez d'autres marqueurs potentiels tels que le virus T-lymphotropique humain (Miura *et al.* 1994; Cassar *et*

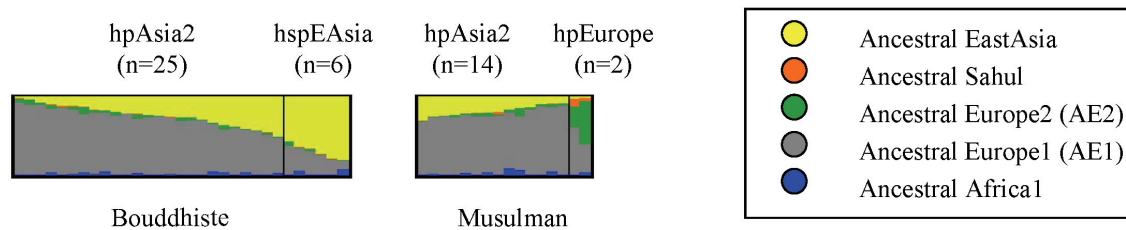
*al.* 2007), l'herpèsvirus humain type 8 (Cassar *et al.* 2007) et le papillomavirus humain (Ho *et al.* 1993).

En l'absence de « gold standard », la bactérie *H. pylori* est un bon candidat pour retracer les migrations humaines puisque :

- elle colonise plus de 50 % de la population mondiale,
- elle était déjà présente dans l'estomac des *Homo sapiens* il y a environ 58 000 ans, avant les grandes migrations de l'homme moderne, partant de l'Afrique de l'est vers l'Asie et l'Europe,
- aucun autre réservoir que l'estomac humain n'a jamais été décrit (Azevedo *et al.* 2007),
- sa diversité génétique, liée à de fréquentes recombinaisons, est une des plus importantes du monde bactérien (Vos and Didelot 2009). L'analyse du polymorphisme génétique de 7 gènes de ménage a montré que les événements génétiques sont si fréquents que pratiquement tous les isolats sont assignés à un nouveau ST et que 50 % du génome peut être remplacé par recombinaison homologue pendant un intervalle de temps compris entre 40 et 2000 ans (Falush, Kraft *et al.* 2001).
- la transmission de l'infection a lieu principalement pendant l'enfance et est le plus souvent intrafamiliale ou au sein de communautés relativement restreintes en nombre, limitant la diffusion de l'infection, ce qui est primordial pour retracer des migrations humaines anciennes (Linz *et al.* 2007; Raymond *et al.* 2008; Schwarz *et al.* 2008).

Wirth *et al.* ont été les premiers à montrer que l'étude de la diversité génétique de 7 gènes de ménage de souches de *H. pylori* est plus résolutive que celle de la diversité génétique humaine (région HVS1 de l'ADNmt et microsatellites) pour retracer les migrations humaines (Wirth *et al.* 2004). Les populations génétiques modernes de *H. pylori* dérivent de six populations ancestrales : AE1, AE2, ancestral EastAsia, ancestral Sahul, ancestral Africa1 et ancestral Africa2 (Falush, Wirth *et al.* 2003; Linz, Balloux *et al.* 2007; Moodley, Linz *et al.* 2009) (cf *Partie I. Généralités : chapitre 6.2.2. Microdiversité*). Wirth *et al.* ont étudié deux populations vivant au Ladakh (nord de l'Inde), l'une bouddhiste et l'autre musulmane. Tandis que les isolats provenant de patients musulmans dérivent de façon presque exclusive de la population ancestrale AE1 (population moderne hpAsia2), mis à part deux souches possédant un haut niveau d'AE2 (population moderne hpEurope), ceux isolés des bouddhistes présentaient un haut niveau de recombinaison entre AE1 et ancestral EastAsia (populations modernes hpAsia2 et hspEAsia), témoin d'une coévolution entre ces deux populations ancestrales pendant des millénaires avant leur arrivée dans la région (Figure 21). La diversité

génétique de *H. pylori*, presque deux fois plus importante que celle de la région HVS1 de l'ADNmt, reflétait l'histoire du Ladakh : i) une migration il y a environ 1000 ans de bouddhistes du Tibet, ii) une migration de musulmans en provenance du Pakistan après le 14<sup>ème</sup> siècle, iii) une coexistence pendant des siècles avec peu de mélange entre les deux communautés du à un isolement culturel et religieux.



**Figure 21. DISTRUCT plot des proportions de nucléotides issues des populations ancestrales composant des souches de *H. pylori* isolées dans le nord de l'Inde (Ladakh), chez des musulmans et des bouddhistes (logiciel STRUCTURE V2.0, modèle linkage)**

Représentation proposée par S. Breurec d'après les données de Wirth *et al.* (Wirth *et al.* 2004) Une ligne verticale pour chaque isolat indique la proportion de nucléotides issus de chaque population ancestrale identifiée par une couleur différente. Une ligne verticale noire sépare les isolats en (sous-) populations génétiques (logiciel Structure, modèle « no linkage »).

Des similarités génétiques entre des souches de *H. pylori* isolées dans certaines populations iraniennes et dans des populations turques, ouzbeks, palestiniennes et israéliennes ont été retrouvées par analyse de la représentation des distances génétiques entre ces populations. Elles reflètent des contacts historiques depuis 2000 ans entre populations iraniennes et populations des pays voisins. De telles relations n'avaient pu être mises en évidence par l'analyse de la diversité génétique de l'ADNmt et du chromosome Y (Latifi-Navid *et al.* 2010).

Ainsi, *H. pylori* représente une alternative prometteuse à l'étude des marqueurs génétiques humains pour l'étude des migrations des différentes populations ainsi que de leurs interactions.



### 2.1.2. *H. pylori* et les migrations humaines en Asie du sud-est et en Océanie

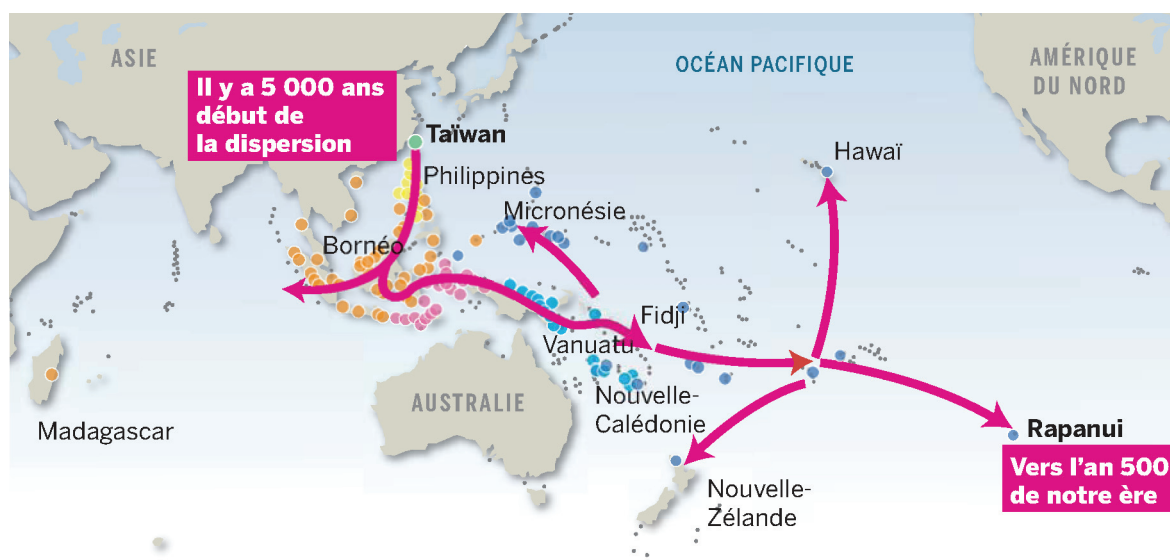
Les études de la diversité génétique de *H. pylori* que nous avons menées en Océanie (Article n°1) et en Asie du sud-est (Article n°2) ont permis d'affiner les connaissances sur les migrations humaines dans ces régions du monde.

#### 2.1.2.1. Océanie

Ma contribution à l'article n°1 n'ayant concerné que la colonisation du Pacifique par les locuteurs des langues austronésiennes, le peuplement préhistorique du Sahul n'a pas été développé ici.

La colonisation de la Polynésie est le sujet de débats au sein de la communauté scientifique depuis des décennies. La famille de langues austronésiennes couvre une aire d'extension considérable, presque la moitié du globe : 26 000 km<sup>2</sup> de l'île de Madagascar dans l'océan Indien, en passant par Taiwan et une partie du sud-est asiatique, principalement l'Indonésie et les Philippines, pour couvrir presque tout le Pacifique et former ce qu'on appelle le «triangle polynésien» dont les sommets sont Hawaï au nord, l'île de Pâques au sud-est et la Nouvelle-Zélande au sud-ouest. Avant la récente expansion des Européens, l'austronésien était la famille linguistique la plus répandue dans le monde. Comment expliquer l'apparement de ces idiomes sur une si vaste étendue ? Où trouver leur origine commune?

Les 1200 langues austronésiennes sont divisées en 10 groupes : 9 groupes sont uniquement parlés par les tribus aborigènes de Taiwan (langues formosanes) et le dernier groupe correspond aux langues malayo-polynésiennes, parlées à l'extérieur de Taiwan. Récemment, Gray *et al.* ont étudié le vocabulaire de 400 de ces langues austronésiennes (Gray *et al.* 2009). En utilisant des méthodes d'analyse statistique issues de la biologie, ils ont comparé les lexiques pour bâtir un arbre généalogique de cette famille de langues. La construction de cette phylogénie dresse un scénario de migrations extrêmement détaillé. La langue à l'origine de tous les idiomes austronésiens était ainsi parlée il y a environ 5 200 ans, à Taiwan (Figure 22). Ce modèle suggère que des agriculteurs parlant des langues austronésiennes ont colonisé l'Océanie, à partir de Taiwan, via les Philippines, la Mélanésie, puis la Polynésie, il y a 5000 ans, en plusieurs « pulsations ». Ces vagues migratoires seraient consécutives à la maîtrise de l'agriculture et donc à l'augmentation de la démographie, au moins pour la période la plus ancienne.



**Figure 22. Migration des Austronésiens à partir de Taiwan en se basant sur la généalogie de 400 langues appartenant aux sous-groupes ou groupes des langues austronésiennes.**

D'après Gray *et al.* (Gray *et al.* 2009).

Les cercles de couleur différente correspondent aux principaux sous-groupes et groupes de langues austronésiennes inclus.

Les données archéologiques confortent-elles cette hypothèse ? A partir du 4<sup>ème</sup> millénaire avant JC environ, des outils de pierre polie et un style de poterie décoré spécifique (Ta-p'en-k'eng) ont fait leur apparition sur l'île de Taiwan. L'archéologie montre qu'un millénaire environ après que la culture « Ta-p'en-k'eng » se soit développée à Taiwan, une culture manifestement dérivée de celle-ci a essaimé de plus en plus loin de cette île, jusqu'à investir le domaine austronésien moderne. Cette culture porte le nom de « Lapita », du nom du site archéologique où elle a été décrite. Plusieurs centaines de sites archéologiques « lapita » ont été retrouvés dans une aire allant de la Nouvelle-Guinée jusqu'aux îles Samoa (archipel Bismarck, îles Salomon, Vanuatu, Nouvelle-Calédonie, Fidji, Tonga, Samoa, Wallis et Futuna) (Spriggs 1996). Les colons « lapita » fabriquaient des poteries avec des décorations extrêmement riches et variées effectuées sur engobe rouge, en utilisant des outils d'obsidienne provenant des volcans et de multiples ornements façonnés dans des coquillages. Ils cultivaient des tubercules (igname, taro), des fruits (noix de coco, banane), élevaient des cochons, des pintades et des chiens. De plus, ils étaient de remarquables navigateurs (pirogues, catamarans). Bien qu'il soit naturellement impossible d'affirmer que les

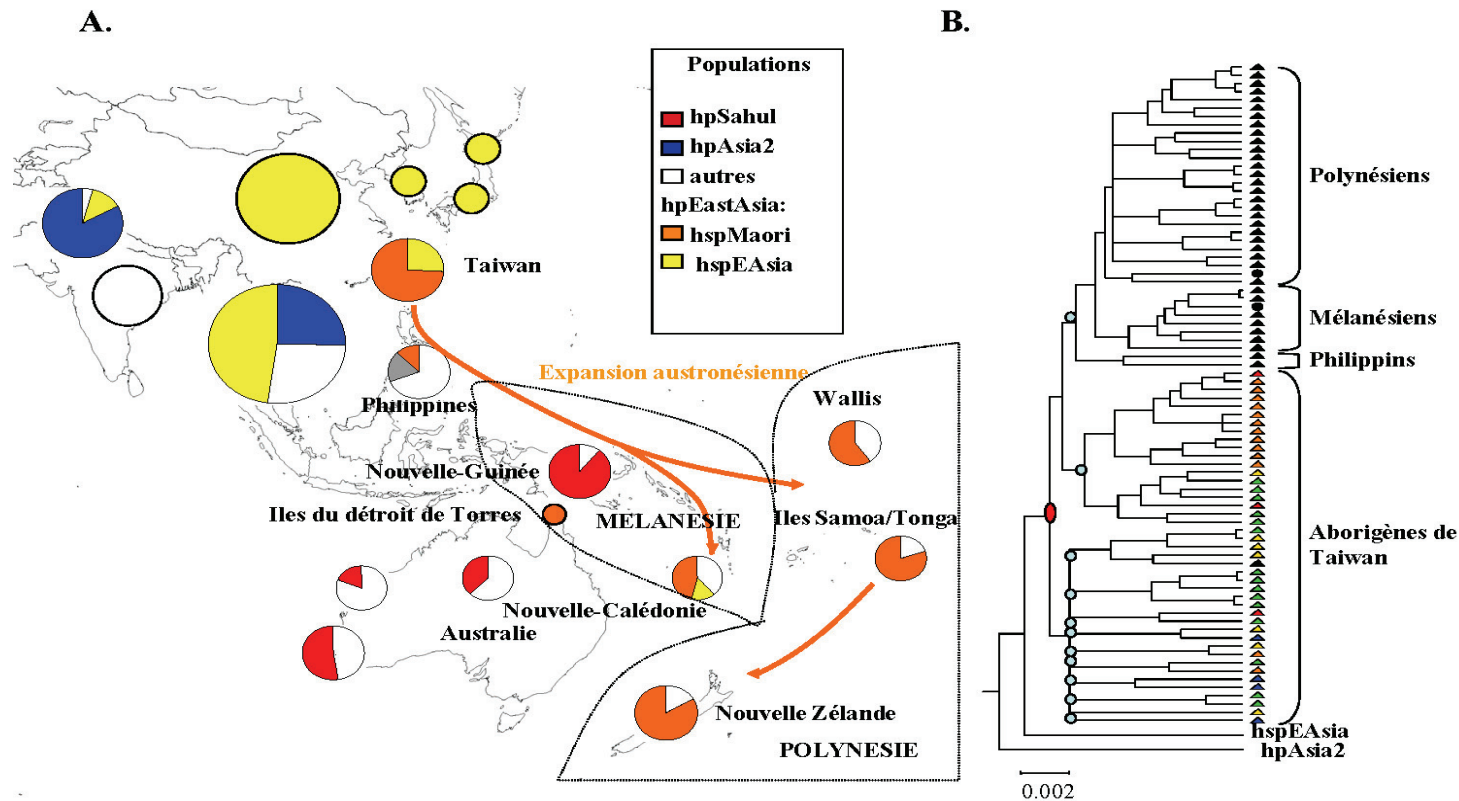
populations qui réalisaient les poteries « lapita » parlait des langues austronésiennes, il est reconnu que la civilisation « lapita » marque l'arrivée des Austronésiens dans la région.

Qu'apportent les marqueurs génétiques humains sur l'origine des Austronésiens ? L'haplogroupe B4a1a de l'ADNmt est retrouvé presque spécifiquement et à une fréquence élevée chez les Austronésiens. Il pourrait être le marqueur génétique de cette migration à partir de Taiwan (Melton *et al.* 1995; Sykes *et al.* 1995; Trejaut *et al.* 2005). Cependant, la plupart des études ont analysé le polymorphisme nucléotidique de la région hypervariable I de l'ADNmt, dont la faible résolution phylogénétique rend l'interprétation souvent difficile, expliquant le manque de consensus entre les résultats. Certaines études indiquent une origine taiwanaise de l'expansion austronésienne en Océanie (Redd *et al.* 1995; Sykes *et al.* 1995; Trejaut *et al.* 2005; Kayser *et al.* 2006), d'autres une origine indonésienne (Richards *et al.* 1998; Oppenheimer and Richards 2001; Hill *et al.* 2007). Récemment, une étude portant sur 157 séquences complètes de l'ADNmt et 4793 séquences de HVS-I issus d'Asie du sud-est continental, d'Asie du sud-est insulaire, de Taiwan, de Mélanésie et de Polynésie indiquait que les Austronésiens seraient arrivés dans l'archipel Bismarck de Papouasie-Nouvelle-Guinée plusieurs milliers d'années avant l'arrivée de la civilisation « lapita » en Mélanésie, excluant ainsi un modèle d'expansion à partir de Taiwan. Pour expliquer la distribution des langues austronésiennes en Océanie, les auteurs avancent l'hypothèse que de petits mouvements migratoires entre Taiwan et l'Asie du sud-est insulaire ont permis l'implantation des langues austronésiennes chez des populations installées de longue date dans l'archipel Bismarck, avant l'expansion en Océanie, sans incidence génétique majeure sur la population autochtone. Cette thèse reste cependant fondée sur des datations purement génétiques et n'est à ce jour pas étayée par des preuves archéologiques (Soares *et al.* 2011).

L'étude du chromosome Y montre principalement une origine mélanésienne (Kayser *et al.* 2000; Underhill *et al.* 2001; Kayser *et al.* 2006). Si la migration est originaire de Taiwan, ces données, outre un long passage des locuteurs des langues austronésiennes en provenance d'Asie par la Mélanésie, avant de coloniser la Polynésie, suggèrent aussi un fort métissage entre Austronésiens et Mélanésiens. L'origine différente de l'ADNmt et du chromosome Y pourrait refléter la résidence matrilocale ancestrale dans la société polynésienne, c'est-à-dire que lors du mariage, le mari rejoint le foyer de sa femme en délaissant le sien.

Qu'apporte l'étude de la diversité génétique de *H. pylori* ? Elle indique sans ambiguïté l'origine taiwanaise des Austronésiens, ce qui recoupe de manière surprenante les données de la linguistique et de l'archéologie:

- le groupe génétique hspMaori, marqueur de cette migration, est retrouvé uniquement chez les peuples parlant des langues austronésiennes (Taiwan, Philippines, Détroit de Torres, Nouvelle-Calédonie, Wallis et Futuna, Iles Tonga/Samoa, Nouvelle-Zélande) (Figure 23A),
- la diversité génétique de la population hspMaori est significativement plus élevée à Taiwan qu'en dehors,
- l'arbre généré par le logiciel CLONAL FRAME montre une migration séquentielle de Taiwan, vers les Philippines, la Mélanésie puis la Polynésie (Figure 23B),
- l'hypothèse d'une transmission de l'infection à *H. pylori* des aborigènes de Taiwan à des populations indigènes installées de longue date dans l'archipel Bismarck, suite à des petits mouvements migratoires entre ces deux régions, avant l'expansion en Océanie, est hautement improbable en raison de l'épidémiologie de l'infection à *H. pylori* (transmission de l'infection principalement pendant l'enfance à la suite de contacts rapprochés), ce qui va à l'encontre de la théorie précédemment développée par Soares *et al.* (Soares *et al.* 2011).



**Figure 23. hspMaori, marqueur de la migration des Austronésiens en Océanie à partir de Taiwan**

S. Breurec (Article n°1).

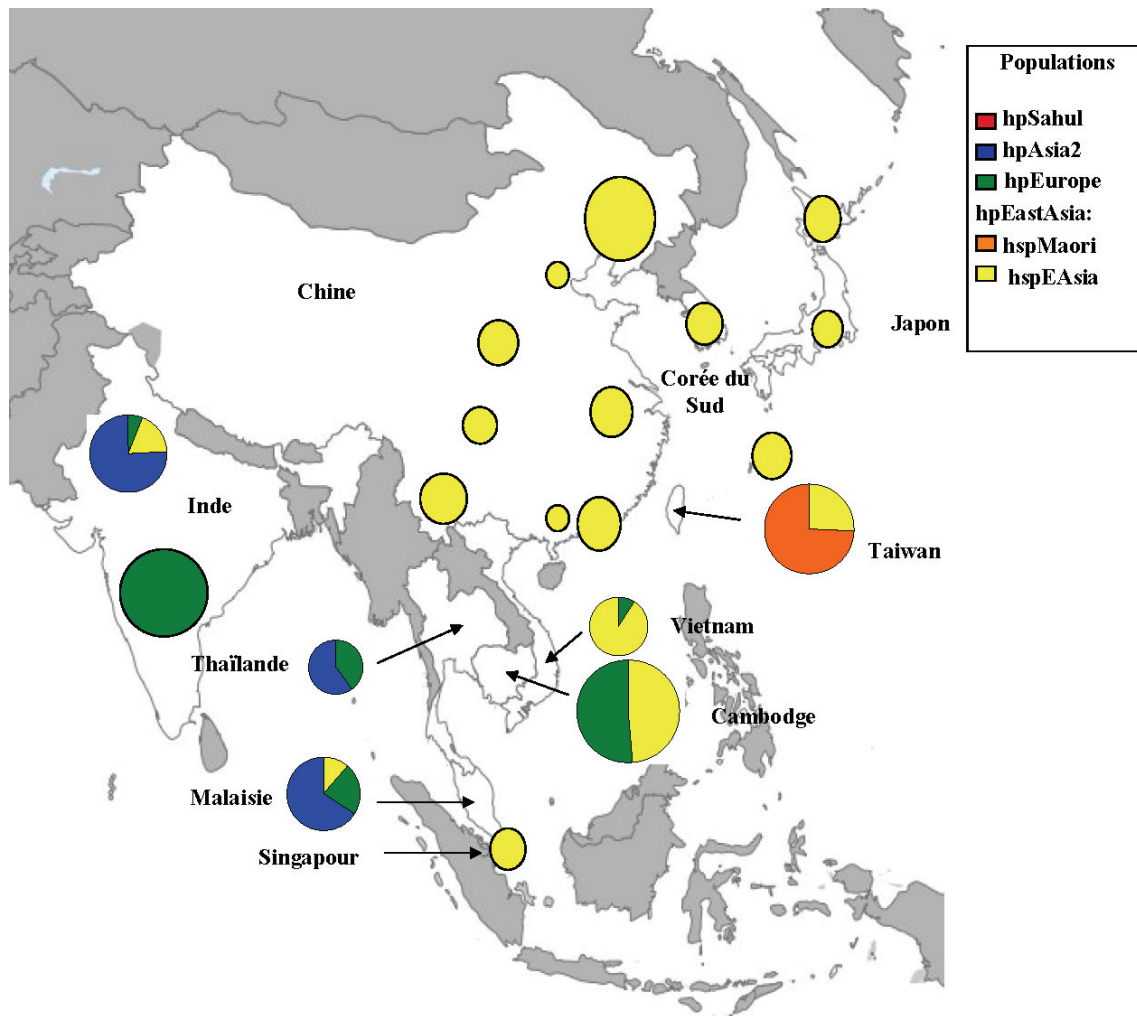
A. Distribution des souches de *H. pylori* selon la (sous-) population génétique associée et l'origine géographique. La taille des cercles est proportionnelle au nombre de souches. Les souches du Vietnam, Cambodge, Thaïlande, Malaisie et Singapour ont été regroupées ensemble.

B. Arbre phylogénétique des souches hspMaori généré par le logiciel CLONAL FRAME (80 % de consensus). Les souches ont été codées par une couleur selon le groupe auquel appartient la langue austronésienne de l'hôte.

Une question reste en suspens sur le mode de colonisation de l'Océanie par des peuples parlant des langues austronésiennes. Deux hypothèses ont été émises: i) une expansion rapide (modèle du « train express ») vers la Mélanésie puis la Polynésie sur la base de la linguistique (Gray and Jordan 2000), ii) une expansion lente supposant des interactions culturelles et génétiques entre peuples des îles mélanésiennes et Austronésiens sur la base de la génétique humaine (Hill *et al.* 2007; Underhill and Kivisild 2007). Le nombre limité d'échantillons bactériens étudiés ne permet pas d'affirmer que les colons en provenance de Taiwan (hspMaori) n'ont pas rencontré en Mélanésie des populations installées depuis bien plus longtemps (hpSahul), entraînant un fort métissage comme le suggère l'étude des marqueurs génétiques humains. Bien que les (sous-) populations génétiques hpSahul et hspMaori n'aient jamais été retrouvées ensemble sur un même site (Figure 23A), elles pourraient encore coexister dans certaines îles du nord de la Mélanésie et sur les côtes de Nouvelle-Guinée. Leur étude pourrait apporter des arguments essentiels pour mieux comprendre les détails de l'histoire humaine dans cette région du monde.

#### **2.1.2.2. Asie du sud-est**

Les populations d'Inde et de Chine, vers le début de notre ère, débordèrent leurs frontières naturelles, venant finalement à la rencontre l'une de l'autre dans la péninsule indochinoise, comme elles le firent, presque au même moment, en Asie centrale. Ces expansions constituent l'un des temps forts de l'histoire de cette région du monde et leurs complexités se reflètent dans la distribution des (sous-) populations génétiques de *H. pylori* en Asie continentale (Figure 24) (Article n°2).



**Figure 24. Populations génétiques de *H. pylori* en Asie continentale, au Japon et à Taiwan**

S. Breurec (Article n°2)

La taille des cercles est proportionnelle au nombre de souches. Les souches hspEAsia isolées de Chinois ayant migré récemment en Thaïlande et en Malaisie n’ont pas été incluses.

#### 2.1.2.2.1. Anciennes migrations venue d’Inde

Les premières sociétés organisées en Asie du sud-est ont en commun d’être des royaumes dits « indianisés », c’est-à-dire sous une forte influence culturelle indienne (Higham 1989). Ces principaux royaumes furent :i) le Royaume Funan, 100 à 550 après JC, sur un territoire correspondant approximativement au sud du Vietnam actuel et au Cambodge actuel, ii) le Royaume Chenla, 550 à 802 après JC, sur un territoire correspondant approximativement au Cambodge actuel, au sud du Laos actuel et du Vietnam actuel, iii) le Royaume d’Angkor, 802 à 1431 après JC, situé selon les périodes sur un territoire

correspondant au Cambodge actuel, au Laos actuel, à la Thaïlande actuelle, au Vietnam actuel et même à la Birmanie et à la Malaisie actuelles. Ces Royaumes n'ont jamais été sous la dépendance d'une métropole indienne. Cependant, ils ont en commun (Higham 1989) l'adoption d'une langue indo-européenne à la cour (sanskrit), autrefois parlée dans le sous-continent indien, l'adoption de religions d'origine indienne (hindouisme pour le Royaume Funan, le shivaïsme pour le Royaume Chenla, l'hindouisme et le bouddhisme pour le Royaume khmer) et une conception de l'État et de la monarchie de type indienne.

L'histoire et les mécanismes de l'indianisation de la péninsule indochinoise sont inconnus, faute de sources historiques dignes de foi. Si les vestiges archéologiques et épigraphiques montrent de façon indubitable que les classes supérieures étaient complètement imprégnées par la culture indienne, l'incertitude demeure immense concernant l'influence exercée par la civilisation indienne sur l'ensemble de la population. L'étude de l'ADNmt humain montre un faible apport génétique d'un point de vue maternel, puisque les haplogroupes retrouvés en Asie de l'ouest ne sont pas retrouvés en Asie du sud-est, suggérant une absence ou un faible métissage entre Indiens et peuples autochtones d'Asie du sud-est (Hill *et al.* 2006; Hill *et al.* 2007; Soares *et al.* 2008). Ainsi, pendant longtemps, on pensa que l'Inde rayonna en Asie du sud-est, essentiellement par ses missionnaires et par ses marchands. Récemment, des récentes excavations archéologiques ont montré la présence de villages d'artisans Indiens probablement originaires du sud de l'Inde, dès le 3<sup>ème</sup> siècle dans la péninsule thaïlandaise. Les contacts entre Indiens et peuples d'Asie du sud-est ont ensuite perduré jusqu'au 15<sup>ème</sup> siècle à travers le commerce et une migration venue d'Inde (Bellina 2002; Bellina-Pryce and Silapanth 2006; Bellina and Silapanth 2006; Bellina 2007).

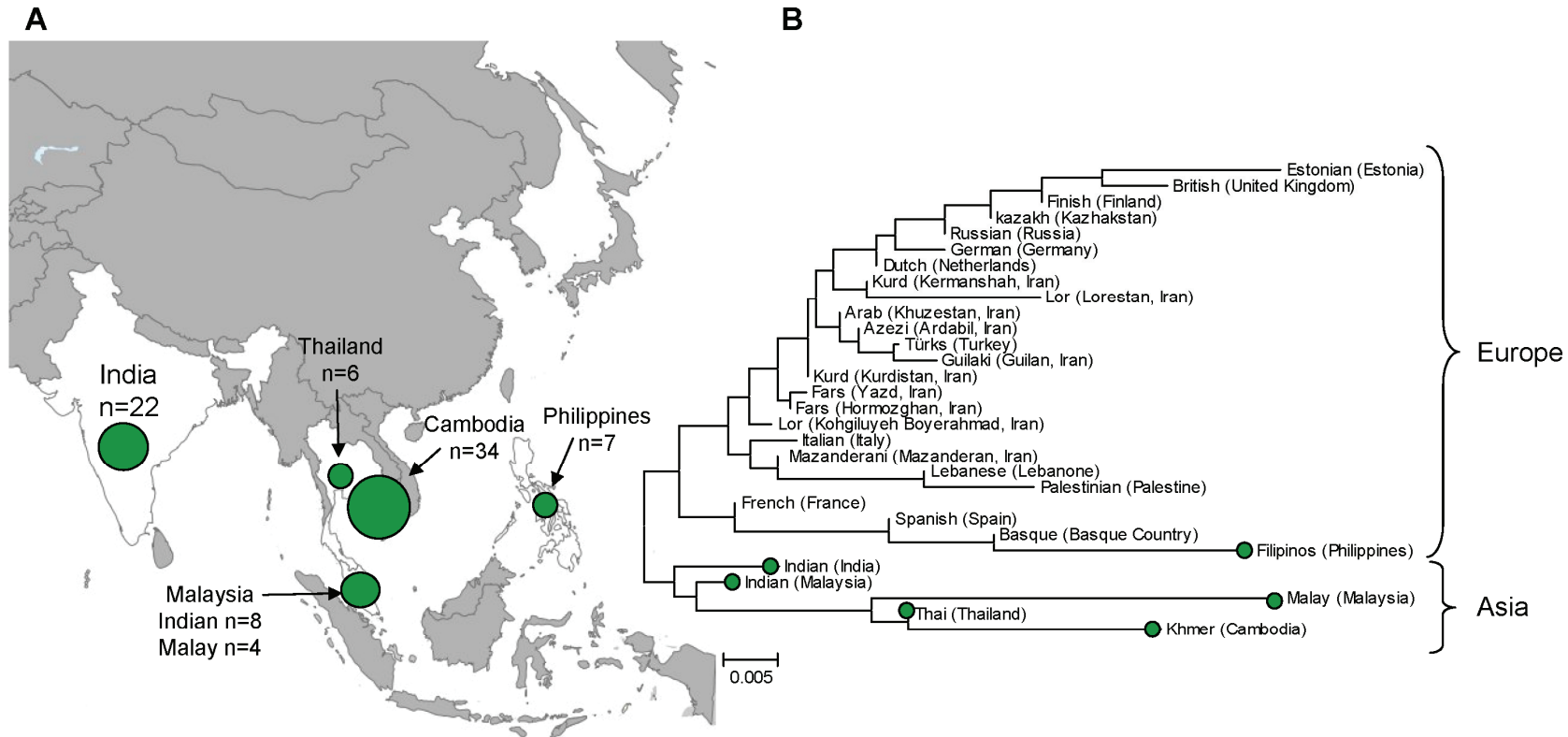
Les données issues de l'étude de la diversité génétique confortent-elles cette hypothèse? Elles recourent de façon surprenante les données archéologiques (Article n°2):

- la population hpEurope est retrouvée à forte fréquence au Cambodge, en Thaïlande et en Malaisie (Figure 24),
- la topologie de l'arbre phylogénétique (Figure 25) montre une migration ancienne d'Indiens, hébergeant *H. pylori* dans leur estomac (population hpEurope), en Malaisie, Thaïlande et Cambodge,
- l'arbre ayant été généré à partir de la matrice Fst, une analyse de type « bootstrap » pour évaluer la robustesse des résultats est impossible. Cependant, l'AMOVA, une alternative, confirme la validité de notre analyse (Article n°2, table S1, p12 de l'article).
- cette migration a laissé une trace plus importante que ne le laisse présager l'étude de l'ADNmt puisque 52 % des souches du Cambodge sont assignées à la population



hpEurope. Cette prévalence est cependant à prendre avec précaution puisqu'une diffusion au sein des populations autochtones des souches hpEurope au détriment des souches hspEAsia est fortement suspectée.

- les Philippines, situés au large des côtes du Vietnam et de l'Indonésie, ont été sous domination espagnole pendant plus de trois siècles (1565-1898). Les isolats de ces îles étaient significativement différenciés de ceux du Cambodge et de Thaïlande, mais pas de ceux issus d'Espagne, indiquant très probablement une introduction récente par les Espagnols (Figure 25).



**Figure 25. La population génétique hpEurope, marqueur d'une migration venue d'Inde**

S. Breurec *et al.* (Article n°2)

A. Distribution des souches hpEurope en Asie.

B. Relations phylogénétiques selon la méthode du Neighbour-Joining à partir de la matrice des Fst (logiciel MEGA 4) des populations hpEurope.

La faible prévalence des souches hpEurope isolées de patients originaires du Vietnam (2/20, 10 %) (Linz, Balloux *et al.* 2007) suggère que la migration venue d'Inde n'a pas concerné le Vietnam. Cependant, les souches étudiées ont été isolées à Hanoï situé dans le nord du Vietnam (Tonkin), région sous domination chinoise (Empire Han) pendant un millénaire à partir du 1<sup>er</sup> siècle après JC. Ce n'est qu'après cette période que les dynasties du nord ont conquis le centre du Vietnam (Annam) puis le sud (Cochinchine). L'étude de la fin de l'extrémité droite du *cag* PAI de souches isolées de patients vietnamiens a montré que 84 % des isolats présentent un motif de type II (Uchida *et al.* 2009), motif caractéristique des souches d'Asie de l'est (hpEastAsia), avec une fréquence de 76 % à Hô-Chi-Minh-Ville dans le sud et de 93 % à Hanoï dans le Nord (Figure 26). En revanche, le motif de type I qui est caractéristique des souches d'Europe (hpEurope) est retrouvé uniquement chez celles isolées à Hô-Chi-Minh-Ville (8/49, 16 %). De façon intéressante, cette ville était avant l'annexion par les Vietnamiens au 17<sup>ème</sup> siècle, le principal port maritime de l'empire khmer connu sous le nom de Prey Nokor. De plus, la Cochinchine a fait parti du royaume du Funan et du Chenla, mais également de l'empire khmer (IX<sup>ème</sup> au XVII<sup>ème</sup> siècle) (Higham 1989). La chaîne montagneuse Annamite, frontière naturelle entre le Cambodge/Laos et le Vietnam (Figure 26), ainsi que l'occupation chinoise, a probablement formé une barrière qui a contenu l'influence indienne dans le Bassin du Mekong, suggérant que les souches hpEurope sont probablement fréquentes chez les habitants de cette région. Des études ultérieures seront nécessaires pour confirmer cette hypothèse.

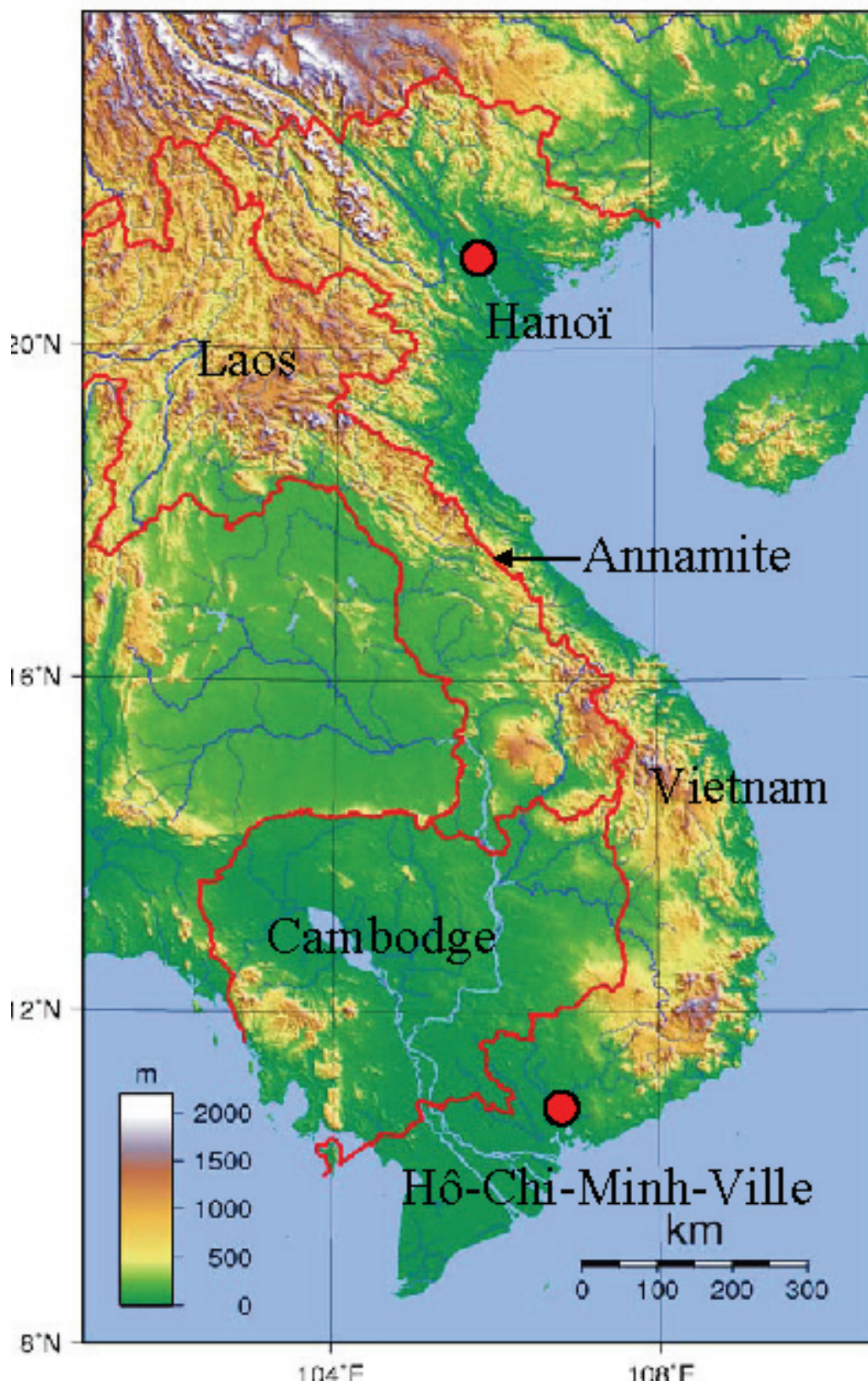


Figure 26. Topographie du Vietnam, Laos et Cambodge

#### 2.1.2.2.2. Anciennes migrations venue de Chine

La distribution fragmentée des peuples parlant les cinq familles de langues majeurs en Asie du sud-est continentale est le reflet de migrations importantes et complexes. Les langues austro-asiatiques, austronésiennes et hmong-mien sont considérées comme les familles de langue les plus anciennes, tandis que les langues tai-kadai et sino-tibétaines sont issues de migrations plus récentes. Actuellement, les langues sino-tibétaines sont les plus répandues en Asie puisque ses locuteurs sont retrouvés en Chine, au Népal, au Bhoutan, en Inde, au Pakistan, au Bangladesh, en Birmanie, en Thaïlande, au Vietnam et au Laos. Les autres familles ont des distributions fragmentées et forment des îlots linguistiques entourés par un océan de locuteurs de langues sino-tibétaines. Pourquoi cette distribution ? La fragmentation géographique de ces langues n'est pas l'effet d'un saupoudrage à travers le paysage asiatique et on peut imaginer qu'il y a eu jadis une distribution continue. Elle est principalement la conséquence de l'expansion de locuteurs de la famille sino-tibétaine depuis 3000 ans à partir du sud de la Chine. Ils ont inondé par vagues successives les populations autochtones issues des migrations antérieures. L'histoire de la dynastie chinoise Zhou de 1100 à 221 avant JC, retrace la conquête et l'absorption de l'immense majorité de cette population de langues non chinoises par des Etats sinophones (Higham 1989).

La famille des langues austro-asiatiques est un groupe linguistique, dont les langues les plus largement parlées sont le khmer (Cambodge) et le vietnamien (Vietnam). Ses 60 millions de locuteurs sont dispersés du Vietnam à l'est, jusqu'à la péninsule malaise au sud et au nord de l'Inde à l'ouest. L'origine géographique des peuples parlant les ancêtres des langues austro-asiatiques est le sujet de débats au sein de la communauté scientifique. Sur la base de données linguistiques, archéologiques et génétiques, trois hypothèses ont été émises sur leur origine ancestrale: i) l'est de l'Inde ou la baie du Bengale, iii) l'Asie du sud-est, ii) le centre ou le sud de la Chine (Jin *et al.* 2001; Sagart *et al.* 2005). Nos résultats sont en faveur d'une migration à partir du sud de la Chine, ce qui est en concordance avec l'hypothèse selon laquelle des agriculteurs maîtrisant la culture du riz et parlant les ancêtres des langues austro-asiatiques auraient migrés de la Chine Centrale (vallée du Yangtsé) vers l'Asie du sud-est il y a 4000 ans, transportant *H. pylori* dans leur estomac (sous-population hspEAsia). Cette théorie a été développée notamment par Charles Higham sur la base des données archéologiques (Higham 1989). Cependant, la topologie de l'arbre phylogénétique générée à partir des valeurs des indices de diversification  $F_{st}$  des populations hspEAsia de *H. Pylori* manque de robustesse pour affirmer l'origine précise (Article n°2, Figure 2C, p5 de l'article). Les expansions chinoises ont très probablement brouillé le signal phylogénétique.

Les langues tai-kadai sont une famille de langues parlées en Asie du sud-est continentale, principalement en Thaïlande et au sud de la Chine. Les Thaïs ont migré du sud de la Chine (Province du Yunnan) vers l'actuelle Thaïlande au début du second millénaire après JC selon les données archéologiques suite à des pressions politiques et économiques, notamment liées à l'arrivée des Mongols qui envahirent le Yunnan et sa capitale au 13<sup>ème</sup> siècle (Eberhard 1977; Higham 1989; Lertrit *et al.* 2008). La population génétique hpAsia 2 est le marqueur de cette migration :

- elle est retrouvée à forte fréquence chez les souches isolées de patients d'ethnie thaï,
- elle est absente chez les souches isolées de patients khmers et vietnamiens, ce qui exclut une introduction ancienne par les Indiens,
- la topologie de l'arbre phylogénétique exclue une introduction récente par les Indiens (Article n°2, Figure 3B, p7 de l'article).

Pour la première fois, nous avons pu retracer les migrations humaines à l'intérieur d'une région géographique aussi restreinte que l'Asie du sud-est, ce qui ouvre des perspectives intéressantes dans d'autres régions du monde où vivent des populations autochtones (Sibérie, Mongolie, Chine, Japon...). Avant cette étude, *H. pylori* avait été uniquement le témoin de migrations historiques intercontinentales ou intracontinentales à large échelle. En Europe, seule zone du globe en dehors de l'Asie pour laquelle suffisamment d'isolats sont disponibles pour une analyse régionale, la complexité de son peuplement avec plusieurs vagues de migration historiques (Semino *et al.* 2000; Chikhi *et al.* 2002) a brouillé la signature génétique de *H. pylori*.

### **2.1.3. Problèmes à résoudre**

#### **2.1.3.1. Dynamique de diffusion différente selon la population génétique**

L'étude de l'ADNmt de populations d'Asie du sud-est montre une absence d'haplogroupes d'Asie de l'ouest, suggérant que la migration ancienne d'Indiens en Asie du sud-est n'explique pas à elle seule la forte fréquence des souches hpEurope en Asie du sud-est (Article n°2). Cette forte prévalence contraste avec celle des souches hspEasia chez les Thaïs (aucune souche) et les Malais (2 sur les 15 souches incluses).

L'étude de l'écologie des plantes a apporté une contribution majeure dans l'élaboration de théories sur l'adaptation d'une espèce à son environnement (Sultan 2001 ;

Richards *et al.* 2005). La capacité de survie et de reproduction dans des conditions environnementales différentes favorise le succès d'envahissement des espèces possédant des génotypes dits « généralistes ». Cette plasticité permet aux organismes d'exprimer des phénotypes avantageux dans une plus large gamme d'environnements et de minimiser les risques d'extinction locale de ses populations en cas de changements environnementaux.

En appliquant ces théories à *H. pylori*, les souches hpEurope sont des souches dites « généralistes », c'est-à-dire qu'elles possèdent la diversité génétique suffisante pour coloniser un large panel d'hôtes, au contraire des souches hspEAsia qui sont des souches dites « spécialistes ». Le résultat est une diffusion au sein des populations autochtones des souches hpEurope au détriment des souches hspEAsia et ce par deux mécanismes: i) compétition durant la colonisation de l'estomac avec dans la situation la plus extrême, une élimination de la souche « spécialiste » par la souche « généraliste », ii) subversion, la souche hspEAsia recombinante devenant de plus en plus « européenne » par acquisition d'ADN exogène de souches hpEurope. Ces dernières sont attendues comme étant génétiquement plus diverses mais également plus « généralistes ». Un tel phénomène avait déjà été mis en évidence en Amérique du sud avec le déplacement de hspAmerind par hpEurope chez des Amérindiens (Dominguez-Bello *et al.* 2008; Kersulyte *et al.* 2011).

La dynamique de diffusion des souches de *H. pylori* au sein d'une population varie selon la population génétique bactérienne, reflétant des capacités différentes de transmission, d'adaptation et de colonisation de l'estomac et perturbant le signal phylogénétique. Il serait intéressant de séquencer complètement le génome d'une ou plusieurs souches hspEAsia recombinantes afin de déterminer la distribution chromosomique des acquisitions d'ADN provenant de souches hpEurope, ainsi que la fréquence selon la catégorie fonctionnelle des gènes impliqués. Les données issues de ce travail permettraient de mieux comprendre la coévolution pendant des millénaires des souches hspEAsia et hpEurope.

### **2.1.3.2. Accélération récente des flux migratoires**

L'accélération récente des flux migratoires avec le développement des moyens de communication influence fortement la distribution de la diversité génétique de *H. pylori* de part le monde, en favorisant le transfert horizontal de souches entre groupes ethniques différents et les recombinaisons entre clones non apparentés de *H. pylori* lors d'infections mixtes (Linz *et al.* 2007) (Articles n°2 et n°3). Tous ces éléments ont commencé à brouiller les frontières génétiques ancestrales de la bactérie.

### **2.1.3.3. Diminution de la prévalence de l'infection**

L'augmentation rapide du niveau socio-économique de certaines sous-populations des pays en voie de développement et l'utilisation de plus en plus répandue des antibiotiques contribue à la diminution de la prévalence de l'infection. Cette baisse devrait continuer dans le futur, conséquence de l'amélioration des conditions de vie dans le monde. En conséquence, l'étude de la diversité génétique de *H. pylori* a beaucoup à apporter dans la connaissance des migrations humaines mais ces études doivent être réalisées avant qu'il ne soit trop tard.

### **2.1.3.4. Difficultés pour obtenir de l'ADN de *H. pylori***

Les biopsies gastriques sont obtenues au moyen d'un geste invasif réalisé lors d'une endoscopie. De plus, la culture bactériologique à partir de laquelle l'ADN sera extrait est délicate puisque réalisée sur un milieu enrichi contenant des antibiotiques avec une incubation de 7 jours à 37°C en atmosphère microaérophile.

## **2.2. *H. pylori* et le cancer gastrique au Sénégal et au Cambodge**

*H. pylori* a été classée parmi les agents carcinogènes de classe I par l'Organisation Mondiale de la Santé. Actuellement, avec approximativement un million de nouveaux cas diagnostiqués chaque année, le CG se situe au quatrième rang des cancers dans le monde et est la deuxième cause de décès parmi les cancers, avec environ 800 000 morts par an. Son pronostic est mauvais avec une survie de 10 à 15 % à 5 ans (Parkin *et al.* 2005). Alors que l'incidence globale et la mortalité par CG sont en diminution dans les pays développés, son incidence parmi les lésions malignes s'accroît dans les pays en développement. Le Centre International de Recherche sur le Cancer a mis en ligne en 2010 GLOBOCAN (Globocan 2010). Cette base de données fournit les estimations mondiales du nombre de nouveaux cas de cancer et de décès par cancer pour 2008.

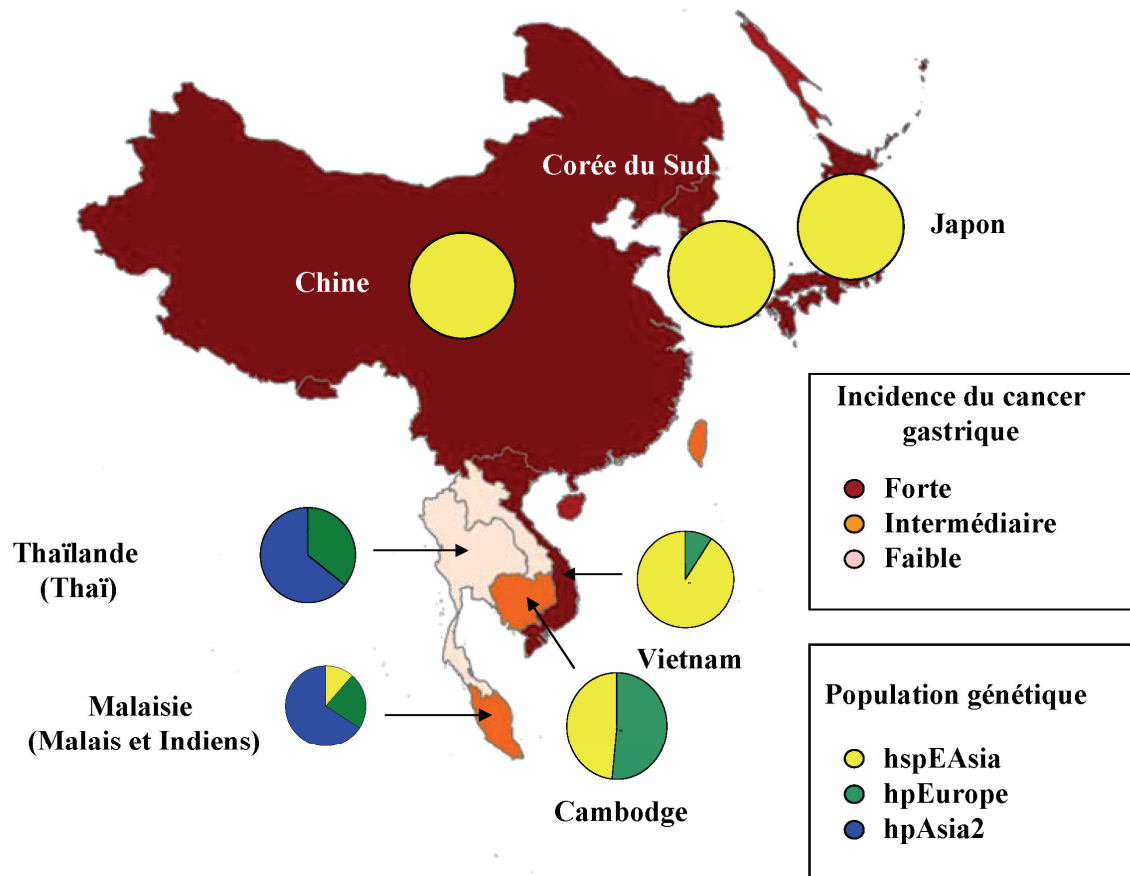
Pour les CG de type sporadique (qui constitue environ 98 % de tous les CG), il a été démontré que les sujets infectés par *H. pylori* ont un risque accru de développer un CG. La réponse inflammatoire de l'hôte serait un élément clef du processus carcinogène et serait influencée par les facteurs de prédisposition génétique de l'hôte, les facteurs environnementaux et les facteurs de virulence de la bactérie. Nous ne nous sommes intéressés qu'aux facteurs de virulence de la bactérie. Les facteurs de prédisposition génétique de l'hôte seront explorés ultérieurement dans le cadre d'une collaboration avec l'équipe de José Machado (IPATIMUP, Medical Faculty of Porto), spécialiste de la thématique. Les difficultés inhérentes à l'évaluation de l'impact des facteurs environnementaux sur la survenue du CG ne



nous ont pas permis de les étudier (incidence faible du CG, disparition de la bactérie fréquente au stade d'atrophie sévère ou de métaplasie intestinale). Pourtant, une telle étude serait intéressante en raison de l'absence de données dans les pays en voie de développement et le contexte socioculturel, ethnique et religieux très différent du Cambodge et du Sénégal.

Nous avons essayé de comprendre la contribution de deux facteurs de virulence majeurs de *H. pylori*, la protéine CagA et la cytotoxine VacA, aux différences géographiques d'incidence du CG en étudiant des souches provenant du Cambodge et du Sénégal. A l'exception du Laos et de la Birmanie, le Cambodge est le seul pays d'Asie pour lequel aucune donnée sur les facteurs de virulence de *H. pylori* n'est disponible, ce qui permet d'avoir une vision d'ensemble sur ce continent. En Afrique, très peu de données existent, la seule étude impliquant un nombre suffisant de CG ayant eu lieu en Afrique du Sud (Argent *et al.* 2004).

Les incidences de CG les plus élevées dans le monde sont observées en Asie, mais des variations de l'incidence existent sur ce continent avec une diminution vers les latitudes plus au sud (Globocan 2010). Cette distribution ne s'explique pas par des différences de prévalence de l'infection à *H. pylori* (Fock and Ang 2010). La Mongolie, le Japon, la Corée du sud, la Chine et le Vietnam sont classés parmi les pays à risque élevé de CG (taux d'incidence standardisé sur l'âge (TIS) 18.9 à 41.4/100 000). Le Cambodge est classé comme un pays avec un risque intermédiaire (TIS 9.8/100 000), tout comme la Birmanie (TIS 11.0/100 000), la Malaisie (TIS 8.4/100 000) et Singapour (TIS 10.9/100 000). Le risque est faible au Bangladesh, au Laos, en Thaïlande et en Inde (TIS 3.5 à 5.2/100 000). Bien que la survenue d'un CG soit le résultat d'un processus complexe, la distribution des (sous-) populations génétiques de *H. pylori* semble fortement corrélée aux variations régionales de l'incidence du CG en Asie (Figure 27) (Article n°3).



**Figure 27. Populations génétiques de *H. pylori* et incidence du cancer gastrique en Asie de l'est et du sud-est**

S. Breurec *et al.*, article n°3.

Les pays ont été classés en risque faible, intermédiaire et fort selon le risque de développer un cancer gastrique à partir des taux d'incidence standardisés sur l'âge, issus de la base de données GLOBOCAN.

En Afrique, le risque de CG est parmi le plus faible au monde, sauf au Mali avec un TIS pour 100 000 habitants de 20.3, mais ce taux est à prendre avec précaution. L'énigme africaine décrite par certains auteurs associe une forte prévalence de l'infection par *H. pylori* à une faible incidence des pathologies gastro-duodénales sévères liée à l'infection (Campbell *et al.* 2001 ; Cataldo *et al.* 2004). Certains auteurs ont évoqué une réponse immunitaire principalement de type Th2 chez les africains pour l'expliquer (cf *Partie I. Généralités : chapitre 5.3.1.3. L'inflammation*) (Mitchell *et al.* 2002). Les parasitoses intestinales (helminthes) ont été avancées comme agent causal puisqu'elles sont endémiques en Afrique et qu'elles induisent une polarisation des lymphocytes T auxiliaires naïfs vers une réponse de

type Th2. Cependant, les helminthiases sont également endémiques dans certaines régions d'Asie ou une incidence forte du CG est observée (Peng *et al.* 1998; Jex *et al.* 2011). Un terrain génétique particulier chez certaines populations africaines pourrait expliquer une réponse immunitaire physiologique de type Th2 plus importante par rapport à d'autres populations. Des études ultérieures sont nécessaires. Une revue de la littérature réalisée sur l'ensemble des études endoscopiques prospectives en Afrique suggère que la survenue de manifestations cliniques graves associées à la présence de *H. pylori* est sous-estimée (Nonaka *et al.* 2005). Le mythe de l'énigme africaine serait dû en partie à la constitution d'échantillons biaisés non représentatifs de la population générale, dans des pays où l'accès aux soins reste difficile et l'espérance de vie relativement courte. Les pays africains, dont la majeure partie des souches est assignée à la population hpAfrica1 (Article n°4), se classent probablement parmi les pays à faible risque de CG. Cependant, le risque n'est probablement pas aussi bas qu'on ne l'a longtemps pensé.

Ainsi le fond génétique des souches pourrait être un marqueur de la virulence de la bactérie bien que des études ultérieures soient nécessaires pour confirmer cette hypothèse.

### **2.2.1. Le gène *cagA***

#### **2.2.1.1. Le statut *cagA* et le cancer gastrique**

Les souches de *H. pylori* sont classées en deux catégories : celles qui possèdent un *cag* PAI complet et fonctionnel et celles qui en sont totalement ou partiellement dépourvues (Covacci, Telford *et al.* 1999). L'étude de la fonctionnalité du *cag* PAI avec la mise en évidence *in vitro* de la sécrétion de la protéine CagA n'a pas été investiguée pour des raisons techniques. Cependant, nous avons pu montrer pour la première fois pour des souches isolées dans un pays d'Afrique que la présence du gène *cagA* était significativement associée au CG (Article n°4) comme dans beaucoup de pays occidentaux (Huang *et al.* 2003). Cette corrélation n'était toutefois pas retrouvée chez des souches isolées de patients cambodgiens et de façon plus large d'Asie de l'est, où quasiment toutes les souches possédaient le gène *cagA* qu'elles soient ou non associées à une pathologie sévère (Article n°3). L'incidence du CG élevée en Asie de l'est ne peut donc être expliquée par la présence ou l'absence du gène *cagA* seul.

### 2.2.1.2. L'extrémité 3' du gène *cagA*, un facteur clef

Des études épidémiologiques, chez l'animal et *in vitro* (cf paragraphe 5.3.1.2. *L'Ilot de pathogénicité cag et la protéine CagA* dans la partie Généralités) suggèrent fortement que les souches présentant un « East Asian CagA » sont plus carcinogènes que celles présentant un « Western CagA », et qu'à l'intérieur de ces dernières, plus le nombre de motifs EPIYA-C est important, plus la souche possède un potentiel carcinogène important.

Il pourrait donc être tentant de vouloir expliquer les différences d'incidence de CG observées dans le monde par la distribution des souches possédant un « East Asian CagA ». Cependant, les incidences de CG peuvent être élevées dans certaines régions où circulent des souches possédant un « Western CagA », montrant la complexité du processus carcinogène. A une échelle plus limitée comme l'Asie de l'est et du sud-est où la diversité génétique humaine est plus limitée, la distribution des TIS correspond parfaitement à la distribution des souches possédant un « East Asian CagA » à quelques incohérences près (Article n°3) (Tableau 3). Si on prend le cas du Vietnam et du Cambodge, deux pays limitrophes dont la structure génétique de la population humaine est proche puisque issue de la même migration, celles des ancêtres des locuteurs des langues austro-asiatiques (Article n°2), l'incidence du CG au Cambodge est environ la moitié de celle du Vietnam (9.8 versus 18.9 pour 100 000 habitants), ce qui correspond à la distribution des souches possédant un « East Asian CagA » (41 % versus 97 %) (Uchida *et al.* 2009). La faible prévalence de ces souches au Cambodge et par extension en Asie du sud-est, à l'exception du Vietnam, pourrait expliquer l'incidence du CG plus faible que celle des pays d'Asie de l'est.

**Tableau 3. Fréquence du type de protéine CagA en Asie de l'est et du sud-est et incidence du cancer gastrique**

Région /Pays	Groupe ethnique	Total <sup>a</sup> N	cagA positive N (%)	Type de CagA		TIS <sup>b</sup>	Référence
				Western N (%)	East Asian N (%)		
<b>Asie de l'est</b>							
Corée du sud	Coréen	108	108 (100)	21 (19)	87 (81)	41.4	(Jones <i>et al.</i> 2009)
Japon	Japonais	78	72 (92)	8 (11)	64 (89)	31.1	(Azuma <i>et al.</i> 2004)
Chine	Chinois	115	89 (77)	15 (17)	74 (83)	29.9	(Li <i>et al.</i> 2009)
<b>Asie du sud-est</b>							
Vietnam	Vietnamien	103	98 (95)	4 (4)	94 (96)	18.9	(Uchida <i>et al.</i> 2009)
Cambodge	Khmer	61	57 (93)	34 (60)	23 (40)	9.8	<i>S. breurec et al</i> Article n°3
Malaisie						8.4	(Mohamed <i>et al.</i> 2009;
	Malais	25 <sup>c</sup>	25	12 (48)	13 (52)		Schmidt <i>et al.</i>
	Indien	14 <sup>c</sup>	14	10 (71)	4 (28)		2009)
	Chinois	67 <sup>c</sup>	67	9 (13)	58 (86)		
Thaïlande	Thaï	179	78 (44)	73 (92)	6 (8)	3.5	(Hirai <i>et al.</i> 2010)

<sup>a</sup> Les souches associées à un CG ou un ulcère peptique n'ont pas été incluses sauf pour le Vietnam et la Malaisie où les données de l'article étaient insuffisantes pour les extraire.

<sup>b</sup> Taux d'incidence standardisé sur l'âge du CG pour 100 000 habitants obtenu à partir de la base de données GLOBOCAN

<sup>c</sup> Seuls les souches *cagA* positives avaient été incluses dans l'article

Au Sénégal, l'extrémité 3' du gène *cagA* semble jouer un rôle déterminant dans la survenue du CG (Article n°4). La présence de deux motifs EPIYA-C était significativement associée à la présence d'un CG, comme démontré récemment par une méta-analyse de l'ensemble des données de la littérature (Xia *et al.* 2009). Bien que la majorité des patients aient des lésions de métaplasie intestinale, la faible proportion de souches avec 2 motifs EPIYA-C (2.6 %) par rapport aux autres continents (11.5 à 30 %) est un des résultats les plus importants de notre étude (Devi *et al.* 2006; Panayotopoulou *et al.* 2007; Reyes-Leon *et al.* 2007; Basso *et al.* 2008; Xia *et al.* 2009). L'analyse du polymorphisme génétique de 7 gènes de ménage selon la méthodologie décrite par Falush *et al.* (Falush *et al.* 2003) de 79 souches choisies au hasard a montré que ces souches étaient assignées à la population génétique hpAfrica1, sous-population hspWAfrica (Article n°4). Cette faible proportion de souches présentant plusieurs motifs EPIYA-C pourrait être une des particularités de cette population génétique. Dans la seule autre étude réalisée en Afrique (Afrique du Sud), 27 % des souches

contenaient 2 motifs EPIYA-C ou plus mais cette différence pourrait être due à la forte fréquence de souches associées au CG (45 %) (Argent, Kidd *et al.* 2004). La faible prévalence des souches avec de multiples motifs EPIYA-C pourrait contribuer à la faible incidence du CG au Sénégal, et par extension en Afrique, en raison de la distribution de la population génétique hpAfrica1.

L'extrémité 3' du gène *cagA* semble jouer un rôle clef dans la survenue du CG et pourrait expliquer au moins partiellement les différences d'incidence de CG observées en Asie du sud-est et de l'est, et en Afrique. Des études ultérieures seront nécessaires pour confirmer cette hypothèse.

### **2.2.2. La cytotoxine vacuolisante VacA et le cancer gastrique**

Le mosaïcisme du gène *vacA* se traduit, *in vitro*, par des niveaux d'activités différents de la toxine. Les souches de type s1/i1/m1-ou-m2 induisent en général un haut niveau de vacuolisation tandis que les souches s1/i2/m2 induisent peu de vacuolisation, et les souches s2/i2-ou-i1/m1-ou-m2 sont dépourvues d'effet cytotoxique (Rhead *et al.* 2007). Il n'est donc pas étonnant que le type s1 ou m1 ait été associé à un risque accru de développer un CG ou un ulcère peptique en Europe (Atherton, Cao *et al.* 1995; van Doorn, Figueiredo *et al.* 1998), au Proche-Orient (Sugimoto *et al.* 2009), en Amérique Latine mais également en Afrique (Sugimoto and Yamaoka 2009). Notre étude au Sénégal confirme l'association entre le type s1 et le CG, le type m1 n'étant cependant pas associé au CG, probablement à cause du faible nombre de cas de notre étude (Article n°4). Au Cambodge et de façon générale en Asie de l'est, le type s1 n'est pas associé à un risque accru de développer une pathologie gastro-duodénale grave puisque la majeure partie des souches est de type s1 (Yamaoka, Kodama *et al.* 1999; Yamaoka, Orito *et al.* 2002).

En Asie de l'est et du sud-est, l'incidence du CG présente un gradient nord-sud qui correspond parfaitement à celui de la fréquence des souches de type m1 (Tableau 4). Cette corrélation n'est cependant valable que pour les régions géographiques où les souches hspEAsia sont endémiques, suggérant que les souches hspEAsia de type m1 pourraient avoir une virulence supérieure. La fréquence de l'allèle *vacA* m1 (58.6 %) chez les souches hspEAsia du Cambodge est similaire à celle retrouvée à Hanoi dans le nord du Vietnam mais inférieure à celles retrouvées au Japon ou en Corée (Tableau 4).

**Tableau 4. Incidence du cancer gastrique et génotype *vacA* m1 et m2 dans les pays d'Asie où les souches *hspEAsia* sont endémiques**

Région géographique/Pays	N <sup>a</sup>	génotype <i>vacA</i>		Total	TIS <sup>b</sup>		Référence
		m1 N (%)	m2 N (%)		Homme	Femme	
<b>Asie de l'est</b>							
Corée du sud	87	83 (95.4)	4 (4.6)	41.4	62.2	24.6	
Japon	83	79 (95.2))	4 (4.8)	31.1	46.8	18.2	(Yamaoka
Chine	27	18 (66.7)	9 (33.3)	29.9	41.3	18.5	2010)
Taiwan	28	9 (32.1)	19 (67.9)	10.4	13.5	7.6	
<b>Asie du sud-est<sup>c</sup></b>							
Vietnam	78 <sup>d</sup>	29 (37.1)	45 (57.9)	18.9	24.4	14.6	(Uchida <i>et al.</i>
Vietnam (Hanoi)	38	19 (50.0)	19 (50.0)	-	27	13.2	2009)
Vietnam (Ho Chi Minh)	40	10 (25.0)	30 (75.0)	-	18.7	8.1	

<sup>a</sup> Les souches associées à un CG ou un ulcère peptique n'ont pas été incluses.

<sup>b</sup> Taux d'incidence standardisé sur l'âge du CG pour 100 000 habitants obtenu à partir de la base de données GLOBOCAN (Globocan 2010) sauf pour Hanoi et Ho Chi Minh, obtenu à partir de la base de données « Cancer Incidence in Five Continents » (Curado *et al.* 2008).

<sup>c</sup> Les souches du Cambodge n'ont pas été incluses puisque seulement 48 % des souches sont assignées à la sous-population *hspEAsia* (Article n°3).

<sup>d</sup> Pour 4 souches du Vietnam, la région m n'a pu être typée.

Le type i1 était associé significativement au CG chez des patients iraniens (Rhead, Letley *et al.* 2007) et était prédictif de l'ulcère peptique chez des patients italiens (Basso *et al.* 2008), tandis qu'en Asie, il n'était pas associé à la survenue de pathologies gastro-duodénales graves (Ogiwara *et al.* 2008). Au Sénégal, il n'est pas un facteur de risque de CG (Article n°4), tout comme au Cambodge, où la majeure partie des souches était de type i1 (93.9 %). Ces données montrent la difficulté d'extrapoler un résultat d'une région géographique à une autre.

### 2.2.3. Diffusion d'allèles *cagA* et *vacA* des souches européennes vers les souches d'Asie de l'est

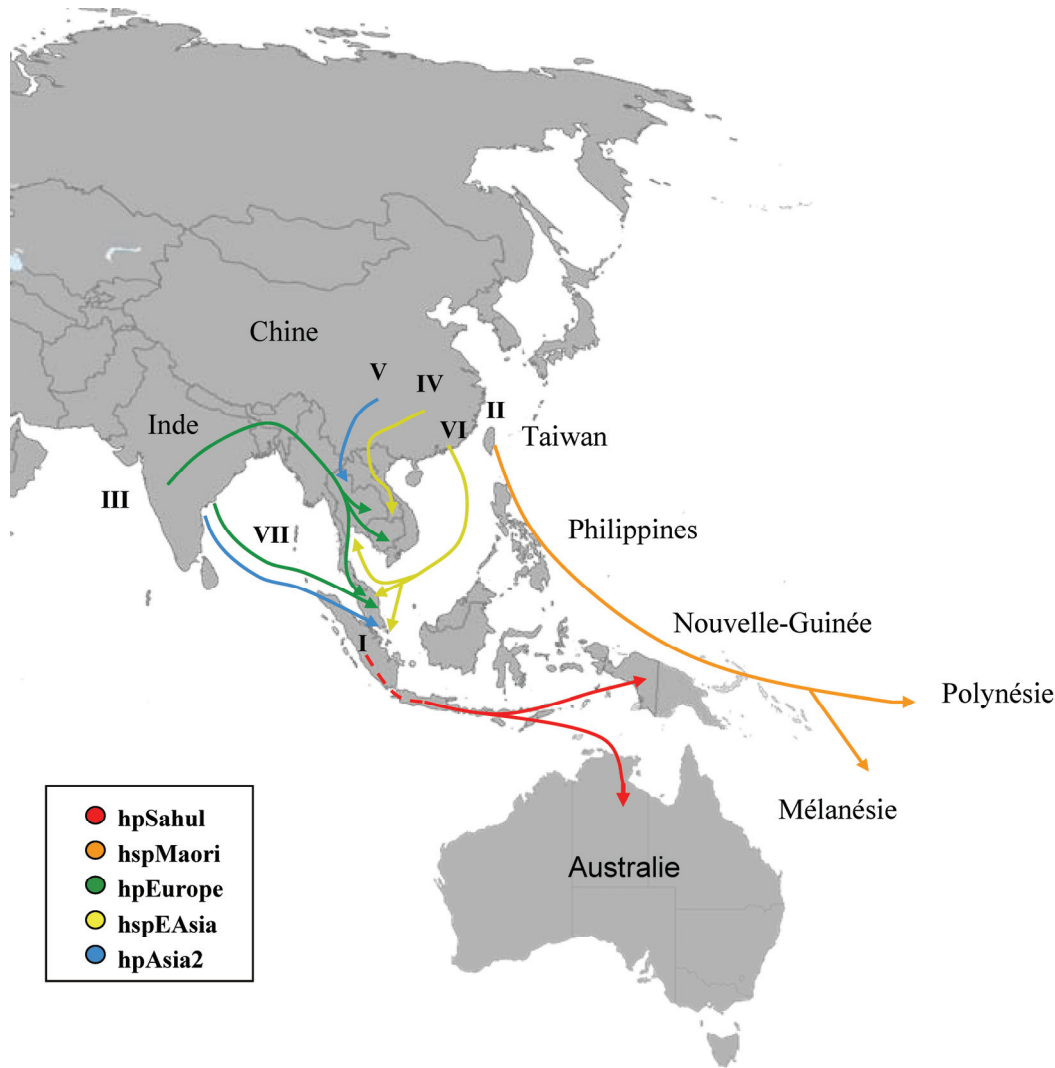
CagA-P active la voie des MAP kinases impliquée dans le contrôle de nombreux processus fondamentaux tels que la prolifération, la différenciation et la mobilité cellulaire (Higashi *et al.* 2002). Un mutant  $\Delta vacA$  augmente significativement l'allongement en « bec de colibri » des cellules AGS induit par la protéine CagA-P (Argent *et al.* 2008). De récents travaux montrent que CagA stimule la translocation nucléaire d'un facteur de transcription, le

facteur nucléaire des lymphocytes T activés (NFAT pour « Nuclear factor of activated T-cells ») qui va induire la transcription de gènes cibles tels que *p21*. Un taux élevé de *p21* pourrait induire le développement d'une métaplasie intestinale, un stade pré-néoplasique. VacA inhibe la translocation nucléaire de NFAT induite par CagA et donc l'induction de la synthèse de *p21* (Yokoyama *et al.* 2005). De plus, CagA inhibe l'apoptose induite par VacA par deux mécanismes complémentaires: i) CagA-P empêche VacA d'atteindre les compartiments endosomaux tardifs, ii) CagA sous forme non-phosphorylée bloque l'apoptose induite par VacA au niveau mitochondrial (Oldani *et al.* 2009). Il a été montré également que des mutants  $\Delta cagA$  et  $\Delta cagE$  augmentaient significativement la vacuolisation des cellules épithéliales induites par VacA (Argent *et al.* 2008). Les protéines VacA et CagA exercent donc un contrôle négatif l'une sur l'autre, modulant leurs activités cellulaires et évitant des dommages excessifs. La diffusion d'allèles *cagA* et *vacA* des souches « européennes » (hpEurope) vers celles d'Asie de l'est (hspEasia) par recombinaison (Article n°3) pourrait entraîner la rupture de l'équilibre entre les effets biologiques de CagA et de VacA et provoquer une dérégulation, ce qui pourrait avoir des conséquences sur l'évolution de la maladie.

### 2.3. Conclusion

*H. pylori* est le « fidèle compagnon » de l'homme depuis au moins 60 000 ans puisque cette bactérie accompagne l'homme dans ses migrations depuis sa sortie d'Afrique. Cette coévolution entre l'homme et *H. pylori*, couplée à l'exceptionnelle diversité génétique de cette bactérie, permet de retracer de nombreuses migrations anciennes ou plus modernes. Les études que nous avons menées en Asie du sud-est et en Océanie ont apporté des détails sur les migrations humaines dans ces régions du monde (Figure 28). Elles confirment la résolution plus élevée de ce marqueur bactérien par comparaison aux marqueurs génétiques humains traditionnels, malgré des échantillonnages beaucoup plus faibles en nombre et en diversité. Cependant, les frontières génétiques ancestrales de la bactérie sont en train de se brouiller. *H. pylori* a encore beaucoup à apporter dans la connaissance de l'histoire des migrations humaines mais ces études doivent être réalisées avant qu'il ne soit trop tard. Depuis plusieurs années, l'amélioration constante des techniques de séquençage a permis de rendre accessible le séquençage complet de génomes à de nombreux laboratoires. Cette révolution devrait se poursuivre dans le futur et l'énorme masse de données générées permettra une meilleure compréhension des migrations humaines d'un point de vue humain mais également bactérien.





**Figure 28. Migrations humaines décrites à partir du polymorphisme génétique de *H. pylori***

S. Breurec *et al.* (Article n°1 et Article n°2)

I. Migration préhistorique de l'Asie vers le Sahul (population hpSahul). II. Migration ancienne des ancêtres des Austronésiens à partir de Taiwan vers la Mélanésie puis la Polynésie (sous-population génétique hspMaori), il y a 5000 ans. III. Migration ancienne d'Indiens vers l'Asie du sud-est principalement entre le 3<sup>ème</sup> et le 15<sup>ème</sup> siècle (population génétique hpEurope). IV. Migration ancienne des ancêtres des Austro-Asiatiques probablement du centre de la Chine vers le Vietnam et le Cambodge avec un peuplement de cette région il y a environ 4000 ans (sous-population génétique hspEAsia). V. Migration ancienne des ancêtres des Thaïs du sud de la Chine vers l'actuelle Thaïlande au début du second millénaire après JC (population génétique hpAsia2). VI. Migration récente de Chinois en Thaïlande, Malaisie et Singapour (sous-population génétique hspEAsia) à partir des provinces du Guangdong et du Fujian. VII. Migration récente d'Indiens en Malaisie (populations génétique hspAsia2 et hpEurope).

La connaissance de la distribution des populations génétiques bactériennes est très importante puisqu'elle semble fortement influencer l'incidence des CG en Asie et en Afrique. En Asie, les souches hspEAsia possédant un « East Asian CagA » et un allèle *vacA* de type m1 semblent plus virulentes que les autres. La faible fréquence des souches possédant l'ensemble des caractéristiques génétiques précédentes au Cambodge (15.2 %), comparé à celles retrouvées au Vietnam et dans les pays d'Asie de l'est, pourrait expliquer l'incidence intermédiaire du CG dans ce pays. Une introgression fréquente d'allèles *cagA* et *vacA* européens (hpEurope) dans des souches d'Asie de l'est (hspEAsia) a été observée. CagA et VacA ayant des effets antagonistes, cette expansion pourrait entraîner la rupture de l'équilibre entre les effets biologiques de ces deux protéines et provoquer une dérégulation, responsable de conséquences graves sur l'évolution de la maladie. L'incidence du CG pourrait augmenter dans les années futures au Cambodge. Au Sénégal, la présence du gène *cagA*, de l'allèle *vacA* s1 et de deux motifs EPIYA-C étaient significativement associées au CG. Les isolats se caractérisaient par la faible prévalence de multiples segments EPIYA-C par rapport à celle retrouvée dans d'autres régions du monde. Cette faible fréquence pourrait être une des particularités de la population génétique hpAfrica1. Elle pourrait expliquer au moins partiellement la faible incidence du CG au Sénégal, et par extension en Afrique, en raison de la large distribution géographique de cette population génétique.

En raison du manque de données issues des pays en voie de développement, les résultats de nos études au Cambodge et au Sénégal sont particulièrement importants. Bien que nos résultats puissent ne pas être représentatifs de la situation globale, l'étude a été mise en place dans une structure hospitalière majeure pour chacun des pays (CHU). Des études ultérieures seront nécessaires pour confirmer nos résultats en se basant sur de larges cohortes de patients, au Cambodge et de façon générale en Afrique, afin de prendre en compte les différents facteurs impliqués dans le processus relatif à la carcinogenèse gastrique : les facteurs de prédisposition génétique de l'hôte, les facteurs environnementaux et la virulence de la bactérie.

## 3. PERSPECTIVES

### 3.1. Etude des migrations humaines

#### 3.1.1. Madagascar

L'île de Madagascar est située à 600 km au large des côtes du Mozambique, beaucoup plus près de l'Afrique que de tout autre continent. L'étude des données linguistiques et archéologiques suggère une colonisation initiale par des peuples parlant des langues austronésiennes vers le 4<sup>ème</sup> siècle après JC, puis une colonisation venue des côtes africaines (Verin 1987; Lewis 2009). Cette migration austronésienne représente le pendant occidental d'un vaste mouvement de populations commencé depuis au moins le troisième millénaire avant notre ère dans l'océan Pacifique. Aujourd'hui, la population de Madagascar peut-être considérée comme le produit d'un métissage poussé entre ces premiers occupants austronésiens et africains. En terme génétique, le métissage est très fort, sauf parmi certaines populations des hautes terres où le phénotype indonésien est préservé (Regueiro *et al.* 2008; Razafindrazaka *et al.* 2010). Cependant, toutes les populations malgaches parlent une langue austronésienne très proche de la langue ma'anyan en vigueur sur l'île indonésienne de Bornéo, séparé de Madagascar par plus de 6000 km (Lewis 2009). Nous avons mis en place une étude des infections à *H. pylori* au CHU de Antananarivo dont les patients sont majoritairement d'ethnie merina (hautes terres). Actuellement 26 souches ont été isolées. Nous espérons apporter des éléments importants sur le peuplement de Madagascar, en corrélant les données de génétique bactérienne avec les données de génétique humaine de l'hôte qui l'héberge. Ce travail sera réalisé ultérieurement en collaboration avec l'équipe de José Machado (IPATIMUP, Medical Faculty of Porto).

#### 3.1.2. Expansion Bantou

En Afrique, les Pygmées et les Khoisans sont des chasseurs-cueilleurs, sans culture ni cheptel. Les Pygmées se distinguent par leur petite taille et vivent en groupes largement éparés à travers la forêt humide d'Afrique centrale. La distribution fragmentaire des 200 000 pygmées, éparpillés au milieu de 120 millions d'agriculteurs bantous, suggèrent qu'ils étaient largement répartis à travers les forêts équatoriales avant qu'ils ne soient déplacés par l'arrivée d'autres populations. Quant aux Khoisans, la région qu'ils occupent en Afrique australe est étonnamment petite pour une population d'anatomie et de langue si distinctes. Toutes les

langues khoisanes sont confinées à l'Afrique australe à deux exceptions près, le hazda et le sandawe, comme échouées en Tanzanie à plus de 1600 km des langues khoisan les plus proches. En outre, des mots khoisan ont été retrouvés beaucoup plus au nord, dans des langues parlées au Kenya (Lewis 2009). A l'instar des Pygmées, les populations khoisanes ont été englouties par les Bantous. Les Bantous sont originaires probablement du Cameroun et du sud-est du Nigeria. Vers - 3000 avant JC, ils commencent à étendre leur territoire dans la forêt équatoriale d'Afrique centrale, puis vers l'est et enfin vers le sud du continent, repoussant progressivement les peuples khoisans qui y vivaient (Diamond 2001). L'étude du polymorphisme génétique d'isolats de *H. pylori* isolés chez des Bantous vivant sur l'axe migratoire historique, de Khoisans et de Pygmées permettrait de mieux comprendre les détails de l'histoire humaine en Afrique. Cette étude n'est pas financée mais pourrait s'appuyer sur le l'Institut Pasteur de Bangui (République centrafricaine) et l'Institut Pasteur de Yaoundé (Cameroun), ainsi que des Instituts partenaires en Afrique du sud.

### **3.2. Structure génétique des gènes *hspA* et *glmM***

L'analyse du polymorphisme de ces 2 gènes montre une très forte structure phyléogéographique (Raymond *et al.* 2004). Nous avons séquencé ces deux gènes chez 100 souches appartenant à l'ensemble des (sous-) populations génétiques décrites par analyse du polymorphisme des 7 gènes de ménage. L'objectif est de comparer les classifications phylogénétiques obtenues et de proposer si les résultats sont concluants, une méthode de classification basée sur le séquençage d'un nombre réduit de gènes. Le travail technique a été effectué, l'analyse des données est en cours de réalisation.

### **3.3. Relation entre la diversité génétique de l'hôte et de la bactérie infectante**

La prévalence de l'infection est beaucoup plus importante dans les pays en voie de développement (80-95 %) que dans les pays industrialisés (15-30 %) (Frenck and Clemens 2003). Cependant, des différences locales de prévalence existent selon l'origine géographique ou ethnique des populations étudiées. Les principaux facteurs de risque d'infection sont la promiscuité, le manque d'hygiène, la qualité de l'eau, un partage du lit pendant l'enfance. D'autres facteurs ont été également identifiés, incluant un faible niveau d'éducation des parents, l'allaitement artificiel, des habitudes culturelles telles que la prémastication, le nombre de personnes dormant dans la même chambre, le nombre de frères et sœurs. Cependant, la relation entre la diversité génétique de la bactérie et de son hôte n'a jamais été étudiée. Pourtant, elle pourrait être un élément clef du mécanisme de l'infection. En outre, une

telle approche permettrait de mieux comprendre les relations très complexes entre l'hôte et ce pathogène.

Ce projet déjà financé concerne uniquement les sites du Sénégal, du Cambodge et de Madagascar et s'inscrit dans le cadre d'une collaboration avec l'équipe de José Machado (IPATIMUP, Medical Faculty of Porto). Il s'appuie sur les données de typage issues de l'analyse du polymorphisme bactérien des 7 gènes de ménage précédemment décrits. La diversité génétique humaine de l'hôte chez qui la souche a été isolée sera étudiée par analyse de 15 microsatellites autosomiques (D3S1358, TH01, D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D, vWA, D8S1179, TPOX et FGA) et de l'ADNmt (fragments HVS-1 et HVS-2). La diversité génétique de l'hôte sera ensuite comparée à celle de la bactérie (Logiciels STRUCTURE et MEGA) afin de déterminer si une relation entre les deux est un élément clef de l'infection humaine (transmission, colonisation et persistance).

Les individus seront également classés selon la présence ou l'absence des principaux génotypes pro-inflammatoires décrits dans la littérature associés à la survenue du CG (IL1  $\beta$ -511C/T, l'intron 2 du gène IL-1RN, IL8-251A/T, TNFA-308G/A) afin de déterminer si il existe une association entre la souche infectante et les génotypes connus pour être associés à un risque de CG.

### 3.4. Etude de la sensibilité des souches de *H. pylori* aux antibiotiques

L'étude de la sensibilité aux antibiotiques des souches isolées (amoxicilline, clarithromycine, métronidazole et lévofloxacine) a été réalisée par la méthode de l'E-test pour l'ensemble des souches du projet selon les recommandations du comité de l'Antibiogramme de la Société Française de Microbiologie. Aucune souche résistante à l'amoxicilline n'a été isolée. Les résultats sont présentés dans le tableau 5 et seront publiés ultérieurement.

**Tableau 5. Résistance des souches de *H. pylori* aux antibiotiques testés**

	Algérie	Sénégal	Madagascar	Cambodge	Grèce	Nouvelle-Calédonie
	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
<b>Amoxicilline</b>	0	0	0	0	0	0
<b>Métronidazole</b>	21 (47)	86 (80)	11 (50)	13 (100)	28 (47)	23 (43)
<b>Clarithromycine</b>	7 (16)	1 (1)	0	3 (23)	18 (31)	5 (9)
<b>Lévofloxacine</b>	1 (2)	16 (15)	0	<sup>1</sup>	4 (7)	<sup>1</sup>
<b>Total</b>	45	108	22	13	59	54

<sup>1</sup> Non réalisé

**Références bibliographiques**

- ABRIL, N., ROLDAN-ARJONA, T., PRIETO-ALAMO, M. J., *et al.* Mutagenesis and DNA repair for alkylation damages in *Escherichia coli* K-12. *Environ Mol Mutagen*, 1992, **19**: 288-296.
- ACHTMAN, M., AZUMA, T., BERG, D. E., *et al.* Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol*, 1999, **32**: 459-470.
- ACHTMAN, M., MORELLI, G., ZHU, P., *et al.* Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc Natl Acad Sci U S A*, 2004, **101**: 17837-17842.
- ACHTMAN, M. and WAGNER, M. Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol*, 2008, **6**: 431-440.
- ADAMS, S. M., KING, T. E., BOSCH, E., *et al.* The case of the unreliable SNP: recurrent back-mutation of Y-chromosomal marker P25 through gene conversion. *Forensic Sci Int*, 2006, **159**: 14-20.
- AKOPYANTS, N. S., FRADKOV, A., DIATCHENKO, L., *et al.* PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. *Proc Natl Acad Sci U S A*, 1998, **95**: 13108-13113.
- ALM, R. A., LING, L. S., MOIR, D. T., *et al.* Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature*, 1999, **397**: 176-180.
- ALM, R. A. and TRUST, T. J. Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J Mol Med*, 1999, **77**: 834-846.
- AMEDEI, A., CAPPON, A., CODOLO, G., *et al.* The neutrophil-activating protein of *Helicobacter pylori* promotes Th1 immune responses. *J Clin Invest*, 2006, **116**: 1092-1101.
- AMIEVA, M. R., VOGELMANN, R., COVACCI, A., *et al.* Disruption of the epithelial apical-junctional complex by *Helicobacter pylori* CagA. *Science*, 2003, **300**: 1430-1434.
- ARGENT, R. H., HALE, J. L., EL-OMAR, E. M., *et al.* Differences in *Helicobacter pylori* CagA tyrosine phosphorylation motif patterns between western and East Asian strains, and influences on interleukin-8 secretion. *J Med Microbiol*, 2008, **57**: 1062-1067.
- ARGENT, R. H., KIDD, M., OWEN, R. J., *et al.* Determinants and consequences of different levels of CagA phosphorylation for clinical isolates of *Helicobacter pylori*. *Gastroenterology*, 2004, **127**: 514-523.
- ARGENT, R. H., THOMAS, R. J., LETLEY, D. P., *et al.* Functional association between the *Helicobacter pylori* virulence factors VacA and CagA. *J Med Microbiol*, 2008, **57**: 145-150.
- ASPHOLM, M., OLFAT, F. O., NORDEN, J., *et al.* SabA is the *H. pylori* hemagglutinin and is polymorphic in binding to sialylated glycans. *PLoS Pathog*, 2006, **2**: e110.
- ATHERTON, J. C. and BLASER, M. J. Coadaptation of *Helicobacter pylori* and humans: ancient history, modern implications. *J Clin Invest*, 2009, **119**: 2475-2487.
- ATHERTON, J. C., CAO, P., PEEK, R. M., JR., *et al.* Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific vacA types with cytotoxin production and peptic ulceration. *J Biol Chem*, 1995, **270**: 17771-17777.
- AXON, A. T. Is *Helicobacter pylori* transmitted by the gastro-oral route? *Aliment Pharmacol Ther*, 1995, **9**: 585-588.
- AZEVEDO, N. F., GUIMARAES, N., FIGUEIREDO, C., *et al.* A new model for the transmission of *Helicobacter pylori*: role of environmental reservoirs as gene pools to increase strain diversity. *Crit Rev Microbiol*, 2007, **33**: 157-169.

- AZEVEDO, N. F., HUNTINGTON, J. and GOODMAN, K. J. The epidemiology of *Helicobacter pylori* and public health implications. *Helicobacter*, 2009, **14 Suppl 1**: 1-7.
- AZUMA, T., YAMAZAKI, S., YAMAKAWA, A., *et al.* Association between diversity in the Src homology 2 domain--containing tyrosine phosphatase binding site of *Helicobacter pylori* CagA protein and gastric atrophy and cancer. *J Infect Dis*, 2004, **189**: 820-827.
- BAAR, C., EPPINGER, M., RADDATZ, G., *et al.* Complete genome sequence and analysis of *Wolinella succinogenes*. *Proc Natl Acad Sci U S A*, 2003, **100**: 11690-11695.
- BACKERT, S. and SELBACH, M. Role of type IV secretion in *Helicobacter pylori* pathogenesis. *Cell Microbiol*, 2008, **10**: 1573-1581.
- BACKSTROM, A., LUNDBERG, C., KERSULYTE, D., *et al.* Metastability of *Helicobacter pylori* bab adhesin genes and dynamics in Lewis b antigen binding. *Proc Natl Acad Sci U S A*, 2004, **101**: 16923-16928.
- BASSO, D., ZAMBON, C. F., LETLEY, D. P., *et al.* Clinical relevance of *Helicobacter pylori* *cagA* and *vacA* gene polymorphisms. *Gastroenterology*, 2008, **135**: 91-99.
- BELLINA-PRYCE, B. and SILAPANTH, P. Weaving cultural identities on trans-Asiatic networks: Upper Thai-Malay peninsula – an early socio-political landscape. *Bulletin de l'Ecole française d'Extrême Orient*, 2006, **93**: 257-294.
- BELLINA, B. Le port protohistorique de Khao Sam Kaeo en Thaïlande péninsulaire: lieu privilégié pour l'étude des premières interactions indiennes et sud-est asiatiques. *Bulletin de l'École française d'Extrême-Orient*, 2002, **89**: 329-343.
- BELLINA, B. Cultural Exchange Between India and Southeast Asia. Production and Distribution of Hard Stone Ornaments (6 c. BC-6 c. AD). 1e éd. Paris: Maison des Sciences de l'Homme, 2007, I-126 p.
- BELLINA, B. and SILAPANTH, P. Khao Sam Kaeo and the Upper Thai Peninsula : understanding the mechanism of early trans-asiatic trade and cultural exchange. *Uncovering Southeast Asia's Past* ed. by E. A. BACUS, I. C. GLOVER and V. C. PIGOTT. Singapore: National University Press, 2006, p. 379-392
- BINSAEED, A. A. Is there a link between seropositivity to *Helicobacter pylori* and hepatitis A virus? A systematic review. *Int J Infect Dis*, 2010, **14**: e567-571.
- BIRD, A. DNA methylation patterns and epigenetic memory. *Genes Dev*, 2002, **16**: 6-21.
- BJORKHOLM, B., SJOLUND, M., FALK, P. G., *et al.* Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*. *Proc Natl Acad Sci U S A*, 2001, **98**: 14607-14612.
- BOCKELMANN, U., DORRIES, H. H., AYUSO-GABELLA, M. N., *et al.* Quantitative PCR monitoring of antibiotic resistance genes and bacterial pathogens in three European artificial groundwater recharge systems. *Appl Environ Microbiol*, 2009, **75**: 154-163.
- BOGAERTS, P., BERHIN, C., NIZET, H., *et al.* Prevalence and mechanisms of resistance to fluoroquinolones in *Helicobacter pylori* strains from patients living in Belgium. *Helicobacter*, 2006, **11**: 441-445.
- BORNSCHEIN, J., SELGRAD, M., WARNECKE, M., *et al.* *Helicobacter pylori* infection is a key risk factor for proximal gastric cancer. *Dig Dis Sci*, 2010, **55**: 3124-3131.
- BOYANOVA, L. and MITOV, I. Geographic map and evolution of primary *Helicobacter pylori* resistance to antibacterial agents. *Expert Rev Anti Infect Ther*, 2010, **8**: 59-70.
- BOYANOVA, L., STANCHEVA, I., SPASSOVA, Z., *et al.* Primary and combined resistance to four antimicrobial agents in *Helicobacter pylori* in Sofia, Bulgaria. *J Med Microbiol*, 2000, **49**: 415-418.
- BROWN, W. M., GEORGE, M., JR. and WILSON, A. C. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A*, 1979, **76**: 1967-1971.

- BRUCE, M. G. and MAAROOS, H. I. Epidemiology of *Helicobacter pylori* infection. *Helicobacter*, 2008, **13 Suppl 1**: 1-6.
- BUNN, J. E., MACKAY, W. G., THOMAS, J. E., *et al.* Detection of *Helicobacter pylori* DNA in drinking water biofilms: implications for transmission in early life. *Lett Appl Microbiol*, 2002, **34**: 450-454.
- BURGERS, R., SCHNEIDER-BRACHERT, W., REISCHL, U., *et al.* *Helicobacter pylori* in human oral cavity and stomach. *Eur J Oral Sci*, 2008, **116**: 297-304.
- BURY-MONE, S., SKOULOUBRIS, S., LABIGNE, A., *et al.* The *Helicobacter pylori* UreI protein: role in adaptation to acidity and identification of residues essential for its activity and for acid activation. *Mol Microbiol*, 2001, **42**: 1021-1034.
- BURY-MONE, S., THIBERGE, J. M., CONTRERAS, M., *et al.* Responsiveness to acidity via metal ion regulators mediates virulence in the gastric pathogen *Helicobacter pylori*. *Mol Microbiol*, 2004, **53**: 623-638.
- BUSLER, V. J., TORRES, V. J., MCCLAIN, M. S., *et al.* Protein-protein interactions among *Helicobacter pylori* *cag* proteins. *J Bacteriol*, 2006, **188**: 4787-4800.
- CABRITA, J., OLEASTRO, M., MATOS, R., *et al.* Features and trends in *Helicobacter pylori* antibiotic resistance in Lisbon area, Portugal (1990-1999). *J Antimicrob Chemother*, 2000, **46**: 1029-1031.
- CAMBAU, E., ALLERHEILIGEN, V., COULON, C., *et al.* Evaluation of a new test, genotype HelicoDR, for molecular detection of antibiotic resistance in *Helicobacter pylori*. *J Clin Microbiol*, 2009, **47**: 3600-3607.
- CAMPBELL, D. I., WARREN, B. F., THOMAS, J. E., *et al.* The African enigma: low prevalence of gastric atrophy, high prevalence of chronic inflammation in West African adults and children. *Helicobacter*, 2001, **6**: 263-267.
- CANEDO, P., CASTANHEIRA-VALE, A. J., LUNET, N., *et al.* The interleukin-8-251\*T/\*A polymorphism is not associated with risk for gastric carcinoma development in a Portuguese population. *Eur J Cancer Prev*, 2008, **17**: 28-32.
- CASSAR, O., AFONSO, P. V., BASSOT, S., *et al.* Novel human herpesvirus 8 subtype D strains in Vanuatu, Melanesia. *Emerg Infect Dis*, 2007, **13**: 1745-1748.
- CASSAR, O., CAPUANO, C., BASSOT, S., *et al.* Human T lymphotropic virus type 1 subtype C melanesian genetic variants of the Vanuatu Archipelago and Solomon Islands share a common ancestor. *J Infect Dis*, 2007, **196**: 510-521.
- CATALDO, F., SIMPORE, J., GRECO, P., *et al.* *Helicobacter pylori* infection in Burkina Faso: an enigma within an enigma. *Dig Liver Dis*, 2004, **36**: 589-593.
- CELLINI, L., GRANDE, R., DI CAMPLI, E., *et al.* Characterization of an *Helicobacter pylori* environmental strain. *J Appl Microbiol*, 2008, **105**: 761-769.
- CENSINI, S., LANGE, C., XIANG, Z., *et al.* *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci U S A*, 1996, **93**: 14648-14653.
- CHAN, A. O., LAM, S. K., WONG, B. C., *et al.* Promoter methylation of E-cadherin gene in gastric mucosa associated with *Helicobacter pylori* infection and in gastric cancer. *Gut*, 2003, **52**: 502-506.
- CHAPUT, C. and BONECCA, I. G. Bases moléculaires de l'interaction de *Helicobacter pylori* avec les cellules épithéliales gastriques. *Hépto-Gastro*, 2006, **13**: 379-388.
- CHEN, L. T., LIN, J. T., TAI, J. J., *et al.* Long-term results of anti-*Helicobacter pylori* therapy in early-stage gastric high-grade transformed MALT lymphoma. *J Natl Cancer Inst*, 2005, **97**: 1345-1353.
- CHERIAN, S., FORBES, D., SANFILIPPO, F., *et al.* *Helicobacter pylori*, helminth infections and growth: a cross-sectional study in a high prevalence population. *Acta Paediatr*, 2009, **98**: 860-864.



- CHIKHI, L., NICHOLS, R. A., BARBUJANI, G., *et al.* Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A*, 2002, **99**: 11008-11013.
- CHOW, T. K., LAMBERT, J. R., WAHLQVIST, M. L., *et al.* *Helicobacter pylori* in Melbourne Chinese immigrants: evidence for oral-oral transmission via chopsticks. *J Gastroenterol Hepatol*, 1995, **10**: 562-569.
- CLEMENS, J., ALBERT, M. J., RAO, M., *et al.* Sociodemographic, hygienic and nutritional correlates of *Helicobacter pylori* infection of young Bangladeshi children. *Pediatr Infect Dis J*, 1996, **15**: 1113-1118.
- COGHLAN, J. G., GILLIGAN, D., HUMPHRIES, H., *et al.* *Campylobacter pylori* and recurrence of duodenal ulcers--a 12-month follow-up study. *Lancet*, 1987, **2**: 1109-1111.
- COHAN, F. M. and PERRY, E. B. A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol*, 2007, **17**: R373-386.
- COOPER, A. M. IL-17 and anti-bacterial immunity: protection versus tissue damage. *Eur J Immunol*, 2009, **39**: 649-652.
- CORREA, P., FONTHAM, E. T., BRAVO, J. C., *et al.* Chemoprevention of gastric dysplasia: randomized trial of antioxidant supplements and anti-*Helicobacter pylori* therapy. *J Natl Cancer Inst*, 2000, **92**: 1881-1888.
- COSTA, K., BACHER, G., ALLMAIER, G., *et al.* The morphological transition of *Helicobacter pylori* cells from spiral to coccoid is preceded by a substantial modification of the cell wall. *J Bacteriol*, 1999, **181**: 3710-3715.
- COUTURIER, M. R., TASCA, E., MONTECUCCO, C., *et al.* Interaction with CagF is required for translocation of CagA into the host via the *Helicobacter pylori* type IV secretion system. *Infect Immun*, 2006, **74**: 273-281.
- COVACCI, A., TELFORD, J. L., DEL GIUDICE, G., *et al.* *Helicobacter pylori* virulence and genetic geography. *Science*, 1999, **284**: 1328-1333.
- COVER, T. L. and BLANKE, S. R. *Helicobacter pylori* VacA, a paradigm for toxin multifunctionality. *Nat Rev Microbiol*, 2005, **3**: 320-332.
- COVER, T. L. and BLASER, M. J. Purification and characterization of the vacuolating toxin from *Helicobacter pylori*. *J Biol Chem*, 1992, **267**: 10570-10575.
- COVER, T. L., HANSON, P. I. and HEUSER, J. E. Acid-induced dissociation of VacA, the *Helicobacter pylori* vacuolating cytotoxin, reveals its pattern of assembly. *J Cell Biol*, 1997, **138**: 759-769.
- CURADO, M. P., EDWARDS, B., SHIN, H. R., *et al.* Cancer Incidence in Five Continents Vol. IX [monographie en ligne]. 1-11-2007. Format PDF. Disponible sur: <http://www.iarc.fr/en/publications/pdfs-online/epi/sp160/index.php>
- D'ELIOS, M. M., MANGHETTI, M., DE CARLI, M., *et al.* T helper 1 effector cells specific for *Helicobacter pylori* in the gastric antrum of patients with peptic ulcer disease. *J Immunol*, 1997, **158**: 962-967.
- DATTA, S., CHATTOPADHYAY, S., PATRA, R., *et al.* Most *Helicobacter pylori* strains of Kolkata in India are resistant to metronidazole but susceptible to other drugs commonly used for eradication and ulcer therapy. *Aliment Pharmacol Ther*, 2005, **22**: 51-57.
- DE BERNARD, M., ARICO, B., PAPINI, E., *et al.* *Helicobacter pylori* toxin VacA induces vacuole formation by acting in the cell cytosol. *Mol Microbiol*, 1997, **26**: 665-674.
- DE BERNARD, M., CAPPON, A., PANCOTTO, L., *et al.* The *Helicobacter pylori* VacA cytotoxin activates RBL-2H3 cells by inducing cytosolic calcium oscillations. *Cell Microbiol*, 2005, **7**: 191-198.
- DE FRANCESCO, V., GIORGIO, F., HASSAN, C., *et al.* Worldwide *H. pylori* antibiotic resistance: a systematic review. *J Gastrointestin Liver Dis*, 2010, **19**: 409-414.

- DE JONGE, R., DURRANI, Z., RIJKEMA, S. G., *et al.* Role of the *Helicobacter pylori* outer-membrane proteins AlpA and AlpB in colonization of the guinea pig stomach. *J Med Microbiol*, 2004, **53**: 375-379.
- DE REUSE, H. and BERESWILL, S. Ten years after the first *Helicobacter pylori* genome: comparative and functional genomics provide new insights in the variability and adaptability of a persistent pathogen. *FEMS Immunol Med Microbiol*, 2007, **50**: 165-176.
- DEBETS-OSENKOPP, Y. J., HERSCHEID, A. J., POT, R. G., *et al.* Prevalence of *Helicobacter pylori* resistance to metronidazole, clarithromycin, amoxicillin, tetracycline and trovafloxacin in The Netherlands. *J Antimicrob Chemother*, 1999, **43**: 511-515.
- DELTOUR, S., CHOPIN, V. and LEPRINCE, D. Epigenetics and cancer. *Med Sci (Paris)*, 2005, **21**: 405-411.
- DERAKHSHAN, M. H., MALEKZADEH, R., WATABE, H., *et al.* Combination of gastric atrophy, reflux symptoms and histological subtype indicates two distinct aetiologies of gastric cardia cancer. *Gut*, 2008, **57**: 298-305.
- DEVI, S. M., AHMED, I., FRANICALACCI, P., *et al.* Ancestral European roots of *Helicobacter pylori* in India. *BMC Genomics*, 2007, **8**: 184.
- DEVI, S. M., AHMED, I., KHAN, A. A., *et al.* Genomes of *Helicobacter pylori* from native Peruvians suggest admixture of ancestral and modern lineages and reveal a western type *cag*-pathogenicity island. *BMC Genomics*, 2006, **7**: 191.
- DIAMOND, J. De l'inégalité parmi les sociétés. Essai sur l'homme et l'environnement dans l'histoire 1e éd. Paris: Gallimard, 2001, I-482 p.
- DOMINGUEZ-BELLO, M. G., PEREZ, M. E., BORTOLINI, M. C., *et al.* Amerindian *Helicobacter pylori* strains go extinct, as european strains expand their host range. *PLoS One*, 2008, **3**: e3307.
- DORE, M. P., BILOTTA, M., VAIRA, D., *et al.* High prevalence of *Helicobacter pylori* infection in shepherds. *Dig Dis Sci*, 1999, **44**: 1161-1164.
- DORE, M. P., SEPULVEDA, A. R., EL-ZIMAITY, H., *et al.* Isolation of *Helicobacter pylori* from sheep-implications for transmission to humans. *Am J Gastroenterol*, 2001, **96**: 1396-1401.
- DORE, M. P., SEPULVEDA, A. R., OSATO, M. S., *et al.* *Helicobacter pylori* in sheep milk. *Lancet*, 1999, **354**: 132.
- DORER, M. S., TALARICO, S. and SALAMA, N. R. *Helicobacter pylori*'s unconventional role in health and disease. *PLoS Pathog*, 2009, **5**: e1000544.
- DOWSETT, S. A. and KOWOLIK, M. J. Oral *Helicobacter pylori*: can we stomach it? *Crit Rev Oral Biol Med*, 2003, **14**: 226-233.
- DUARTE, I., SANTOS, A., SOUSA, H., *et al.* G-308A TNF-alpha polymorphism is associated with an increased risk of invasive cervical cancer. *Biochem Biophys Res Commun*, 2005, **334**: 588-592.
- EATON, K. A., CATRENICH, C. E., MAKIN, K. M., *et al.* Virulence of coccoid and bacillary forms of *Helicobacter pylori* in gnotobiotic piglets. *J Infect Dis*, 1995, **171**: 459-462.
- EATON, K. A. and KRAKOWKA, S. Effect of gastric pH on urease-dependent colonization of gnotobiotic piglets by *Helicobacter pylori*. *Infect Immun*, 1994, **62**: 3604-3607.
- EATON, K. A., MORGAN, D. R. and KRAKOWKA, S. Motility as a factor in the colonisation of gnotobiotic piglets by *Helicobacter pylori*. *J Med Microbiol*, 1992, **37**: 123-127.

- EATON, K. A., SUERBAUM, S., JOSEPHANS, C., *et al.* Colonization of gnotobiotic piglets by *Helicobacter pylori* deficient in two flagellin genes. *Infect Immun*, 1996, **64**: 2445-2448.
- EBERHARD, W. A history of China. 4e éd. Berkeley: University of California Press, 1977, I-382 p.
- EL-OMAR, E. M., RABKIN, C. S., GAMMON, M. D., *et al.* Increased risk of noncardia gastric cancer associated with proinflammatory cytokine gene polymorphisms. *Gastroenterology*, 2003, **124**: 1193-1201.
- ELVISS, N. C., LAWSON, A. J. and OWEN, R. J. Application of 3'-mismatched reverse primer PCR compared with real-time PCR and PCR-RFLP for the rapid detection of 23S rDNA mutations associated with clarithromycin resistance in *Helicobacter pylori*. *Int J Antimicrob Agents*, 2004, **23**: 349-355.
- EPPINGER, M., BAAR, C., LINZ, B., *et al.* Who ate whom? Adaptive *Helicobacter* genomic changes that accompanied a host jump from early humans to large felines. *PLoS Genet*, 2006, **2**: e120.
- EPPINGER, M., BAAR, C., RADDATZ, G., *et al.* Comparative analysis of four *Campylobacteriales*. *Nat Rev Microbiol*, 2004, **2**: 872-885.
- ERMAK, T. H., GIANNASCA, P. J., NICHOLS, R., *et al.* Immunization of mice with urease vaccine affords protection against *Helicobacter pylori* infection in the absence of antibodies and is mediated by MHC class II-restricted responses. *J Exp Med*, 1998, **188**: 2277-2288.
- ERNST, P. B. and GOLD, B. D. The disease spectrum of *Helicobacter pylori*: the immunopathogenesis of gastroduodenal ulcer and gastric cancer. *Annu Rev Microbiol*, 2000, **54**: 615-640.
- FALLONE, C. A. Epidemiology of the antibiotic resistance of *Helicobacter pylori* in Canada. *Can J Gastroenterol*, 2000, **14**: 879-882.
- FALUSH, D., KRAFT, C., TAYLOR, N. S., *et al.* Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc Natl Acad Sci U S A*, 2001, **98**: 15056-15061.
- FALUSH, D., STEPHENS, M. and PRITCHARD, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 2003, **164**: 1567-1587.
- FALUSH, D., WIRTH, T., LINZ, B., *et al.* Traces of human migrations in *Helicobacter pylori* populations. *Science*, 2003, **299**: 1582-1585.
- FAN, X. G., CHUA, A., LI, T. G., *et al.* Survival of *Helicobacter pylori* in milk and tap water. *J Gastroenterol Hepatol*, 1998, **13**: 1096-1098.
- FERRERO, R. L., CUSSAC, V., COURCOUX, P., *et al.* Construction of isogenic urease-negative mutants of *Helicobacter pylori* by allelic exchange. *J Bacteriol*, 1992, **174**: 4212-4217.
- FERRERO, R. L., THIBERGE, J. M., HUERRE, M., *et al.* Immune responses of specific-pathogen-free mice to chronic *Helicobacter pylori* (strain SS1) infection. *Infect Immun*, 1998, **66**: 1349-1355.
- FIGUEIREDO, C., MACHADO, J. C., PHAROAH, P., *et al.* *Helicobacter pylori* and interleukin 1 genotyping: an opportunity to identify high-risk individuals for gastric carcinoma. *J Natl Cancer Inst*, 2002, **94**: 1680-1687.
- FOCK, K. M. and ANG, T. L. Epidemiology of *Helicobacter pylori* infection and gastric cancer in Asia. *J Gastroenterol Hepatol*, 2010, **25**: 479-486.

- FONTANA, C., FAVARO, M., MINELLI, S., *et al.* New site of modification of 23S rRNA associated with clarithromycin resistance of *Helicobacter pylori* clinical isolates. *Antimicrob Agents Chemother*, 2002, **46**: 3765-3769.
- FORD, A. C. and AXON, A. T. Epidemiology of *Helicobacter pylori* infection and public health implications. *Helicobacter*, 2010, **15 Suppl 1**: 1-6.
- FORD, A. C., DELANEY, B. C., FORMAN, D., *et al.* Eradication therapy in *Helicobacter pylori* positive peptic ulcer disease: systematic review and economic analysis. *Am J Gastroenterol*, 2004, **99**: 1833-1855.
- FOX, J. G. Non-human reservoirs of *Helicobacter pylori*. *Aliment Pharmacol Ther*, 1995, **9 Suppl 2**: 93-103.
- FRENCK, R. W., JR. and CLEMENS, J. *Helicobacter* in the developing world. *Microbes Infect*, 2003, **5**: 705-713.
- FUCCIO, L., ZAGARI, R. M., EUSEBI, L. H., *et al.* Meta-analysis: can *Helicobacter pylori* eradication treatment reduce the risk for gastric cancer? *Ann Intern Med*, 2009, **151**: 121-128.
- FUJIMURA, S., KATO, S. and KAWAMURA, T. *Helicobacter pylori* in Japanese river water and its prevalence in Japanese children. *Lett Appl Microbiol*, 2004, **38**: 517-521.
- FUJIMURA, S., KAWAMURA, T., KATO, S., *et al.* Detection of *Helicobacter pylori* in cow's milk. *Lett Appl Microbiol*, 2002, **35**: 504-507.
- GAGNEUX, S., DERIEMER, K., VAN, T., *et al.* Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A*, 2006, **103**: 2869-2873.
- GALMICHE, A., RASSOW, J., DOYE, A., *et al.* The N-terminal 34 kDa fragment of *Helicobacter pylori* vacuolating cytotoxin targets mitochondria and induces cytochrome c release. *EMBO J*, 2000, **19**: 6361-6370.
- GEBERT, B., FISCHER, W., WEISS, E., *et al.* *Helicobacter pylori* vacuolating cytotoxin inhibits T lymphocyte activation. *Science*, 2003, **301**: 1099-1102.
- GERRITS, M. M., GODOY, A. P., KUIPERS, E. J., *et al.* Multiple mutations in or adjacent to the conserved penicillin-binding protein motifs of the penicillin-binding protein 1A confer amoxicillin resistance to *Helicobacter pylori*. *Helicobacter*, 2006, **11**: 181-187.
- GERRITS, M. M., SCHUIJFFEL, D., VAN ZWET, A. A., *et al.* Alterations in penicillin-binding protein 1A confer resistance to beta-lactam antibiotics in *Helicobacter pylori*. *Antimicrob Agents Chemother*, 2002, **46**: 2229-2233.
- GHIARA, P., MARCHETTI, M., BLASER, M. J., *et al.* Role of the *Helicobacter pylori* virulence factors vacuolating cytotoxin, CagA, and urease in a mouse model of disease. *Infect Immun*, 1995, **63**: 4154-4160.
- GHIL, H. M., YOO, J. H., JUNG, W. S., *et al.* Survey of *Helicobacter* infection in domestic and feral cats in Korea. *J Vet Sci*, 2009, **10**: 67-72.
- GILES, R. E., BLANC, H., CANN, H. M., *et al.* Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A*, 1980, **77**: 6715-6719.
- GISBERT, J. P. The recurrence of *Helicobacter pylori* infection: incidence and variables influencing it. A critical review. *Am J Gastroenterol*, 2005, **100**: 2083-2099.
- GLOBOCAN 2010 [Ressource électronique]. [Lyon]: International Agency for research on cancer, 2008. [réf. 1er décembre 2010]. France. Disponible sur: <http://globocan.iarc.fr>
- GLUPCZYNSKI, Y., BROUTET, N., CANTAGREL, A., *et al.* Comparison of the E test and agar dilution method for antimicrobial susceptibility testing of *Helicobacter pylori*. *Eur J Clin Microbiol Infect Dis*, 2002, **21**: 549-552.
- GLYNN, M. K., FRIEDMAN, C. R., GOLD, B. D., *et al.* Seroincidence of *Helicobacter pylori* infection in a cohort of rural Bolivian children: acquisition and analysis of possible risk factors. *Clin Infect Dis*, 2002, **35**: 1059-1065.

- GONZALEZ, C. A. and LOPEZ-CARRILLO, L. *Helicobacter pylori*, nutrition and smoking interactions: their impact in gastric carcinogenesis. *Scand J Gastroenterol*, 2010, **45**: 6-14.
- GONZALEZ, C. A., SALA, N. and CAPELLA, G. Genetic susceptibility and gastric cancer risk. *Int J Cancer*, 2002, **100**: 249-260.
- GOODMAN, K. J., CORREA, P., TENGANA AUX, H. J., *et al.* *Helicobacter pylori* infection in the Colombian Andes: a population-based study of transmission pathways. *Am J Epidemiol*, 1996, **144**: 290-299.
- GOODWIN, A., KERSULYTE, D., SISSON, G., *et al.* Metronidazole resistance in *Helicobacter pylori* is due to null mutations in a gene (*rdxA*) that encodes an oxygen-insensitive NADPH nitroreductase. *Mol Microbiol*, 1998, **28**: 383-393.
- GOODWIN, A. C., WEINBERGER, D. M., FORD, C. B., *et al.* Expression of the *Helicobacter pylori* adhesin SabA is controlled via phase variation and the ArsRS signal transduction system. *Microbiology*, 2008, **154**: 2231-2240.
- GOODWIN, C., ARMSTRONG, J., CHILVERS, T., *et al.* Transfer of *Campylobacter pyloridis* and *Campylobacter mustelae* to *Helicobacter* gen. nov. as *Helicobacter pylori* comb. nov. and *Helicobacter mustelae* comb. nov., respectively. *Int. J. Syst. Bacteriol.*, 1989, **39**: 397-405.
- GRAY, R. D., DRUMMOND, A. J. and GREENHILL, S. J. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*, 2009, **323**: 479-483.
- GRAY, R. D. and JORDAN, F. M. Language trees support the express-train sequence of Austronesian expansion. *Nature*, 2000, **405**: 1052-1055.
- GRESSMANN, H., LINZ, B., GHAI, R., *et al.* Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet*, 2005, **1**: e43.
- GUILFORD, P., HOPKINS, J., HARRAWAY, J., *et al.* E-cadherin germline mutations in familial gastric cancer. *Nature*, 1998, **392**: 402-405.
- GUNDOGDU, O., BENTLEY, S. D., HOLDEN, M. T., *et al.* Re-annotation and re-analysis of the *Campylobacter jejuni* NCTC11168 genome sequence. *BMC Genomics*, 2007, **8**: 162.
- HA, N. C., OH, S. T., SUNG, J. Y., *et al.* Supramolecular assembly and acid resistance of *Helicobacter pylori* urease. *Nat Struct Biol*, 2001, **8**: 505-509.
- HAN, Y. H., LIU, W. Z., SHI, Y. Z., *et al.* Comparative genomics profiling of clinical isolates of *Helicobacter pylori* in Chinese populations using DNA microarray. *J Microbiol*, 2007, **45**: 21-28.
- HAO, Q., LI, Y., ZHANG, Z. J., *et al.* New mutation points in 23S rRNA gene associated with *Helicobacter pylori* resistance to clarithromycin in northeast China. *World J Gastroenterol*, 2004, **10**: 1075-1077.
- HATAKEYAMA, M. *Helicobacter pylori* CagA--a potential bacterial oncoprotein that functionally mimics the mammalian Gab family of adaptor proteins. *Microbes Infect*, 2003, **5**: 143-150.
- HATAKEYAMA, M. *Helicobacter pylori* and gastric carcinogenesis. *J Gastroenterol*, 2009, **44**: 239-248.
- HAUKIOJA, A., IHALIN, R., LOIMARANTA, V., *et al.* Sensitivity of *Helicobacter pylori* to an innate defence mechanism, the lactoperoxidase system, in buffer and in human whole saliva. *J Med Microbiol*, 2004, **53**: 855-860.
- HIGASHI, H., TSUTSUMI, R., FUJITA, A., *et al.* Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci U S A*, 2002, **99**: 14428-14433.
- HIGASHI, H., TSUTSUMI, R., MUTO, S., *et al.* SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. *Science*, 2002, **295**: 683-686.

- HIGASHI, H., YOKOYAMA, K., FUJII, Y., *et al.* EPIYA motif is a membrane-targeting signal of *Helicobacter pylori* virulence factor CagA in mammalian cells. *J Biol Chem*, 2005, **280**: 23130-23137.
- HIGHAM, C. The archaeology of mainland Southeast asia: from 10,000 B.C. to the fall of Angkoréd. Cambridge: Cambridge University Press, 1989, I-404 p.
- HILDEBRAND, P., MEYER-WYSS, B. M., MOSSI, S., *et al.* Risk among gastroenterologists of acquiring *Helicobacter pylori* infection: case-control study. *BMJ*, 2000, **321**: 149.
- HILL, C., SOARES, P., MORMINA, M., *et al.* A mitochondrial stratigraphy for island southeast Asia. *Am J Hum Genet*, 2007, **80**: 29-43.
- HILL, C., SOARES, P., MORMINA, M., *et al.* Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol Biol Evol*, 2006, **23**: 2480-2491.
- HIRAI, I., SASAKI, T., KIMOTO, A., *et al.* Infection of less virulent *Helicobacter pylori* strains in asymptomatic healthy individuals in Thailand as a potential contributing factor to the Asian enigma. *Microbes Infect*, 2010, **12**: 227-230.
- HISATSUNE, J., NAKAYAMA, M., ISOMOTO, H., *et al.* Molecular characterization of *Helicobacter pylori* VacA induction of IL-8 in U937 cells reveals a prominent role for p38MAPK in activating transcription factor-2, cAMP response element binding protein, and NF-kappaB activation. *J Immunol*, 2008, **180**: 5017-5027.
- HO, L., CHAN, S. Y., BURK, R. D., *et al.* The genetic drift of human papillomavirus type 16 is a means of reconstructing prehistoric viral spread and the movement of ancient human populations. *J Virol*, 1993, **67**: 6413-6423.
- HONDA, K., OHKUSA, T., TAKASHIMIZU, I., *et al.* High risk of *Helicobacter pylori* infection in young Japanese dentists. *J Gastroenterol Hepatol*, 2001, **16**: 862-865.
- HOPKINS, R. J., VIAL, P. A., FERRECCIO, C., *et al.* Seroprevalence of *Helicobacter pylori* in Chile: vegetables may serve as one route of transmission. *J Infect Dis*, 1993, **168**: 222-226.
- HUANG, J. Q., ZHENG, G. F., SUMANAC, K., *et al.* Meta-analysis of the relationship between *cagA* seropositivity and gastric cancer. *Gastroenterology*, 2003, **125**: 1636-1644.
- HUANG, S., KANG, J. and BLASER, M. J. Antimutator role of the DNA glycosylase *mutY* gene in *Helicobacter pylori*. *J Bacteriol*, 2006, **188**: 6224-6234.
- HUMBERT, O., PRUDHOMME, M., HAKENBECK, R., *et al.* Homeologous recombination and mismatch repair during transformation in *Streptococcus pneumoniae*: saturation of the Hex mismatch repair system. *Proc Natl Acad Sci U S A*, 1995, **92**: 9052-9056.
- HWANG, I. R., KODAMA, T., KIKUCHI, S., *et al.* Effect of interleukin 1 polymorphisms on gastric mucosal interleukin 1 beta production in *Helicobacter pylori* infection. *Gastroenterology*, 2002, **123**: 1793-1803.
- ISHIKAWA, S., OHTA, T. and HATAKEYAMA, M. Stability of *Helicobacter pylori* CagA oncoprotein in human gastric epithelial cells. *FEBS Lett*, 2009, **583**: 2414-2418.
- ISOMOTO, H., MOSS, J. and HIRAYAMA, T. Pleiotropic actions of *Helicobacter pylori* vacuolating cytotoxin, VacA. *Tohoku J Exp Med*, 2010, **220**: 3-14.
- ISRAEL, D. A., LOU, A. S. and BLASER, M. J. Characteristics of *Helicobacter pylori* natural transformation. *FEMS Microbiol Lett*, 2000, **186**: 275-280.
- ISRAEL, D. A., SALAMA, N., KRISHNA, U., *et al.* *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc Natl Acad Sci U S A*, 2001, **98**: 14625-14630.
- IVIE, S. E., MCCLAIN, M. S., TORRES, V. J., *et al.* *Helicobacter pylori* VacA subdomain required for intracellular toxin activity and assembly of functional oligomeric complexes. *Infect Immun*, 2008, **76**: 2843-2851.

- IWAMOTO, H., CZAJKOWSKY, D. M., COVER, T. L., *et al.* VacA from *Helicobacter pylori*: a hexameric chloride channel. *FEBS Lett*, 1999, **450**: 101-104.
- JAFRI, W., YAKOOB, J., ABID, S., *et al.* *Helicobacter pylori* infection in children: population-based age-specific prevalence and risk factors in a developing country. *Acta Paediatr*, 2010, **99**: 279-282.
- JANZON, A., BHUIYAN, T., LUNDGREN, A., *et al.* Presence of high numbers of transcriptionally active *Helicobacter pylori* in vomitus from Bangladeshi patients suffering from acute gastroenteritis. *Helicobacter*, 2009, **14**: 237-247.
- JEONG, J. Y., MUKHOPADHYAY, A. K., DAILIDIENE, D., *et al.* Sequential inactivation of *rdxA* (HP0954) and *frxA* (HP0642) nitroreductase genes causes moderate and high-level metronidazole resistance in *Helicobacter pylori*. *J Bacteriol*, 2000, **182**: 5082-5090.
- JEX, A. R., LIM, Y. A., J, M. B., *et al.* Soil-transmitted helminths of humans in Southeast Asia-towards integrated control. *Adv Parasitol*, 2011, **74**: 231-265.
- JIANG, Q., HIRATSUKA, K. and TAYLOR, D. E. Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity. *Mol Microbiol*, 1996, **20**: 833-842.
- JIANG, X. and DOYLE, M. P. Effect of environmental and substrate factors on survival and growth of *Helicobacter pylori*. *J Food Prot*, 1998, **61**: 929-933.
- JIN, L., SEIELSTAD, M. and XIAO, C. Genetic, Linguistic and Archaeological Perspectives on Human Diversity in Southeast Asia. 1e éd. River Edge, New Jersey: World Scientific Publishing Co, 2001, I-172 p.
- JOBLING, M. A. and TYLER-SMITH, C. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet*, 2003, **4**: 598-612.
- JONES, K. R., JOO, Y. M., JANG, S., *et al.* Polymorphism in the CagA EPIYA motif impacts development of gastric cancer. *J Clin Microbiol*, 2009, **47**: 959-968.
- JOSEPHANS, C., LABIGNE, A. and SUERBAUM, S. Comparative ultrastructural and functional studies of *Helicobacter pylori* and *Helicobacter mustelae* flagellin mutants: both flagellin subunits, FlaA and FlaB, are necessary for full motility in *Helicobacter* species. *J Bacteriol*, 1995, **177**: 3010-3020.
- KAAKOUSH, N. O., ASECIO, C., MEGRAUD, F., *et al.* A redox basis for metronidazole resistance in *Helicobacter pylori*. *Antimicrob Agents Chemother*, 2009, **53**: 1884-1891.
- KANG, J. and BLASER, M. J. Bacterial populations as perfect gases: genomic integrity and diversification tensions in *Helicobacter pylori*. *Nat Rev Microbiol*, 2006, **4**: 826-836.
- KANG, J. and BLASER, M. J. UvrD helicase suppresses recombination and DNA damage-induced deletions. *J Bacteriol*, 2006, **188**: 5450-5459.
- KANSAU, I., RAYMOND, J., BINGEN, E., *et al.* Genotyping of *Helicobacter pylori* isolates by sequencing of PCR products and comparison with the RAPD technique. *Res Microbiol*, 1996, **147**: 661-669.
- KAYSER, M., BRAUER, S., CORDAUX, R., *et al.* Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol Biol Evol*, 2006, **23**: 2234-2244.
- KAYSER, M., BRAUER, S., WEISS, G., *et al.* Melanesian origin of Polynesian Y chromosomes. *Curr Biol*, 2000, **10**: 1237-1246.
- KENNEMANN, L., DIDELOT, X., AEBISCHER, T., *et al.* *Helicobacter pylori* genome evolution during human infection. *Proc Natl Acad Sci U S A*, 2011, **108**: 5033-5038.
- KERSULYTE, D., KALIA, A., GILMAN, R. H., *et al.* *Helicobacter pylori* from Peruvian amerindians: traces of human migrations in strains from remote Amazon, and genome sequence of an Amerind strain. *PLoS One*, 2011, **5**: e15076.

- KHAN, R., NAHAR, S., SULTANA, J., *et al.* T2182C mutation in 23S rRNA is associated with clarithromycin resistance in *Helicobacter pylori* isolates obtained in Bangladesh. *Antimicrob Agents Chemother*, 2004, **48**: 3567-3569.
- KIDD, M., LASTOVICA, A. J., ATHERTON, J. C., *et al.* Heterogeneity in the *Helicobacter pylori vacA* and *cagA* genes: association with gastroduodenal disease in South Africa? *Gut*, 1999, **45**: 499-502.
- KIM, J. M., KIM, J. S., JUNG, H. C., *et al.* Distribution of antibiotic MICs for *Helicobacter pylori* strains over a 16-year period in patients from Seoul, South Korea. *Antimicrob Agents Chemother*, 2004, **48**: 4843-4847.
- KIM, J. M., KIM, J. S., KIM, N., *et al.* Distribution of fluoroquinolone MICs in *Helicobacter pylori* strains from Korean patients. *J Antimicrob Chemother*, 2005, **56**: 965-967.
- KIM, J. S., CHANG, J. H., CHUNG, S. I., *et al.* Molecular cloning and characterization of the *Helicobacter pylori fliD* gene, an essential factor in flagellar structure and motility. *J Bacteriol*, 1999, **181**: 6969-6976.
- KIMURA, M., GOTO, S., WADA, A., *et al.* Vacuolating cytotoxin purified from *Helicobacter pylori* causes mitochondrial damage in human gastric cells. *Microb Pathog*, 1999, **26**: 45-52.
- KIST, M. *Helicobacter pylori*: primary antimicrobial resistance and first-line treatment strategies. *Euro Surveill*, 2007, **12**: E1-2.
- KIVI, M., TINDBERG, Y., SORBERG, M., *et al.* Concordance of *Helicobacter pylori* strains within families. *J Clin Microbiol*, 2003, **41**: 5604-5608.
- KLEIN, P. D., GRAHAM, D. Y., GAILLOUR, A., *et al.* Water source as risk factor for *Helicobacter pylori* infection in Peruvian children. Gastrointestinal Physiology Working Group. *Lancet*, 1991, **337**: 1503-1506.
- KOBAYASHI, I., MURAKAMI, K., KATO, M., *et al.* Changing antimicrobial susceptibility epidemiology of *Helicobacter pylori* strains in Japan between 2002 and 2005. *J Clin Microbiol*, 2007, **45**: 4006-4010.
- KONG, H., LIN, L. F., PORTER, N., *et al.* Functional analysis of putative restriction-modification system genes in the *Helicobacter pylori* J99 genome. *Nucleic Acids Res*, 2000, **28**: 3216-3223.
- KONTUREK, P. C., KANIA, J., KONTUREK, J. W., *et al.* *H. pylori* infection, atrophic gastritis, cytokines, gastrin, COX-2, PPAR $\gamma$  and impaired apoptosis in gastric carcinogenesis. *Med Sci Monit*, 2003, **9**: SR53-66.
- KONTUREK, P. C., KONTUREK, S. J., STARZYSKA, T., *et al.* *Helicobacter pylori*-gastrin link in MALT lymphoma. *Aliment Pharmacol Ther*, 2000, **14**: 1311-1318.
- KOSTRZYNSKA, M., BETTS, J. D., AUSTIN, J. W., *et al.* Identification, characterization, and spatial localization of two flagellin species in *Helicobacter pylori* flagella. *J Bacteriol*, 1991, **173**: 937-946.
- KRAFT, C., STACK, A., JOSEPHANS, C., *et al.* Genomic changes during chronic *Helicobacter pylori* infection. *J Bacteriol*, 2006, **188**: 249-254.
- KRAFT, C. and SUERBAUM, S. Mutation and recombination in *Helicobacter pylori*: mechanisms and role in generating strain diversity. *Int J Med Microbiol*, 2005, **295**: 299-305.
- KRIENITZ, W. Ueber das Auftreten von Spirockäten verschiedener Form im Mageninhalt bei Carcinoma Ventriculi. *Deutsche Med Wchnschr*, 1906, **28**: 872-890.
- KUDO, T., NURGALIEVA, Z. Z., CONNER, M. E., *et al.* Correlation between *Helicobacter pylori* OipA protein expression and *oipA* gene switch status. *J Clin Microbiol*, 2004, **42**: 2279-2281.
- KUIPERS, E. J. Review article: exploring the link between *Helicobacter pylori* and gastric cancer. *Aliment Pharmacol Ther*, 1999, **13 Suppl 1**: 3-11.



- KUIPERS, E. J., ISRAEL, D. A., KUSTERS, J. G., *et al.* Evidence for a conjugation-like mechanism of DNA transfer in *Helicobacter pylori*. *J Bacteriol*, 1998, **180**: 2901-2905.
- KULICK, S., MOCCIA, C., DIDELOT, X., *et al.* Mosaic DNA imports with interspersions of recipient sequence after natural transformation of *Helicobacter pylori*. *PLoS One*, 2008, **3**: e3797.
- KURASHIMA, Y., MURATA-KAMIYA, N., KIKUCHI, K., *et al.* Deregulation of beta-catenin signal by *Helicobacter pylori* CagA requires the CagA-multimerization sequence. *Int J Cancer*, 2008, **122**: 823-831.
- KUSTERS, J. G., GERRITS, M. M., VAN STRIJP, J. A., *et al.* Coccoid forms of *Helicobacter pylori* are the morphologic manifestation of cell death. *Infect Immun*, 1997, **65**: 3672-3679.
- KWOK, T., ZABLER, D., URMAN, S., *et al.* *Helicobacter* exploits integrin for type IV secretion and kinase activation. *Nature*, 2007, **449**: 862-866.
- KWON, D. H., EL-ZAATARI, F. A., KATO, M., *et al.* Analysis of *rdxA* and involvement of additional genes encoding NAD(P)H flavin oxidoreductase (*FrxA*) and ferredoxin-like protein (*FdxB*) in metronidazole resistance of *Helicobacter pylori*. *Antimicrob Agents Chemother*, 2000, **44**: 2133-2142.
- LAMB, A., YANG, X. D., TSANG, Y. H., *et al.* *Helicobacter pylori* CagA activates NF-kappaB by targeting TAK1 for TRAF6-mediated Lys 63 ubiquitination. *EMBO Rep*, 2009, **10**: 1242-1249.
- LATIFI-NAVID, S., GHORASHI, S. A., SIAVOSHI, F., *et al.* Ethnic and geographic differentiation of *Helicobacter pylori* within Iran. *PLoS One*, 2010, **5**: e9645.
- LEE, J. H., SHIN, J. H., ROE, I. H., *et al.* Impact of clarithromycin resistance on eradication of *Helicobacter pylori* in infected adults. *Antimicrob Agents Chemother*, 2005, **49**: 1600-1603.
- LEHOURS, P., MENARD, A., DUPOUY, S., *et al.* Evaluation of the association of nine *Helicobacter pylori* virulence factors with strains involved in low-grade gastric mucosa-associated lymphoid tissue lymphoma. *Infect Immun*, 2004, **72**: 880-888.
- LEHOURS, P. and YILMAZ, O. Epidemiology of *Helicobacter pylori* infection. *Helicobacter*, 2007, **12 Suppl 1**: 1-3.
- LEHOURS, P., ZHENG, Z., SKOGLUND, A., *et al.* Is there a link between the lipopolysaccharide of *Helicobacter pylori* gastric MALT lymphoma associated strains and lymphoma pathogenesis? *PLoS One*, 2009, **4**: e7297.
- LEODOLTER, A., KULIG, M., BRASCH, H., *et al.* A meta-analysis comparing eradication, healing and relapse rates in patients with *Helicobacter pylori*-associated gastric or duodenal ulcer. *Aliment Pharmacol Ther*, 2001, **15**: 1949-1958.
- LERTRIT, P., POOLSUWAN, S., THOSARAT, R., *et al.* Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. *Am J Phys Anthropol*, 2008, **137**: 425-440.
- LEUNG, W. K., LIN, S. R., CHING, J. Y., *et al.* Factors predicting progression of gastric intestinal metaplasia: results of a randomised trial on *Helicobacter pylori* eradication. *Gut*, 2004, **53**: 1244-1249.
- LEUNG, W. K., NG, E. K., CHAN, W. Y., *et al.* Risk factors associated with the development of intestinal metaplasia in first-degree relatives of gastric cancer patients. *Cancer Epidemiol Biomarkers Prev*, 2005, **14**: 2982-2986.
- LEUNG, W. K., SIU, K. L., KWOK, C. K., *et al.* Isolation of *Helicobacter pylori* from vomitus in children and its implication in gastro-oral transmission. *Am J Gastroenterol*, 1999, **94**: 2881-2884.

- LEUNG, W. K., SUNG, J. J., LING, T. K., *et al.* Use of chopsticks for eating and *Helicobacter pylori* infection. *Dig Dis Sci*, 1999, **44**: 1173-1176.
- LEUNK, R. D., JOHNSON, P. T., DAVID, B. C., *et al.* Cytotoxic activity in broth-culture filtrates of *Campylobacter pylori*. *J Med Microbiol*, 1988, **26**: 93-99.
- LEWIS, P. *Ethnologue: Languages of the World*. 16e éd. Dallas: SIL International, 2009, I-1248 p.
- LEY, C., MOHAR, A., GUARNER, J., *et al.* *Helicobacter pylori* eradication and gastric preneoplastic conditions: a randomized, double-blind, placebo-controlled trial. *Cancer Epidemiol Biomarkers Prev*, 2004, **13**: 4-10.
- LI, J., OU, Z., WANG, F., *et al.* Distinctiveness of the *cagA* genotype in children and adults with peptic symptoms in South China. *Helicobacter*, 2009, **14**: 248-255.
- LI, J. Z., ABSHER, D. M., TANG, H., *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 2008, **319**: 1100-1104.
- LIN, L. F., POSFAI, J., ROBERTS, R. J., *et al.* Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. *Proc Natl Acad Sci U S A*, 2001, **98**: 2740-2745.
- LIN, S. K., LAMBERT, J. R., SCHEMBRI, M. A., *et al.* The prevalence of *Helicobacter pylori* in practising dental staff and dental students. *Aust Dent J*, 1998, **43**: 35-39.
- LIN, S. K., LAMBERT, J. R., SCHEMBRI, M. A., *et al.* *Helicobacter pylori* prevalence in endoscopy and medical staff. *J Gastroenterol Hepatol*, 1994, **9**: 319-324.
- LINTON, D., KARLYSHEV, A. V. and WREN, B. W. Deciphering *Campylobacter jejuni* cell surface interactions from the genome sequence. *Curr Opin Microbiol*, 2001, **4**: 35-40.
- LINZ, B., BALLOUX, F., MOODLEY, Y., *et al.* An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*, 2007, **445**: 915-918.
- LOSTER, B. W., CZESNIKIEWICZ-GUZIŁ, M., BIELANSKI, W., *et al.* Prevalence and characterization of *Helicobacter pylori* infection and colonization in dentists. *J Physiol Pharmacol*, 2009, **60 Suppl 8**: 13-18.
- LOZNIĘWSKI, A., HARISTOY, X., RASKO, D. A., *et al.* Influence of Lewis antigen expression by *Helicobacter pylori* on bacterial internalization by gastric epithelial cells. *Infect Immun*, 2003, **71**: 2902-2906.
- LU, H., MURATA-KAMIYA, N., SAITO, Y., *et al.* Role of partitioning-defective 1/microtubule affinity-regulating kinases in the morphogenetic activity of *Helicobacter pylori* CagA. *J Biol Chem*, 2009, **284**: 23024-23036.
- LU, H. S., SAITO, Y., UMEDA, M., *et al.* Structural and functional diversity in the PAR1b/MARK2-binding region of *Helicobacter pylori* CagA. *Cancer Sci*, 2008, **99**: 2004-2011.
- LU, Y., REDLINGER, T. E., AVITIA, R., *et al.* Isolation and genotyping of *Helicobacter pylori* from untreated municipal wastewater. *Appl Environ Microbiol*, 2002, **68**: 1436-1439.
- LUMAN, W., ZHAO, Y., NG, H. S., *et al.* *Helicobacter pylori* infection is unlikely to be transmitted between partners: evidence from genotypic study in partners of infected patients. *Eur J Gastroenterol Hepatol*, 2002, **14**: 521-528.
- LUNDGREN, A., STROMBERG, E., SJOLING, A., *et al.* Mucosal FOXP3-expressing CD4+ CD25high regulatory T cells in *Helicobacter pylori*-infected patients. *Infect Immun*, 2005, **73**: 523-531.
- LUNDGREN, A., SURI-PAYER, E., ENARSSON, K., *et al.* *Helicobacter pylori*-specific CD4+ CD25high regulatory T cells suppress memory T-cell responses to *H. pylori* in infected individuals. *Infect Immun*, 2003, **71**: 1755-1762.

- LUPETTI, P., HEUSER, J. E., MANETTI, R., *et al.* Oligomeric and subunit structure of the *Helicobacter pylori* vacuolating cytotoxin. *J Cell Biol*, 1996, **133**: 801-807.
- LUZZA, F., PARRELLO, T., MONTELEONE, G., *et al.* Up-regulation of IL-17 is associated with bioactive IL-8 expression in *Helicobacter pylori*-infected human gastric mucosa. *J Immunol*, 2000, **165**: 5332-5337.
- MACHADO, A. M., FIGUEIREDO, C., SERUCA, R., *et al.* *Helicobacter pylori* infection generates genetic instability in gastric cells. *Biochim Biophys Acta*, 2010, **1806**: 58-65.
- MACHADO, J. C., FIGUEIREDO, C., CANEDO, P., *et al.* A proinflammatory genetic profile increases the risk for chronic atrophic gastritis and gastric carcinoma. *Gastroenterology*, 2003, **125**: 364-371.
- MACHADO, J. C., PHAROAH, P., SOUSA, S., *et al.* Interleukin 1B and interleukin 1RN polymorphisms are associated with increased risk of gastric carcinoma. *Gastroenterology*, 2001, **121**: 823-829.
- MAEKITA, T., NAKAZAWA, K., MIHARA, M., *et al.* High levels of aberrant DNA methylation in *Helicobacter pylori*-infected gastric mucosae and its possible association with gastric cancer risk. *Clin Cancer Res*, 2006, **12**: 989-995.
- MALATY, H. M., GRAHAM, D. Y., ISAKSSON, I., *et al.* Co-twin study of the effect of environment and dietary elements on acquisition of *Helicobacter pylori* infection. *Am J Epidemiol*, 1998, **148**: 793-797.
- MALFERTHEINER, P., MEGRAUD, F., O'MORAIN, C., *et al.* Current concepts in the management of *Helicobacter pylori* infection: the Maastricht III Consensus Report. *Gut*, 2007, **56**: 772-781.
- MALFERTHEINER, P., MEGRAUD, F., O'MORAIN, C., *et al.* Current concepts in the management of *Helicobacter pylori* infection--the Maastricht 2-2000 Consensus Report. *Aliment Pharmacol Ther*, 2002, **16**: 167-180.
- MALFERTHEINER, P., SIPPONEN, P., NAUMANN, M., *et al.* *Helicobacter pylori* eradication has the potential to prevent gastric cancer: a state-of-the-art critique. *Am J Gastroenterol*, 2005, **100**: 2100-2115.
- MARSHALL, B. J., GOODWIN, C. S., WARREN, J. R., *et al.* Prospective double-blind trial of duodenal ulcer relapse after eradication of *Campylobacter pylori*. *Lancet*, 1988, **2**: 1437-1442.
- MCCLAIN, M. S., SHAFFER, C. L., ISRAEL, D. A., *et al.* Genome sequence analysis of *Helicobacter pylori* strains associated with gastric ulceration and gastric cancer. *BMC Genomics*, 2009, **10**: 3.
- MEGRAUD, F. H pylori antibiotic resistance: prevalence, importance, and advances in testing. *Gut*, 2004, **53**: 1374-1384.
- MEGRAUD, F. *Helicobacter pylori* and antibiotic resistance. *Gut*, 2007, **56**: 1502.
- MEGRAUD, F., LEHN, N., LIND, T., *et al.* Antimicrobial susceptibility testing of *Helicobacter pylori* in a large multicenter trial: the MACH 2 study. *Antimicrob Agents Chemother*, 1999, **43**: 2747-2752.
- MEGRAUD, F. and LEHOURS, P. *Helicobacter pylori* detection and antimicrobial susceptibility testing. *Clin Microbiol Rev*, 2007, **20**: 280-322.
- MELTON, T., PETERSON, R., REDD, A. J., *et al.* Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am J Hum Genet*, 1995, **57**: 403-414.
- MERA, R., FONTHAM, E. T., BRAVO, L. E., *et al.* Long term follow up of patients treated for *Helicobacter pylori* infection. *Gut*, 2005, **54**: 1536-1540.
- MERRELL, D. S., GOODRICH, M. L., OTTO, G., *et al.* pH-regulated gene expression of the gastric pathogen *Helicobacter pylori*. *Infect Immun*, 2003, **71**: 3529-3539.

- MEYER-TER-VEHN, T., COVACCI, A., KIST, M., *et al.* *Helicobacter pylori* activates mitogen-activated protein kinase cascades and induces expression of the proto-oncogenes c-fos and c-jun. *J Biol Chem*, 2000, **275**: 16064-16072.
- MEYER, J. M., SILLIMAN, N. P., WANG, W., *et al.* Risk factors for *Helicobacter pylori* resistance in the United States: the surveillance of *H. pylori* antimicrobial resistance partnership (SHARP) study, 1993-1999. *Ann Intern Med*, 2002, **136**: 13-24.
- MIEHLKE, S., KIRSCH, C., AGHA-AMIRI, K., *et al.* The *Helicobacter pylori vacA s1, m1* genotype and *cagA* is associated with gastric carcinoma in Germany. *Int J Cancer*, 2000, **87**: 322-327.
- MITCHELL, H. M., ALLY, R., WADEE, A., *et al.* Major differences in the IgG subclass response to *Helicobacter pylori* in the first and third worlds. *Scand J Gastroenterol*, 2002, **37**: 517-522.
- MITCHELL, H. M., LI, Y. Y., HU, P. J., *et al.* Epidemiology of *Helicobacter pylori* in southern China: identification of early childhood as the critical period for acquisition. *J Infect Dis*, 1992, **166**: 149-153.
- MIURA, M., OHNISHI, N., TANAKA, S., *et al.* Differential oncogenic potential of geographically distinct *Helicobacter pylori* CagA isoforms in mice. *Int J Cancer*, 2009, **125**: 2497-2504.
- MIURA, T., FUKUNAGA, T., IGARASHI, T., *et al.* Phylogenetic subtypes of human T-lymphotropic virus type I and their relations to the anthropological background. *Proc Natl Acad Sci U S A*, 1994, **91**: 1124-1127.
- MIYACHI, H., MIKI, I., AOYAMA, N., *et al.* Primary levofloxacin resistance and *gyrA/B* mutations among *Helicobacter pylori* in Japan. *Helicobacter*, 2006, **11**: 243-249.
- MIZOGUCHI, H., FUJIOKA, T. and NASU, M. Evidence for viability of coccoid forms of *Helicobacter pylori*. *J Gastroenterol*, 1999, **34 Suppl 11**: 32-36.
- MOHAMED, R., HANAFIAH, A., ROSE, I. M., *et al.* *Helicobacter pylori cagA* gene variants in Malaysians of different ethnicity. *Eur J Clin Microbiol Infect Dis*, 2009, **28**: 865-869.
- MOLINARI, M., SALIO, M., GALLI, C., *et al.* Selective inhibition of Ii-dependent antigen presentation by *Helicobacter pylori* toxin VacA. *J Exp Med*, 1998, **187**: 135-140.
- MOLLENHAUER-REKTORSCHKE, M., HANAUER, G., SACHS, G., *et al.* Expression of UreI is required for intragastric transit and colonization of gerbil gastric mucosa by *Helicobacter pylori*. *Res Microbiol*, 2002, **153**: 659-666.
- MONOT, M., HONORE, N., GARNIER, T., *et al.* On the origin of leprosy. *Science*, 2005, **308**: 1040-1042.
- MONSTEIN, H. J. and JONASSON, J. Differential virulence-gene mRNA expression in coccoid forms of *Helicobacter pylori*. *Biochem Biophys Res Commun*, 2001, **285**: 530-536.
- MONTALBAN, C. and NORMAN, F. Treatment of gastric mucosa-associated lymphoid tissue lymphoma: *Helicobacter pylori* eradication and beyond. *Expert Rev Anticancer Ther*, 2006, **6**: 361-371.
- MOODLEY, Y., LINZ, B., YAMAOKA, Y., *et al.* The peopling of the Pacific from a bacterial perspective. *Science*, 2009, **323**: 527-530.
- MOORE, R. A., BECKTHOLD, B., WONG, S., *et al.* Nucleotide sequence of the *gyrA* gene and characterization of ciprofloxacin-resistant mutants of *Helicobacter pylori*. *Antimicrob Agents Chemother*, 1995, **39**: 107-111.
- MORAN, A. P., LINDNER, B. and WALSH, E. J. Structural characterization of the lipid A component of *Helicobacter pylori* rough- and smooth-form lipopolysaccharides. *J Bacteriol*, 1997, **179**: 6453-6463.

- MORAN, A. P. and PRENDERGAST, M. M. Molecular mimicry in *Campylobacter jejuni* and *Helicobacter pylori* lipopolysaccharides: contribution of gastrointestinal infections to autoimmunity. *J Autoimmun*, 2001, **16**: 241-256.
- MORELLI, G., DIDELOT, X., KUSECEK, B., *et al.* Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families. *PLoS Genet*, 2010, **6**: e1001036.
- MUOTIALA, A., HELANDER, I. M., PYHALA, L., *et al.* Low biological activity of *Helicobacter pylori* lipopolysaccharide. *Infect Immun*, 1992, **60**: 1714-1716.
- MURATA-KAMIYA, N., KURASHIMA, Y., TEISHIKATA, Y., *et al.* *Helicobacter pylori* CagA interacts with E-cadherin and deregulates the beta-catenin signal that promotes intestinal transdifferentiation in gastric epithelial cells. *Oncogene*, 2007, **26**: 4617-4626.
- NAHAR, S., MUKHOPADHYAY, A. K., KHAN, R., *et al.* Antimicrobial susceptibility of *Helicobacter pylori* strains isolated in Bangladesh. *J Clin Microbiol*, 2004, **42**: 4856-4858.
- NAITO, M., YAMAZAKI, T., TSUTSUMI, R., *et al.* Influence of EPIYA-repeat polymorphism on the phosphorylation-dependent biological activity of *Helicobacter pylori* CagA. *Gastroenterology*, 2006, **130**: 1181-1190.
- NAKAMURA, S., MATSUMOTO, T., SUEKANE, H., *et al.* Long-term clinical outcome of *Helicobacter pylori* eradication for gastric mucosa-associated lymphoid tissue lymphoma with a reference to second-line treatment. *Cancer*, 2005, **104**: 532-540.
- NDIP, R. N., MALANGE, A. E., AKOACHERE, J. F., *et al.* *Helicobacter pylori* antigens in the faeces of asymptomatic children in the Buea and Limbe health districts of Cameroon: a pilot study. *Trop Med Int Health*, 2004, **9**: 1036-1040.
- NILSSON, H. O., BLOM, J., ABU-AL-SOUD, W., *et al.* Effect of cold starvation, acid stress, and nutrients on metabolic activity of *Helicobacter pylori*. *Appl Environ Microbiol*, 2002, **68**: 11-19.
- NOMURA, A. M., KOLONEL, L. N., MIKI, K., *et al.* *Helicobacter pylori*, pepsinogen, and gastric adenocarcinoma in Hawaii. *J Infect Dis*, 2005, **191**: 2075-2081.
- NONAKA, L., CONNELL, S. R. and TAYLOR, D. E. 16S rRNA Mutations That Confer Tetracycline Resistance in *Helicobacter pylori* Decrease Drug Binding in *Escherichia coli* Ribosomes. *J Bacteriol*, 2005, **187**: 3708-3712.
- NORMARK, S., NILSSON, C., NORMARK, B. H., *et al.* Persistent infection with *Helicobacter pylori* and the development of gastric cancer. *Adv Cancer Res*, 2003, **90**: 63-89.
- O'TOOLE, P. W., SNELLING, W. J., CANCHAYA, C., *et al.* Comparative genomics and proteomics of *Helicobacter mustelae*, an ulcerogenic and carcinogenic gastric pathogen. *BMC Genomics*, 2010, **11**: 164.
- ODENBREIT, S., TILL, M., HOFREUTER, D., *et al.* Genetic and functional characterization of the *alpAB* gene locus essential for the adhesion of *Helicobacter pylori* to human gastric tissue. *Mol Microbiol*, 1999, **31**: 1537-1548.
- OGIWARA, H., GRAHAM, D. Y. and YAMAOKA, Y. *vacA* i-region subtyping. *Gastroenterology*, 2008, **134**: 1267; author reply 1268.
- OHNO, T., SUGIMOTO, M., NAGASHIMA, A., *et al.* Relationship between *Helicobacter pylori* *hopQ* genotype and clinical outcome in Asian and Western populations. *J Gastroenterol Hepatol*, 2009, **24**: 462-468.
- OLBERMANN, P., JOSEPHANS, C., MOODLEY, Y., *et al.* A global overview of the genetic and functional diversity in the *Helicobacter pylori* *cag* pathogenicity island. *PLoS Genet*, 2010, **6**: e1001069.

- OLDANI, A., CORMONT, M., HOFMAN, V., *et al.* *Helicobacter pylori* counteracts the apoptotic action of its VacA toxin by injecting the CagA protein into gastric epithelial cells. *PLoS Pathog*, 2009, **5**: e1000603.
- OPPENHEIMER, S. J. and RICHARDS, M. Polynesian origins. Slow boat to Melanesia? *Nature*, 2001, **410**: 166-167.
- OTTEMANN, K. M. and LOWENTHAL, A. C. *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. *Infect Immun*, 2002, **70**: 1984-1990.
- OWEN, R. J. and XERRY, J. Tracing clonality of *Helicobacter pylori* infecting family members from analysis of DNA sequences of three housekeeping genes (*ureI*, *atpA* and *ahpC*), deduced amino acid sequences, and pathogenicity-associated markers (*cagA* and *vacA*). *J Med Microbiol*, 2003, **52**: 515-524.
- PANAYOTOPOULOU, E. G., SGOURAS, D. N., PAPADAKOS, K., *et al.* Strategy to characterize the number and type of repeating EPIYA phosphorylation motifs in the carboxyl terminus of CagA protein in *Helicobacter pylori* clinical isolates. *J Clin Microbiol*, 2007, **45**: 488-495.
- PANNELL, J. R. and PUJOL, B. The paradoxical spread of a new Y chromosome - a novel explanation. *Trends Ecol Evol*, 2009, **24**: 59-63.
- PANTHEL, K., JECHLINGER, W., MATIS, A., *et al.* Generation of *Helicobacter pylori* ghosts by PhiX protein E-mediated inactivation and their evaluation as vaccine candidates. *Infect Immun*, 2003, **71**: 109-116.
- PAPINI, E., DE BERNARD, M., MILIA, E., *et al.* Cellular vacuoles induced by *Helicobacter pylori* originate from late endosomal compartments. *Proc Natl Acad Sci U S A*, 1994, **91**: 9720-9724.
- PARK, S. Y., YOO, E. J., CHO, N. Y., *et al.* Comparison of CpG island hypermethylation and repetitive DNA hypomethylation in premalignant stages of gastric cancer, stratified for *Helicobacter pylori* infection. *J Pathol*, 2009, **219**: 410-416.
- PARKHILL, J., WREN, B. W., MUNGALL, K., *et al.* The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature*, 2000, **403**: 665-668.
- PARKIN, D. M., BRAY, F., FERLAY, J., *et al.* Global cancer statistics, 2002. *CA Cancer J Clin*, 2005, **55**: 74-108.
- PARSONNET, J. The incidence of *Helicobacter pylori* infection. *Aliment Pharmacol Ther*, 1995, **9 Suppl 2**: 45-51.
- PARSONNET, J., FRIEDMAN, G. D., VANDERSTEEN, D. P., *et al.* *Helicobacter pylori* infection and the risk of gastric carcinoma. *N Engl J Med*, 1991, **325**: 1127-1131.
- PATTIS, I., WEISS, E., LAUGKS, R., *et al.* The *Helicobacter pylori* CagF protein is a type IV secretion chaperone-like molecule that binds close to the C-terminal secretion signal of the CagA effector protein. *Microbiology*, 2007, **153**: 2896-2909.
- PENG, W., ZHOU, X. and CROMPTON, D. W. Ascariasis in China. *Adv Parasitol*, 1998, **41**: 109-148.
- PERRY, S., DE LA LUZ SANCHEZ, M., YANG, S., *et al.* Gastroenteritis and transmission of *Helicobacter pylori* infection in households. *Emerg Infect Dis*, 2006, **12**: 1701-1708.
- PFLOCK, M., KENNARD, S., FINSTERER, N., *et al.* Acid-responsive gene regulation in the human pathogen *Helicobacter pylori*. *J Biotechnol*, 2006, **126**: 52-60.
- PINTO, A. V., MATHIEU, A., MARSIN, S., *et al.* Suppression of homologous and homeologous recombination by the bacterial MutS2 protein. *Mol Cell*, 2005, **17**: 113-120.

- PIQUERES, P., MORENO, Y., ALONSO, J. L., *et al.* A combination of direct viable count and fluorescent in situ hybridization for estimating *Helicobacter pylori* cell viability. *Res Microbiol*, 2006, **157**: 345-349.
- POMS, R. E. and TATINI, S. R. Survival of *Helicobacter pylori* in ready-to-eat foods at 4 degrees C. *Int J Food Microbiol*, 2001, **63**: 281-286.
- PRAZERES MAGALHAES, P., DE MAGALHAES QUEIROZ, D. M., CAMPOS BARBOSA, D. V., *et al.* *Helicobacter pylori* primary resistance to metronidazole and clarithromycin in Brazil. *Antimicrob Agents Chemother*, 2002, **46**: 2021-2023.
- QUERALT, N., BARTOLOME, R. and ARAUJO, R. Detection of *Helicobacter pylori* DNA in human faeces and water with different levels of faecal pollution in the north-east of Spain. *J Appl Microbiol*, 2005, **98**: 889-895.
- QUINTANA-MURCI, L., HARMANT, C., QUACH, H., *et al.* Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. *Am J Hum Genet*, 2010, **86**: 611-620.
- RAYMOND, J., LAMARQUE, D., KALACH, N., *et al.* High level of antimicrobial resistance in French *Helicobacter pylori* isolates. *Helicobacter*, 2010, **15**: 21-27.
- RAYMOND, J., THIBERGE, J. M., CHEVALIER, C., *et al.* Genetic and transmission analysis of *Helicobacter pylori* strains within a family. *Emerg Infect Dis*, 2004, **10**: 1816-1821.
- RAYMOND, J., THIBERGE, J. M., KALACH, N., *et al.* Using macro-arrays to study routes of infection of *Helicobacter pylori* in three families. *PLoS One*, 2008, **3**: e2259.
- RAZAFINDRAZAKA, H., RICAUT, F. X., COX, M. P., *et al.* Complete mitochondrial DNA sequences provide new insights into the Polynesian motif and the peopling of Madagascar. *Eur J Hum Genet*, 2010, **18**: 575-581.
- REDD, A. J., TAKEZAKI, N., SHERRY, S. T., *et al.* Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol*, 1995, **12**: 604-615.
- REGUEIRO, M., MIRABAL, S., LACAU, H., *et al.* Austronesian genetic signature in East African Madagascar and Polynesia. *J Hum Genet*, 2008, **53**: 106-120.
- REPPING, S., VAN DAALLEN, S. K., BROWN, L. G., *et al.* High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat Genet*, 2006, **38**: 463-467.
- REYES-LEON, A., ATHERTON, J. C., ARGENT, R. H., *et al.* Heterogeneity in the activity of Mexican *Helicobacter pylori* strains in gastric epithelial cells and its association with diversity in the *cagA* gene. *Infect Immun*, 2007, **75**: 3445-3454.
- REYRAT, J. M., LANZAVECCHIA, S., LUPETTI, P., *et al.* 3D imaging of the 58 kDa cell binding subunit of the *Helicobacter pylori* cytotoxin. *J Mol Biol*, 1999, **290**: 459-470.
- RHEAD, J. L., LETLEY, D. P., MOHAMMADI, M., *et al.* A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology*, 2007, **133**: 926-936.
- RICHARDS, C. L., PENNING, S. C. and DONOVAN, L. A. Habitat range and phenotypic variation in salt marsh plants. *Plant Ecology*, 2005, **176**: 263-273.
- RICHARDS, M., OPPENHEIMER, S. and SYKES, B. mtDNA suggests Polynesian origins in Eastern Indonesia. *Am J Hum Genet*, 1998, **63**: 1234-1236.
- RIMBARA, E., NOGUCHI, N., KAWAI, T., *et al.* Mutations in penicillin-binding proteins 1, 2 and 3 are responsible for amoxicillin resistance in *Helicobacter pylori*. *J Antimicrob Chemother*, 2008, **61**: 995-998.
- ROBINSON, K., KENEFECK, R., PIDGEON, E. L., *et al.* *Helicobacter pylori*-induced peptic ulcer disease is associated with inadequate regulatory T cell responses. *Gut*, 2008, **57**: 1375-1385.

- ROKKAS, T., SECHOPOULOS, P., PISTIOLAS, D., *et al.* *Helicobacter pylori* infection and gastric histology in first-degree relatives of gastric cancer patients: a meta-analysis. *Eur J Gastroenterol Hepatol*, 2010, **22**: 1128-1133.
- ROMANO, M., MARMO, R., CUOMO, A., *et al.* Pretreatment antimicrobial susceptibility testing is cost saving in the eradication of *Helicobacter pylori*. *Clin Gastroenterol Hepatol*, 2003, **1**: 273-278.
- ROMO-GONZALEZ, C., SALAMA, N. R., BURGONO-FERREIRA, J., *et al.* Differences in genome content among *Helicobacter pylori* isolates from patients with gastritis, duodenal ulcer, or gastric cancer reveal novel disease-associated genes. *Infect Immun*, 2009, **77**: 2201-2211.
- ROUMAGNAC, P., WEILL, F. X., DOLECEK, C., *et al.* Evolutionary history of *Salmonella typhi*. *Science*, 2006, **314**: 1301-1304.
- RUDI, J., KOLB, C., MAIWALD, M., *et al.* Diversity of *Helicobacter pylori vacA* and *cagA* genes and relationship to VacA and CagA protein expression, cytotoxin production, and associated diseases. *J Clin Microbiol*, 1998, **36**: 944-948.
- RUSSO, F., NOTARNICOLA, M., DI MATTEO, G., *et al.* Detection of *Helicobacter pylori cagA* gene by polymerase chain reaction in faecal samples. *Eur J Gastroenterol Hepatol*, 1999, **11**: 251-256.
- SAADAT, I., HIGASHI, H., OBUSE, C., *et al.* *Helicobacter pylori* CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. *Nature*, 2007, **447**: 330-333.
- SAGART, L., BLENCH, R. and SANCHEZ-MAZAS, A. The peopling of East Asia: putting together archaeology, linguistics and genetics. 1e éd. London: Routledge Curzon, 2005, I-360 p.
- SALAMA, N., GUILLEMIN, K., MCDANIEL, T. K., *et al.* A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci U S A*, 2000, **97**: 14668-14673.
- SALAMA, N. R., OTTO, G., TOMPKINS, L., *et al.* Vacuolating cytotoxin of *Helicobacter pylori* plays a role during colonization in a mouse model of infection. *Infect Immun*, 2001, **69**: 730-736.
- SALAUN, L., LINZ, B., SUERBAUM, S., *et al.* The diversity within an expanded and redefined repertoire of phase-variable genes in *Helicobacter pylori*. *Microbiology*, 2004, **150**: 817-830.
- SALEM, A. H., BADR, F. M., GABALLAH, M. F., *et al.* The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. *Am J Hum Genet*, 1996, **59**: 741-743.
- SCHMIDT, H. M., GOH, K. L., FOCK, K. M., *et al.* Distinct *cagA* EPIYA motifs are associated with ethnic diversity in Malaysia and Singapore. *Helicobacter*, 2009, **14**: 256-263.
- SCHMITT, W. and HAAS, R. Genetic analysis of the *Helicobacter pylori* vacuolating cytotoxin: structural similarities with the IgA protease type of exported protein. *Mol Microbiol*, 1994, **12**: 307-319.
- SCHNEIDER, B. G., PENG, D. F., CAMARGO, M. C., *et al.* Promoter DNA hypermethylation in gastric biopsies from subjects at high and low risk for gastric cancer. *Int J Cancer*, 2010, **127**: 2588-2597.
- SCHOFIELD, M. J. and HSIEH, P. DNA mismatch repair: molecular mechanisms and biological function. *Annu Rev Microbiol*, 2003, **57**: 579-608.
- SCHREIBER, S., KONRADT, M., GROLL, C., *et al.* The spatial orientation of *Helicobacter pylori* in the gastric mucus. *Proc Natl Acad Sci U S A*, 2004, **101**: 5024-5029.



- SCHREIBER, S., NGUYEN, T. H., STUBEN, M., *et al.* Demonstration of a pH gradient in the gastric gland of the acid-secreting guinea pig mucosa. *Am J Physiol Gastrointest Liver Physiol*, 2000, **279**: G597-604.
- SCHREIBER, S., STUBEN, M., JOSENHANS, C., *et al.* In vivo distribution of *Helicobacter felis* in the gastric mucus of the mouse: experimental method and results. *Infect Immun*, 1999, **67**: 5151-5156.
- SCHWARZ, S., MORELLI, G., KUSECEK, B., *et al.* Horizontal versus familial transmission of *Helicobacter pylori*. *PLoS Pathog*, 2008, **4**: e1000180.
- SECK, A., MBENGUE, M., GASSAMA-SOW, A., *et al.* Antibiotic susceptibility of *Helicobacter pylori* isolates in Dakar, Senegal. *J Infect Dev Ctries*, 2009, **3**: 137-140.
- SEGAL, E. D., CHA, J., LO, J., *et al.* Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proc Natl Acad Sci U S A*, 1999, **96**: 14559-14564.
- SELBACH, M., MOESE, S., HURWITZ, R., *et al.* The *Helicobacter pylori* CagA protein induces cortactin dephosphorylation and actin rearrangement by c-Src inactivation. *EMBO J*, 2003, **22**: 515-528.
- SELBACH, M., PAUL, F. E., BRANDT, S., *et al.* Host cell interactome of tyrosine-phosphorylated bacterial proteins. *Cell Host Microbe*, 2009, **5**: 397-403.
- SEMINO, O., PASSARINO, G., OEFNER, P. J., *et al.* The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science*, 2000, **290**: 1155-1159.
- SHACKELTON, L. A., RAMBAUT, A., PYBUS, O. G., *et al.* JC virus evolution and its association with human populations. *J Virol*, 2006, **80**: 9928-9933.
- SHARMA, S. A., TUMMURU, M. K., BLASER, M. J., *et al.* Activation of IL-8 gene expression by *Helicobacter pylori* is regulated by transcription factor nuclear factor-kappa B in gastric epithelial cells. *J Immunol*, 1998, **160**: 2401-2407.
- SHIRAI, K., OHMIYA, N., TAGUCHI, A., *et al.* Interleukin-8 gene polymorphism associated with susceptibility to non-cardia gastric carcinoma with microsatellite instability. *J Gastroenterol Hepatol*, 2006, **21**: 1129-1135.
- SIMOONS-SMIT, I. M., APPELMELK, B. J., VERBOOM, T., *et al.* Typing of *Helicobacter pylori* with monoclonal antibodies against Lewis antigens in lipopolysaccharide. *J Clin Microbiol*, 1996, **34**: 2196-2200.
- SKOULOUBRIS, S., THIBERGE, J. M., LABIGNE, A., *et al.* The *Helicobacter pylori* UreI protein is not involved in urease activity but is essential for bacterial survival in vivo. *Infect Immun*, 1998, **66**: 4517-4521.
- SMOOT, D. T., MOBLEY, H. L., CHIPPENDALE, G. R., *et al.* *Helicobacter pylori* urease activity is toxic to human gastric epithelial cells. *Infect Immun*, 1990, **58**: 1992-1994.
- SMYTHIES, L. E., WAITES, K. B., LINDSEY, J. R., *et al.* *Helicobacter pylori*-induced mucosal inflammation is Th1 mediated and exacerbated in IL-4, but not IFN-gamma, gene-deficient mice. *J Immunol*, 2000, **165**: 1022-1029.
- SOARES, P., RITO, T., TREJAUT, J., *et al.* Ancient voyaging and Polynesian origins. *Am J Hum Genet*, 2011, **88**: 239-247.
- SOARES, P., TREJAUT, J. A., LOO, J. H., *et al.* Climate change and postglacial human dispersals in southeast Asia. *Mol Biol Evol*, 2008, **25**: 1209-1218.
- SOLNICK, J. V. and SCHAUER, D. B. Emergence of diverse *Helicobacter* species in the pathogenesis of gastric and enterohepatic diseases. *Clin Microbiol Rev*, 2001, **14**: 59-97.
- SORBERG, M., NILSSON, M., HANBERGER, H., *et al.* Morphologic conversion of *Helicobacter pylori* from bacillary to coccoid form. *Eur J Clin Microbiol Infect Dis*, 1996, **15**: 216-219.

- SPRIGGS, M. Prehistoric Mongoloid Dispersals. *What is Southeast Asian about Lapita* éd. by A. TAKERU and E. J. E. SZATHMARY. London: Oxford University Press, 1996, p. 322-346
- STYER, C. M., HANSEN, L. M., COOKE, C. L., *et al.* Expression of the BabA adhesin during experimental infection with *Helicobacter pylori*. *Infect Immun*, 2010, **78**: 1593-1600.
- SUERBAUM, S., JOSEPHANS, C., STERZENBACH, T., *et al.* The complete genome sequence of the carcinogenic bacterium *Helicobacter hepaticus*. *Proc Natl Acad Sci U S A*, 2003, **100**: 7901-7906.
- SUERBAUM, S. and MICHETTI, P. *Helicobacter pylori* infection. *N Engl J Med*, 2002, **347**: 1175-1186.
- SUERBAUM, S., SMITH, J. M., BAPUMIA, K., *et al.* Free recombination within *Helicobacter pylori*. *Proc Natl Acad Sci U S A*, 1998, **95**: 12619-12624.
- SUGIMOTO, M. and YAMAOKA, Y. The association of *vacA* genotype and *Helicobacter pylori*-related disease in Latin American and African populations. *Clin Microbiol Infect*, 2009, **15**: 835-842.
- SUGIMOTO, M., ZALI, M. R. and YAMAOKA, Y. The association of *vacA* genotypes and *Helicobacter pylori*-related gastroduodenal diseases in the Middle East. *Eur J Clin Microbiol Infect Dis*, 2009, **28**: 1227-1236.
- SULTAN, S. E. Phenotypic plasticity for fitness components in *Polygonum* species of contrasting ecological breadth. *Ecology*, 2001 **82**: 328-343.
- SUNDRUD, M. S., TORRES, V. J., UNUTMAZ, D., *et al.* Inhibition of primary human T cell proliferation by *Helicobacter pylori* vacuolating toxin (VacA) is independent of VacA effects on IL-2 secretion. *Proc Natl Acad Sci U S A*, 2004, **101**: 7727-7732.
- SUPAJATURA, V., USHIO, H., WADA, A., *et al.* Cutting edge: VacA, a vacuolating cytotoxin of *Helicobacter pylori*, directly activates mast cells for migration and production of proinflammatory cytokines. *J Immunol*, 2002, **168**: 2603-2607.
- SUZUKI, M., MIMURO, H., SUZUKI, T., *et al.* Interaction of CagA with Crk plays an important role in *Helicobacter pylori*-induced loss of gastric epithelial cell adhesion. *J Exp Med*, 2005, **202**: 1235-1247.
- SYKES, B., LEIBOFF, A., LOW-BEER, J., *et al.* The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *Am J Hum Genet*, 1995, **57**: 1463-1475.
- SZABO, I., BRUTSCHE, S., TOMBOLA, F., *et al.* Formation of anion-selective channels in the cell plasma membrane by the toxin VacA of *Helicobacter pylori* is required for its biological activity. *EMBO J*, 1999, **18**: 5517-5527.
- TAKEUCHI, H., SHIRAI, M., AKADA, J. K., *et al.* Nucleotide sequence and characterization of *cdrA*, a cell division-related gene of *Helicobacter pylori*. *J Bacteriol*, 1998, **180**: 5263-5268.
- TAMMER, I., BRANDT, S., HARTIG, R., *et al.* Activation of Abl by *Helicobacter pylori*: a novel kinase for CagA and crucial mediator of host cell scattering. *Gastroenterology*, 2007, **132**: 1309-1319.
- TANKOVIC, J., LAMARQUE, D., DELCHIER, J. C., *et al.* Frequent association between alteration of the *rdxA* gene and metronidazole resistance in French and North African isolates of *Helicobacter pylori*. *Antimicrob Agents Chemother*, 2000, **44**: 608-613.
- TANKOVIC, J., LASCOLS, C., SCULO, Q., *et al.* Single and double mutations in *gyrA* but not in *gyrB* are associated with low- and high-level fluoroquinolone resistance in *Helicobacter pylori*. *Antimicrob Agents Chemother*, 2003, **47**: 3942-3944.

- TARTAGLIA, M., NIEMEYER, C. M., FRAGALE, A., *et al.* Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat Genet*, 2003, **34**: 148-150.
- TAY, C. Y., MITCHELL, H., DONG, Q., *et al.* Population structure of *Helicobacter pylori* among ethnic groups in Malaysia: recent acquisition of the bacterium by the Malay population. *BMC Microbiol*, 2009, **9**: 126.
- TAYLOR, D. E., EATON, M., CHANG, N., *et al.* Construction of a *Helicobacter pylori* genome map and demonstration of diversity at the genome level. *J Bacteriol*, 1992, **174**: 6800-6806.
- TELFORD, J. L., GHIARA, P., DELL'ORCO, M., *et al.* Gene structure of the *Helicobacter pylori* cytotoxin and evidence of its key role in gastric disease. *J Exp Med*, 1994, **179**: 1653-1658.
- THIBERGE, J. M., BOURSAUX-EUDE, C., LEHOURS, P., *et al.* From array-based hybridization of *Helicobacter pylori* isolates to the complete genome sequence of an isolate associated with MALT lymphoma. *BMC Genomics*, 2010, **11**: 368.
- TOMB, J. F., WHITE, O., KERLAVAGE, A. R., *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, 1997, **388**: 539-547.
- TORACCHIO, S., ACETO, G. M., MARIANI-COSTANTINI, R., *et al.* Identification of a novel mutation affecting domain V of the 23S rRNA gene in *Helicobacter pylori*. *Helicobacter*, 2004, **9**: 396-399.
- TORACCHIO, S. and MARZIO, L. Primary and secondary antibiotic resistance of *Helicobacter pylori* strains isolated in central Italy during the years 1998-2002. *Dig Liver Dis*, 2003, **35**: 541-545.
- TORRES, V. J., VANCOMPERNOLLE, S. E., SUNDRUD, M. S., *et al.* *Helicobacter pylori* vacuolating cytotoxin inhibits activation-induced proliferation of human T and B lymphocyte subsets. *J Immunol*, 2007, **179**: 5433-5440.
- TORRONI, A., ACHILLI, A., MACAULAY, V., *et al.* Harvesting the fruit of the human mtDNA tree. *Trends Genet*, 2006, **22**: 339-345.
- TOYODA, T., TSUKAMOTO, T., TAKASU, S., *et al.* Anti-inflammatory effects of caffeic acid phenethyl ester (CAPE), a nuclear factor-kappaB inhibitor, on *Helicobacter pylori*-induced gastritis in Mongolian gerbils. *Int J Cancer*, 2009, **125**: 1786-1795.
- TREJAUT, J. A., KIVISILD, T., LOO, J. H., *et al.* Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol*, 2005, **3**: e247.
- TSUTSUMI, R., HIGASHI, H., HIGUCHI, M., *et al.* Attenuation of *Helicobacter pylori* CagA x SHP-2 signaling by interaction between CagA and C-terminal Src kinase. *J Biol Chem*, 2003, **278**: 3664-3670.
- TUMMURU, M. K., SHARMA, S. A. and BLASER, M. J. *Helicobacter pylori* *picB*, a homologue of the Bordetella pertussis toxin secretion protein, is required for induction of IL-8 in gastric epithelial cells. *Mol Microbiol*, 1995, **18**: 867-876.
- TUO, B., SONG, P., WEN, G., *et al.* *Helicobacter pylori* vacuolating cytotoxin inhibits duodenal bicarbonate secretion by a histamine-dependent mechanism in mice. *J Infect Dis*, 2009, **199**: 505-512.
- TURUTOGLU, H. and MUDUL, S. Investigation of *Helicobacter pylori* in raw sheep milk samples. *J Vet Med B Infect Dis Vet Public Health*, 2002, **49**: 308-309.
- UCHIDA, T., NGUYEN, L. T., TAKAYAMA, A., *et al.* Analysis of virulence factors of *Helicobacter pylori* isolated from a Vietnamese population. *BMC Microbiol*, 2009, **9**: 175.
- UEMURA, N., OKAMOTO, S., YAMAMOTO, S., *et al.* *Helicobacter pylori* infection and the development of gastric cancer. *N Engl J Med*, 2001, **345**: 784-789.

- UMEDA, M., MURATA-KAMIYA, N., SAITO, Y., *et al.* *Helicobacter pylori* CagA causes mitotic impairment and induces chromosomal instability. *J Biol Chem*, 2009, **284**: 22166-22172.
- UNDERHILL, P. A. and KIVISILD, T. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. 2007, **41**: 539-564.
- UNDERHILL, P. A., PASSARINO, G., LIN, A. A., *et al.* Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. *Hum Mutat*, 2001, **17**: 271-280.
- VAN DOORN, L. J., FIGUEIREDO, C., MEGRAUD, F., *et al.* Geographic distribution of *vacA* allelic types of *Helicobacter pylori*. *Gastroenterology*, 1999, **116**: 823-830.
- VERIN, P. The History of Civilisation in North Madagascar *Journal des Africanistes*, 1987, **57**: 291-292.
- VERSALOVIC, J., SHORTRIDGE, D., KIBLER, K., *et al.* Mutations in 23S rRNA are associated with clarithromycin resistance in *Helicobacter pylori*. *Antimicrob Agents Chemother*, 1996, **40**: 477-480.
- VINCENT, P., GOTTRAND, F., PERNES, P., *et al.* High prevalence of *Helicobacter pylori* infection in cohabiting children. Epidemiology of a cluster, with special emphasis on molecular typing. *Gut*, 1994, **35**: 313-316.
- VOS, M. and DIDELOT, X. A comparison of homologous recombination rates in bacteria and archaea. *ISME J*, 2009, **3**: 199-208.
- WANG, G., WILSON, T. J., JIANG, Q., *et al.* Spontaneous mutations that confer antibiotic resistance in *Helicobacter pylori*. *Antimicrob Agents Chemother*, 2001, **45**: 727-733.
- WEEKS, D. L., ESKANDARI, S., SCOTT, D. R., *et al.* A H<sup>+</sup>-gated urea channel: the link between *Helicobacter pylori* urease and gastric colonization. *Science*, 2000, **287**: 482-485.
- WEN, Y., MARCUS, E. A., MATRUBUTHAM, U., *et al.* Acid-adaptive genes of *Helicobacter pylori*. *Infect Immun*, 2003, **71**: 5921-5939.
- VIALA, J., CHAPUT, C., BONECA, I. G., *et al.* Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori* *cag* pathogenicity island. *Nat Immunol*, 2004, **5**: 1166-1174.
- WIEDEMANN, T., LOELL, E., MUELLER, S., *et al.* *Helicobacter pylori* *cag*-Pathogenicity island-dependent early immunological response triggers later precancerous gastric changes in Mongolian gerbils. *PLoS One*, 2009, **4**: e4754.
- WILLHITE, D. C. and BLANKE, S. R. *Helicobacter pylori* vacuolating cytotoxin enters cells, localizes to the mitochondria, and induces mitochondrial membrane permeability changes correlated to toxin channel activity. *Cell Microbiol*, 2004, **6**: 143-154.
- WILLHITE, D. C., COVER, T. L. and BLANKE, S. R. Cellular vacuolation and mitochondrial cytochrome c release are independent outcomes of *Helicobacter pylori* vacuolating cytotoxin activity that are each dependent on membrane channel formation. *J Biol Chem*, 2003, **278**: 48204-48209.
- WIRTH, H. P., YANG, M., KARITA, M., *et al.* Expression of the human cell surface glycoconjugates Lewis x and Lewis y by *Helicobacter pylori* isolates is related to *cagA* status. *Infect Immun*, 1996, **64**: 4598-4605.
- WIRTH, T., WANG, X., LINZ, B., *et al.* Distinguishing human ethnic groups by means of sequences from *Helicobacter pylori*: lessons from Ladakh. *Proc Natl Acad Sci U S A*, 2004, **101**: 4746-4751.
- WOLFFE, A. P. and MATZKE, M. A. Epigenetics: regulation through repression. *Science*, 1999, **286**: 481-486.

- WONG, B. C., LAM, S. K., WONG, W. M., *et al.* *Helicobacter pylori* eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. *JAMA*, 2004, **291**: 187-194.
- WU, C. Y., KUO, K. N., WU, M. S., *et al.* Early *Helicobacter pylori* eradication decreases risk of gastric cancer in patients with peptic ulcer disease. *Gastroenterology*, 2009, **137**: 1641-1648 e1641-1642.
- WU, J. J., SHEU, B. S., HUANG, A. H., *et al.* Characterization of *flgK* gene and FlgK protein required for *H. pylori* colonization--from cloning to clinical relevance. *World J Gastroenterol*, 2006, **12**: 3989-3993.
- WUNDISCH, T., THIEDE, C., MORGNER, A., *et al.* Long-term follow-up of gastric MALT lymphoma after *Helicobacter pylori* eradication. *J Clin Oncol*, 2005, **23**: 8018-8024.
- XIA, Y., YAMAOKA, Y., ZHU, Q., *et al.* A comprehensive sequence and disease correlation analyses for the C-terminal region of CagA protein of *Helicobacter pylori*. *PLoS ONE*, 2009, **4**: e7736.
- YAMAOKA, Y. Mechanisms of disease: *Helicobacter pylori* virulence factors. *Nat Rev Gastroenterol Hepatol*, 2010, **7**: 629-641.
- YAMAOKA, Y., KODAMA, T., GUTIERREZ, O., *et al.* Relationship between *Helicobacter pylori* *iceA*, *cagA*, and *vacA* status and clinical outcome: studies in four different countries. *J Clin Microbiol*, 1999, **37**: 2274-2279.
- YAMAOKA, Y., KWON, D. H. and GRAHAM, D. Y. A M(r) 34,000 proinflammatory outer membrane protein (*oipA*) of *Helicobacter pylori*. *Proc Natl Acad Sci U S A*, 2000, **97**: 7533-7538.
- YAMAOKA, Y., OJO, O., FUJIMOTO, S., *et al.* *Helicobacter pylori* outer membrane proteins and gastroduodenal disease. *Gut*, 2006, **55**: 775-781.
- YANG, Y. J., YANG, H. B., WU, J. J., *et al.* Persistent *H. pylori* colonization in early acquisition age of mice related with higher gastric sialylated Lewis x, IL-10, but lower interferon-gamma expressions. *J Biomed Sci*, 2008.
- YOGO, Y., SUGIMOTO, C., ZHENG, H. Y., *et al.* JC virus genotyping offers a new paradigm in the study of human populations. *Rev Med Virol*, 2004, **14**: 179-191.
- YOKOYAMA, K., HIGASHI, H., ISHIKAWA, S., *et al.* Functional antagonism between *Helicobacter pylori* CagA and vacuolating toxin VacA in control of the NFAT signaling pathway in gastric epithelial cells. *Proc Natl Acad Sci U S A*, 2005, **102**: 9661-9666.
- YOUNG, K. A., AKYON, Y., RAMPTON, D. S., *et al.* Quantitative culture of *Helicobacter pylori* from gastric juice: the potential for transmission. *J Med Microbiol*, 2000, **49**: 343-347.
- ZHANG, D. H., ZHOU, L. Y., LIN, S. R., *et al.* Recent changes in the prevalence of *Helicobacter pylori* infection among children and adults in high- or low-incidence regions of gastric cancer in China. *Chin Med J*, 2009, **122**: 1759-1763.
- ZHENG, P. Y. and JONES, N. L. *Helicobacter pylori* strains expressing the vacuolating cytotoxin interrupt phagosome maturation in macrophages by recruiting and retaining TACO (coronin 1) protein. *Cell Microbiol*, 2003, **5**: 25-40.
- ZHOU, L., SUNG, J. J., LIN, S., *et al.* A five-year follow-up study on the pathological changes of gastric mucosa after *H. pylori* eradication. *Chin Med J*, 2003, **116**: 11-14.
- ZULLO, A., HASSAN, C., ANDRIANI, A., *et al.* Eradication therapy for *Helicobacter pylori* in patients with gastric MALT lymphoma: a pooled data analysis. *Am J Gastroenterol*, 2009, **104**: 1932-1937; quiz 1938.
- ZULLO, A., PERNA, F., HASSAN, C., *et al.* Primary antibiotic resistance in *Helicobacter pylori* strains isolated in northern and central Italy. *Aliment Pharmacol Ther*, 2007, **25**: 1429-1434.

## **Annexes**

### **Annexe I: Article n°5: CagA and VacA polymorphisms are associated with distinct pathological features in *Helicobacter pylori*-infected adults with peptic and non-peptic ulcer disease**

Panayotopoulou EG, Sgouras DN, Papadakos KS, Petraki K, **Breurec S**, Michopoulos S, Mantzaris G, Papatheodoridis G, Mentis A, Archimandritis A. J Clin Microbiol. 2010. 48: 2237-2239.

## CagA and VacA Polymorphisms Are Associated with Distinct Pathological Features in *Helicobacter pylori*-Infected Adults with Peptic Ulcer and Non-Peptic Ulcer Disease<sup>∇</sup>

Effrosini G. Panayotopoulou,<sup>1</sup> Dionyssios N. Sgouras,<sup>1\*</sup> Konstantinos S. Papadakos,<sup>1</sup> Kalliopi Petraki,<sup>1</sup> Sébastien Breurec,<sup>2</sup> Spyros Michopoulos,<sup>3</sup> Gerassimos Mantzaris,<sup>4</sup> George Papatheodoridis,<sup>5</sup> Andreas Mentis,<sup>1</sup> and Athanasios Archimandritis<sup>5</sup>

Laboratory of Medical Microbiology, Hellenic Pasteur Institute, Athens, Greece<sup>1</sup>; Laboratoire de Biologie Médicale, Institut Pasteur, Dakar, Senegal<sup>2</sup>; Gastroenterology Clinic, Alexandra Hospital, Athens, Greece<sup>3</sup>; Gastroenterology Clinic, Evangelismos Hospital, Athens, Greece<sup>4</sup>; and Second Department of Internal Medicine, Athens University School of Medicine, Athens, Greece<sup>5</sup>

Received 30 March 2010/Accepted 5 April 2010

**Polymorphic variability in *Helicobacter pylori* factors CagA and VacA contributes to bacterial virulence. The presence of one CagA EPIYA-C site is an independent risk factor for gastroduodenal ulceration (odds ratio [OR], 4.647; 95% confidence interval [CI], 2.037 to 10.602), while the presence of the *vacA* i1 allele is a risk factor for increased activity (OR, 5.310; 95% CI, 2.295 to 12.287) and severity of gastritis (OR, 3.862; 95% CI, 1.728 to 8.632).**

*Helicobacter pylori*, colonizing the gastric mucosa of 35 to 70% of people worldwide, is the etiologic factor for peptic ulcer development and increases the risk for gastric cancer. *H. pylori* pathogenesis is exerted via distinct virulent factors such as the secreted cytotoxin VacA (vacuolating cytotoxin A), the *cag* pathogenicity island (*cagPAI*) encoding the type IV secretion system (T4SS), and the cytotoxin-associated gene A (CagA) protein (6). We analyzed *H. pylori* clinical isolates from the antrum of 144 Greek adults (mean age  $\pm$  standard deviation [SD], 52.6  $\pm$  13.7 years; 78 male) diagnosed with peptic ulcer (gastric,  $n = 21$ ; duodenal,  $n = 44$ ) and non-peptic ulcer disease (nonulcer dyspepsia,  $n = 61$ ; esophagitis,  $n = 18$ ) on the basis of functional CagA EPIYA motifs as well as *vacA* alleles for signal, intermediate, and middle regions, as described previously (13, 15), and assessed putative associations with disease parameters and gastric inflammatory response.

Approximately 27% of the strains were found to be *cagA* negative with complete absence of the *cagPAI*. Among the 96 *cagA*-positive isolates, 15 (10.4%) lacked a functional T4SS as they induced minimal interleukin-8 (IL-8) levels (Fig. 1A), and no phosphorylated CagA was detected (Fig. 1B) following infection of gastric epithelial AGS cells (15). Infection with strains possessing a functional T4SS led to significantly higher IL-8 secretion, irrespective of the number of EPIYA-C sites, and to CagA phosphorylation (Fig. 1A and B). Hence, for univariate and multivariate logistic regression analysis, *cagA*-positive isolates with a nonfunctional T4SS were grouped together with *cagA*-negative cases, comprising the “None” category. In single *H. pylori* strain infections, the majority of isolates ( $n = 59$ , 41.0%) were of the ABC EPIYA type (15), with a second EPIYA-C repeat observed in 19 (13.2%) strains,

while ABCCC strains were also identified ( $n = 2$ , 1.4%). In 11 cases (7.7%), the presence of mixed infection by isogenic strains differing solely with regard to the number of EPIYA-C repeats was identified as shown before (13).

The dominant *vacA* polymorphisms for the signal, intermediate, and middle regions were s1, i1, and m2, respectively, as reported for Western-type *H. pylori* strains (5, 17). More specifically, 102 (70.8%) isolates were identified as *vacA* s1, with 57 (39.6%) carrying the *vacA* m1 allele simultaneously. No strain with *vacA* s2/m1 was recorded. Of the 91 *vacA* i1 strains, 54 (59.3%) were also typed as *vacA* s1/m1, whereas 38/53 (71.7%) *vacA* i2 strains were s2/m2 ( $P < 0.001$ ). Depending on the *vacA* genotype, strains were further classified into three categories (7, 14), namely, nonvacuolation (s2/i2/m2, s1/i2/m1, s1/i2/m2, and s2/i1/m2), low vacuolation (s1/i1/m2), and high vacuolation (s1/i1/m1). Vacuolating *vacA* s1/i1/m1 or s1/i1/m2 types were present in strains harboring functional CagA variants with more EPIYA-C phosphorylation repeats, with frequencies reaching approximately 90% in cases of multiple infections, whereas *cagPAI*-defective strains were almost exclusively related to a nonvacuolating *vacA* genotype ( $P < 0.001$ ).

*vacA* s1 and i1 polymorphisms were found to be associated with marked chronic inflammatory infiltration and activity of chronic gastritis in the antrum (Table 1). No association was observed with the density of *H. pylori* colonization or the presence of gastric atrophy and intestinal metaplasia (IM), even though in 41/57 (71.9%) of recorded IM the strain carried the *vacA* i1 allele ( $P = 0.111$ ). However, the risk for IM development increased 2-fold upon infection with *vacA* m1 strains (odds ratio [OR], 2.182; 95% confidence interval [CI], 1.098 to 4.338;  $P = 0.026$ ). Heavy *H. pylori* colonization (OR of 6.866 and 95% CI of 3.072 to 15.344 [ $P < 0.001$ ] and OR of 8.476 and 95% CI of 3.633 to 19.777 [ $P < 0.001$ ], respectively) and infection with *vacA* i1 strains (OR of 3.862 and 95% CI of 1.728 to 8.632 [ $P = 0.001$ ] and OR of 5.310 and 95% CI of 2.295 to

\* Corresponding author. Mailing address: Laboratory of Medical Microbiology, Hellenic Pasteur Institute, Athens, Greece. Phone: 302106478824. Fax: 302106478832. E-mail: sgouras@pasteur.gr.

<sup>∇</sup> Published ahead of print on 14 April 2010.

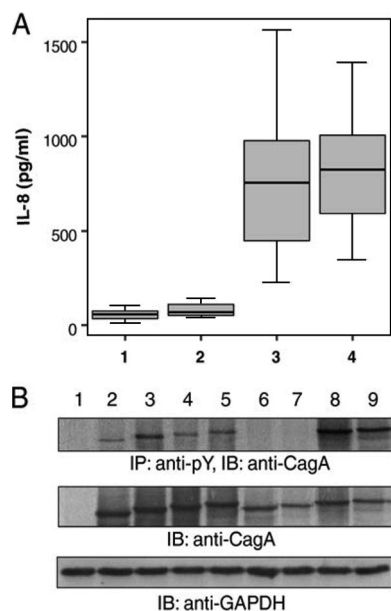


FIG. 1. (A) Levels of secreted IL-8 following infection of gastric epithelial AGS cells with *H. pylori* clinical strains (1, *cagA* negative; 2, *cagPAI* defective; 3, 1 EPIYA-C repeat; 4,  $\geq 2$  EPIYA-C repeats). No difference was observed between *cagA*-negative and *cagPAI*-defective strains ( $U = 32.500$  and  $P = 0.201$  by the Mann-Whitney U test). *cagA*- and *cagPAI*-positive strains induced higher levels of IL-8 than *cagPAI*-defective strains ( $U = 0.000$  and  $P < 0.0001$  by the Mann-Whitney U test), irrespective of the number of EPIYA-C sites ( $U = 224.500$  and  $P = 0.627$  by the Mann-Whitney U test). (B) Tyrosine phosphorylation and expression patterns of CagA protein following infection of AGS cells with representative *H. pylori* clinical strains. CagA tyrosine phosphorylation was detected by immunoblotting (IB) following immunoprecipitation (IP) with PY20 antiphosphotyrosine antibody. The expression of GAPDH (glyceraldehyde-3-phosphate dehydrogenase) was utilized as a total protein loading control. Lanes: 1, CagA-negative clinical isolate; 2 to 5, CagA-positive isolates with functional *cagPAI* harboring 2 (AB), 3 (ABC), 4 (ABCC), and 5 (ABCCC) motifs in CagA, respectively; 6 and 7, CagA-positive *H. pylori* strains carrying 3 (ABC) and 4 (ABCC) EPIYA motifs with defective *cagPAI*, respectively, as depicted by the absence of phosphorylated CagA; 8 and 9, CagA-positive strains with 5 (ABCCC) EPIYA motifs.

12,287 [ $P = 0.001$ ], respectively) were recognized as independent risk factors for the development of severe chronic inflammatory infiltration and marked activity of chronic gastritis in the antrum. This is the first report associating *vacA* intermediate region polymorphisms with increased activity of antral gastritis. *vacA* s1 strains of Western origin have previously been associated with more-severe gastric inflammation (5, 11, 16). The only reports relating specific *cagA* polymorphisms with histological lesions involve strains of East Asian origin carrying the ABD EPIYA sites, which present distinct biological properties compared to Western-type ABC EPIYA motifs (4).

The development of gastric ulcers (GU) or duodenal ulcers (DU) was associated with the occurrence of *H. pylori* strains harboring functional EPIYA-C repeats in CagA ( $P = 0.001$ ) as well as with the *vacA* s1 allele ( $P = 0.004$ ) and marked activity of chronic gastritis ( $P = 0.014$ ). Despite that, CagA EPIYA polymorphisms were found to be the only independent risk factor for ulcer disease (Table 2), with over 50% of strains with 1 (57.6%) or 2 or more (52.4%) EPIYA-C repeats found to be isolated from ulcer cases (8 GU/26 DU and 4 GU/7 DU, respectively) and the majority (8/11, 72.8%) of multiple infections with isogenic strains (4 GU and 4 DU) (Table 2). To date, infection with *cagA*-positive *H. pylori* has been well associated with gastroduodenal ulcers (2, 6, 12), whereas variability in the EPIYA phosphorylation sites and in particular CagA variants with an increased number of EPIYA-C repeats or of East Asian origin have been reported to augment the risk for gastric adenocarcinoma (1, 4, 8–10, 20). In our study, we observed that the presence of single or multiple infecting strains rather than the number of EPIYA-C sites in CagA *per se* is probably crucial in determining the type of gastric disease, since the majority of mixed infections with isogenic strains expressing CagA with various numbers of EPIYA-C repeats were isolated from peptic ulcer patients. Previous reports relate ulcer lesions with the presence of *vacA* s1 and i1 types (3, 5, 14), although in our sample, the association of VacA determinants with peptic

TABLE 1. Univariate logistic regression analysis showing association of *vacA* and *cagA* polymorphisms with severity and activity of chronic gastritis in the antrum of 144 Greek adults

Risk factor	Chronic inflammatory infiltration				Activity of chronic gastritis			
	No. (%) of isolates with mild/moderate severity	No. (%) of isolates with marked severity	OR (95% CI) <sup>a</sup>	<i>P</i>	No. (%) of isolates with mild/moderate activity	No. (%) of isolates with marked activity	OR (95% CI) <sup>a</sup>	<i>P</i>
<i>vacA</i> alleles								
s2	25 (17.4)	17 (11.8)	Reference		27 (18.8)	15 (10.4)	Reference	
s1	34 (23.6)	68 (47.2)	2.941 (1.402–6.171)	0.004	38 (26.4)	64 (44.4)	3.032 (1.435–6.405)	0.004
i2	30 (20.8)	23 (16.0)	Reference		34 (23.6)	19 (13.2)	Reference	
i1	29 (20.1)	62 (43.1)	2.789 (1.385–5.614)	0.004	31 (21.5)	60 (41.7)	3.463 (1.704–7.040)	0.001
<i>cagA</i> EPIYA status								
None	32 (22.2)	21 (14.6)	Reference		33 (22.9)	20 (13.9)	Reference	
1 EPIYA-C repeat	18 (12.5)	41 (28.5)	3.471 (1.589–7.58)	0.002	23 (16.0)	36 (25.0)	2.583 (1.204–5.539)	0.015
$\geq 2$ EPIYA-C repeat	6 (4.2)	15 (10.4)	3.810 (1.274–11.389)	0.017	6 (4.2)	15 (10.4)	4.125 (1.376–12.363)	0.011
Mixed infections	3 (2.1)	8 (5.6)	4.063 (0.966–17.091)	0.056	3 (2.1)	8 (5.6)	4.4 (1.044–18.542)	0.044
Vacuolation potential								
None	32 (22.2)	24 (16.7)	Reference		35 (24.3)	21 (14.6)	Reference	
Low	12 (8.3)	22 (15.3)	2.444 (1.014–5.895)	0.047	11 (7.6)	23 (16.0)	3.485 (1.418–8.566)	0.007
High	15 (10.4)	39 (27.1)	3.467 (1.563–7.690)	0.002	19 (13.2)	35 (24.3)	3.070 (1.411–6.681)	0.005

<sup>a</sup> Reference, used as the reference category for the calculation of risk in each case.



TABLE 2. Multivariate logistic regression model depicting parameters relating to the development of peptic ulcers

Risk factor	No. (%) of cases with non-peptic ulcers	No. (%) of cases with peptic ulcers	OR (95% CI) <sup>a</sup>	P
<i>cagA</i> EPIYA status				
Defective <i>cagPAI</i>	41 (28.5)	12 (8.3)	Reference	<0.001
1 EPIYA-C repeat	25 (17.4)	34 (23.6)	4.647 (2.037–10.602)	0.015
≥2 EPIYA-C repeat	10 (6.9)	11 (7.6)	3.758 (1.288–10.969)	0.003
Mixed infections	3 (2.1)	8 (5.6)	9.111 (2.085–39.810)	<0.001
<i>vacA</i> alleles				
s2	30 (20.8)	11 (7.6)	Reference	
s1	49 (34.0)	54 (37.5)		0.24
i2	34 (23.6)	19 (13.2)	Reference	
i1	45 (31.3)	46 (31.9)		0.76
Activity of chronic gastritis				
Mild	11 (7.6)	1 (0.7)	Reference	
Moderate	32 (22.2)	21 (14.6)		0.383
Marked	36 (25.0)	43 (29.9)		0.129

<sup>a</sup> Reference, used as the reference category for the calculation of risk in each case.

ulcer disease was not sustained through multivariate analysis, possibly reflecting geographical differences in the prevalence of the various genotypes (17–19).

Collectively, our data indicate a distinct yet coordinated activity of virulence factors associated with *H. pylori* pathogenesis, with CagA contributing to the development of particular disease phenotypes, such as peptic ulcer, and VacA differentially affecting the inflammatory process. Our findings emphasize the necessity to meticulously assess the functionality of virulence factors in *H. pylori* clinical strains so as to discern the true biological significance that lies beneath the plasticity of the *H. pylori* genome.

**Nucleotide sequence accession numbers.** Partial *cagA* nucleotide sequences were submitted to the GenBank/EMBL/DDBJ databases under accession numbers FN561978 to FN562025.

This study was supported by a Hellenic Pasteur Institute grant (HPI-922616) and in part by an Actions Concertées Inter Pasteuriennes (ACIP) program entitled “Etude de la fréquence de l’infection par *Helicobacter pylori* (*Hp*) des patients présentant une pathologie gastro-duodénale, de la sensibilité des souches aux antibiotiques et de la diversité de la région 3’ du gène *cagA*.”

#### REFERENCES

- Argent, R. H., M. Kidd, R. J. Owen, R. J. Thomas, M. C. Limb, and J. C. Atherton. 2004. Determinants and consequences of different levels of CagA phosphorylation for clinical isolates of *Helicobacter pylori*. *Gastroenterology* **127**:514–523.
- Atherton, J. C. 2006. The pathogenesis of *Helicobacter pylori*-induced gastro-duodenal diseases. *Annu. Rev. Pathol.* **1**:63–96.
- Atherton, J. C., P. Cao, R. M. Peek, Jr., M. K. Tummuru, M. J. Blaser, and T. L. Cover. 1995. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific *vacA* types with cytotoxin production and peptic ulceration. *J. Biol. Chem.* **270**:17771–17777.
- Azuma, T., S. Yamazaki, A. Yamakawa, M. Ohtani, A. Muramatsu, H. Suto, Y. Ito, M. Dojo, Y. Yamazaki, M. Kuriyama, Y. Keida, H. Higashi, and M. Hatakeyama. 2004. Association between diversity in the Src homology 2 domain-containing tyrosine phosphatase binding site of *Helicobacter pylori* CagA protein and gastric atrophy and cancer. *J. Infect. Dis.* **189**:820–827.
- Basso, D., C. F. Zambon, D. P. Letley, A. Stranges, A. Marchet, J. L. Rhead, S. Schiavon, G. Guariso, M. Ceroti, D. Nitti, M. Rugge, M. Plebani, and J. C. Atherton. 2008. Clinical relevance of *Helicobacter pylori* *cagA* and *vacA* gene polymorphisms. *Gastroenterology* **135**:91–99.
- Blaser, M. J., and J. C. Atherton. 2004. *Helicobacter pylori* persistence: biology and disease. *J. Clin. Invest.* **113**:321–333.
- Cover, T. L., and S. R. Blanke. 2005. *Helicobacter pylori* VacA, a paradigm for toxin multifunctionality. *Nat. Rev. Microbiol.* **3**:320–332.
- Higashi, H., R. Tsutsumi, A. Fujita, S. Yamazaki, M. Asaka, T. Azuma, and M. Hatakeyama. 2002. Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc. Natl. Acad. Sci. U. S. A.* **99**:14428–14433.
- Jones, K. R., Y. M. Joo, S. Jang, Y. J. Yoo, H. S. Lee, I. S. Chung, C. H. Olsen, J. M. Whitmore, D. S. Merrell, and J. H. Cha. 2009. Polymorphism in the CagA EPIYA motif impacts development of gastric cancer. *J. Clin. Microbiol.* **47**:959–968.
- Naito, M., T. Yamazaki, R. Tsutsumi, H. Higashi, K. Onoe, S. Yamazaki, T. Azuma, and M. Hatakeyama. 2006. Influence of EPIYA-repeat polymorphism on the phosphorylation-dependent biological activity of *Helicobacter pylori* CagA. *Gastroenterology* **130**:1181–1190.
- Nogueira, C., C. Figueiredo, F. Carneiro, A. T. Gomes, R. Barreira, P. Figueira, C. Salgado, L. Belo, A. Peixoto, J. C. Bravo, L. E. Bravo, J. L. Realpe, A. P. Plaisier, W. G. Quint, B. Ruiz, P. Correa, and L. J. van Doorn. 2001. *Helicobacter pylori* genotypes may determine gastric histopathology. *Am. J. Pathol.* **158**:647–654.
- Nomura, A. M., G. I. Perez-Perez, J. Lee, G. Stemmermann, and M. J. Blaser. 2002. Relation between *Helicobacter pylori* *cagA* status and risk of peptic ulcer disease. *Am. J. Epidemiol.* **155**:1054–1059.
- Panayotopoulou, E. G., D. N. Sgouras, K. Papadakos, A. Kalliaropoulos, G. Papatheodoridis, A. F. Mentis, and A. J. Archimandritis. 2007. Strategy to characterize the number and type of repeating EPIYA phosphorylation motifs in the carboxyl terminus of CagA protein in *Helicobacter pylori* clinical isolates. *J. Clin. Microbiol.* **45**:488–495.
- Rhead, J. L., D. P. Letley, M. Mohammadi, N. Hussein, M. A. Mohagheghi, M. Eshagh Hosseini, and J. C. Atherton. 2007. A new *Helicobacter pylori* vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer. *Gastroenterology* **133**:926–936.
- Sgouras, D. N., E. G. Panayotopoulou, K. Papadakos, B. Martinez-Gonzalez, A. Roumbani, J. Panayiotou, C. vanVliet-Constantinidou, A. F. Mentis, and E. Roma-Giannikou. 2009. CagA and VacA polymorphisms do not correlate with severity of histopathological lesions in *Helicobacter pylori*-infected Greek children. *J. Clin. Microbiol.* **47**:2426–2434.
- Soltermann, A., S. Koetzer, F. Eigenmann, and P. Komminoth. 2007. Correlation of *Helicobacter pylori* virulence genotypes *vacA* and *cagA* with histological parameters of gastritis and patient’s age. *Mod. Pathol.* **20**:878–883.
- Van Doorn, L. J., C. Figueiredo, F. Megraud, S. Pena, P. Midolo, D. M. Queiroz, F. Carneiro, B. Vanderborcht, M. D. Pegado, R. Sanna, W. De Boer, P. M. Schneberger, P. Correa, E. K. Ng, J. Atherton, M. J. Blaser, and W. G. Quint. 1999. Geographic distribution of *vacA* allelic types of *Helicobacter pylori*. *Gastroenterology* **116**:823–830.
- van Doorn, L.-J., C. Figueiredo, R. Sanna, M. J. Blaser, and W. G. Quint. 1999. Distinct variants of *Helicobacter pylori* *cagA* are associated with *vacA* subtypes. *J. Clin. Microbiol.* **37**:2306–2311.
- Yamaoka, Y., T. Kodama, O. Gutierrez, J. G. Kim, K. Kashima, and D. Y. Graham. 1999. Relationship between *Helicobacter pylori* *iceA*, *cagA*, and *vacA* status and clinical outcome: studies in four different countries. *J. Clin. Microbiol.* **37**:2274–2279.
- Yamaoka, Y., T. Kodama, K. Kashima, D. Y. Graham, and A. R. Sepulveda. 1998. Variants of the 3’ region of the *cagA* gene in *Helicobacter pylori* isolates from patients with different *H. pylori*-associated diseases. *J. Clin. Microbiol.* **36**:2258–2263.

## **Annexe II : Logiciel STRUCTURE (Falush *et al.* 2003)**

### **1. Objectifs**

Le programme STRUCTURE implémente une méthode de clustering pour l'inférence de la structure génétique de populations à partir de données génétiques sous la forme de marqueurs indépendants. L'objectif est d'assigner les individus à une population génétique et d'estimer leur degré d'introgession/hybridation.

### **2. Définitions utiles**

**Génotype** : ensemble ou partie donnée de la composition génétique d'un individu.

**Locus** : emplacement physique précis et invariable sur un chromosome.

**Prior** : en statistique bayésienne, il s'agit de la distribution de probabilité estimée a priori, c'est-à-dire avant d'avoir pris en compte les données.

### **3. Hypothèse de départ**

Afin d'inférer la structure génétique des populations, les deux hypothèses suivantes sont émises :

1- toutes les populations inférées sont considérées être à l'équilibre de Hardy-Weinberg, une théorie qui postule que la fréquence des allèles et des génotypes est constante d'une génération à l'autre.

Pour que l'équilibre existe, plusieurs hypothèses sont décrites :

- taille infinie,
- population panmictique (croisements aléatoires entre individus),
- absence de mutation sur les allèles considérés,
- absence de sélection d'individus,
- absence de migration,
- les générations ne se chevauchent pas,

2- un équilibre de liaison entre loci au sein des populations existe. La notion d'équilibre de liaison génétique entre deux loci reflète le fait que les allèles présents à l'un des locus s'associent au hasard aux allèles présents à l'autre locus. Ainsi, la fréquence d'observation au sein d'une population d'une combinaison d'allèles est égale au produit des fréquences d'observations individuelles des allèles.

#### **4. Les modélisations proposées par STRUCTURE et applicable à la génétique de population de *H. pylori***

1- Modèle « no linkage » : tous les gènes de chaque individu proviennent d'une unique population parmi les K populations possibles.

2- Modèle « linkage » : ce modèle permet de prendre en compte le déséquilibre de liaison lié à l'introgession c'est-à-dire les corrélations entre loci génétiquement liés du fait d'un évènement d'ad mixture récent entre les populations. Le modèle de base est le suivant : si l'on remonte de t générations dans le passé, il y a eu un évènement d'ad mixture qui a mélangé les K populations. Un chromosome d'un individu descendant de ce mélange sera alors constitué de différentes « portions » héritées des différentes populations.

#### **5. Principe**

STRUCTURE propose de déterminer la structure génétique des populations à l'aide d'une inférence dite "bayésienne".

##### **5.1. Inférence de la structure génétique**

Le principe est le suivant : soient X le vecteur aléatoire correspondant aux génotypes des individus échantillonnés, Z correspondant aux populations d'origine des individus, Q aux proportions des loci provenant des différentes populations et P aux fréquences alléliques dans chacune des populations. Sous les hypothèses précédemment citées, la loi de  $X|Z, P, Q$  est complètement connue mais il est impossible de calculer toutes les combinaisons (Z, P, Q) possibles.

Après avoir spécifié les priors Z, P et Q, un algorithme de simulation par MCMC a été choisi pour obtenir des couples de solutions approchées (Z1, P1, Q1), (Z2, P2, Q2)...

##### **5.2. Inférence du nombre K de populations**

Selon la règle de Bayes, la détermination du nombre K de populations se fait à posteriori, en sachant :  $P(K|X) \propto P(X|K)P(K)$ . Cependant, la distribution ainsi définie à posteriori peut varier de manière notable en fonction des hypothèses de modélisation et l'estimation de  $P(X|K)$  est une tâche délicate. Ainsi, une méthode d'estimation ad hoc a été préférée, supposant que  $X|K$  suit une loi exponentielle. On choisit alors pour K la valeur qui maximise la vraisemblance.

### **Annexe III : Valorisation scientifique de la thèse**

#### **1. Communications orales**

1. **Breurec S.** L'histoire des migrations humaines en Asie du sud-est et Océanie déduite de l'étude de la diversité génétique de *Helicobacter pylori*. Paris, France. Séminaire du Département Infection et Epidémiologie de l'Institut Pasteur, le 24 juin 2011.
2. **Breurec S.** Le peuplement de l'Asie du sud-est d'un point de vue microbien. Journée Nationale "Actualités 2011" du Groupe d'Etudes Français des *Helicobacters*. Bordeaux, France, le 28 janvier 2011.
3. **Breurec S.** Genetic diversity of *Helicobacter pylori* strains. Séminaire annuel Pathogenomics Era-Net *Helicobacter pylori*. Porto, Portugal, le 16 Septembre 2009.
4. **Breurec S.** Genetic diversity of *Helicobacter pylori* strains. Séminaire annuel Pathogenomics Era-Net *Helicobacter pylori*. Hanovre, Allemagne, le 28 Novembre 2008.
5. **Breurec S.** Genetic diversity of *Helicobacter pylori* strains from Africa, Asia, Europe and Oceania. Conference "Research on infectious diseases: a global challenge". Paris, France, le 27 juin 2008.
6. **Breurec S.** Etude de la fréquence de l'infection par *Helicobacter pylori* des patients présentant une pathologie gastro-duodénale, de la sensibilité des souches aux antibiotiques et de la diversité de la région 3' du gène *cagA*. X<sup>èmes</sup> journées de gastro-entérologie et d'hépatologie d'Afrique francophone. Dakar, Sénégal, le 1er décembre 2007.

#### **2. Communications affichées**

1. **Breurec S,** Michel R, Seck A, Brisse S, Fall C, Garin B, Huerre M, Labigne A, Mbengue M, Thiberge JM, Dia D, Raymond J. Clinical impact of *cagA* and *vacA* gene polymorphisms in *Helicobacter pylori* strains from Senegalese patients. X<sup>ème</sup> journées de l'école doctorale. Innovation thérapeutique : du Fondamental à l'Appliqué. Paris, France, 25-26 Mai 2010.
2. **Breurec S,** Thiberge JM, Garin B, Hem S, Seck A, Rougier Y, Monchy D, Labigne A. *Helicobacter pylori* strain specific coding sequences distribution in the genome of clinical strains, isolated among patients from various ethnic groups and geographical locations. 14<sup>th</sup> International Workshop on *Campylobacter*, *Helicobacter* and Related Organisms. Rotterdam, Hollande, 2-5 Septembre 2007.

## Index

- acidité gastrique.....51  
adhésines .....57, 67, 82  
ADN mitochondrial.....115, 119, 122, 127, 133  
Ancestral Europe 1 .....72, 118  
Ancestral Europe 2 .....72, 118  
aliments .....24  
amoxicilline .....29, 31  
average nucleotide identity.....61  
Australonésiens .....122, 125, 144  
BabA.....57  
*Campylobacter jejuni* .....67, 82  
CagA.....46, 47, 48, 49, 50, 51, 58, 68, 89, 109, 110, 113, 136, 138, 139, 142, 145, 173  
Cambodge.....104, 105, 106, 107, 109, 110, 126, 127, 128, 130, 132, 136, 139, 140, 141, 142, 144, 145, 148  
cancer gastrique.....16, 17, 19, 20, 21, 22, 32, 33, 38, 43, 47, 48, 51, 58, 59, 64, 106, 109, 110, 112, 113, 135, 136, 137, 138, 139, 140, 141, 142, 145, 148  
chromosome Y .....115, 116, 119, 122  
clarithromycine.....30, 31  
CLONAL FRAME .....100, 102, 123  
core .....76  
dynamique de diffusion.....134  
East Asian CagA .....46, 47, 110, 139  
écotypes.....61, 63  
EPIYA.....46, 47, 48, 109, 110, 112, 113, 114, 139, 140, 145  
facteur de nécrose tumorale alpha...33, 49, 50  
facteurs de risque.....28  
facteurs environnementaux .....32, 34, 145  
flagelles .....51, 52, 53, 55  
fluoroquinolones.....31, 32  
génomme .....64, 67, 68, 76, 78, 84, 88, 89  
GLOBOCAN.....135  
*Helicobacter acinonychis* 64, 68, 74, 77, 83  
*Helicobacter hepaticus*.....64, 67, 74  
*Helicobacter mustelae*.....64, 68  
hybridation ADN-ADN.....60, 61  
hpAfrica1.....94, 138, 140, 145  
hpAfrica2.....72, 78, 94  
hpAsia2.....69, 72, 101, 106, 107, 110, 118, 144  
hpEastAsia.....69, 72, 94, 105, 130  
hpEurope.....69, 71, 72, 74, 94, 105, 106, 107, 109, 110, 118, 127, 128, 130, 133, 134, 143, 144  
hpNEAfrica.....69, 94  
hpSahul .....94, 100, 101, 125, 144  
hspAmerind.....69, 71, 74, 94, 134  
hspEAsia.....69, 71, 94, 105, 106, 107, 109, 110, 118, 125, 128, 132, 134, 141, 144, 145  
hspMaori.....69, 71, 94, 101, 102, 110, 123, 125, 144  
hspSAfrica .....69, 71, 94  
hspWAfrica.....69, 71, 94, 140  
îlot de pathogénicité *cag*.....43, 50, 66, 67, 76, 78, 89, 110, 130, 138  
interleukine-1 bêta .....32, 33, 42  
interleukine-8.....33, 42, 47, 48, 49, 50, 51, 58  
incidence .....17, 20, 22  
Inde.....69, 71, 72, 94, 104, 105, 106, 118, 125, 127, 130, 132, 136  
infection mixte .....27, 112, 113  
infection récurrente.....22  
inflammation.....41, 43, 49, 51, 58  
interspersed sequences of the recipient... 84  
langues austro-asiatiques.....105, 107, 132, 139  
langues austronésiennes.....99, 101, 120, 122, 123, 125, 146  
lésions pré-néoplasiques .....19, 20, 21  
lipopolysaccharide .....56, 57  
lymphocytes T régulateurs.....51  
lymphome du MALT .....16, 19, 78  
macrodiversité.....64, 76, 77  
Malaisie....105, 106, 107, 127, 136, 140, 144  
Marshall B. ....14  
métronidazole .....30, 31, 32  
microdiversité .....64, 68, 77, 78, 81, 89  
migrations humaines.....64, 71, 72, 73, 99, 102, 105, 107, 115, 117, 118, 120, 133, 135, 143  
MultiLocus Sequence Analysis .....61  
mutations.....69, 80, 81, 82, 84, 85, 86, 88  
*Mycobacterium leprae* .....117  
*Mycobacterium tuberculosis*.....117  
Nod1 .....50  
Papillomavirus humain .....118

- peptidoglycane .....50  
prévalence.....21  
prix Nobel.....14  
réassortiment .....80, 83  
recombinaisons.....64, 80, 84, 86, 87, 88,  
90, 110, 113, 118, 134  
réparation par excision de base .....86  
réparation par excision de nucléotides ....86  
réparation des mésappariements .....86, 87  
réservoir.....22, 24  
SabA .....55, 57  
Sénégal.....69, 94, 95, 113, 136, 140, 141,  
142, 145, 148  
SHP-2 .....47, 48  
souche généraliste .....106, 134  
souche spécialiste .....106, 134  
STRUCTURE .....69, 72, 99, 109, 174  
système de restriction-modification ..76, 83  
Taiwan.....99, 101, 102, 106, 120, 121,  
122, 123, 125, 142, 144  
taux de mutation ..... 81, 87, 88  
taux de recombinaison ..... 88  
Thaïlande.....38, 69, 104, 105, 106, 107,  
127, 128, 132, 133, 136, 140, 144  
transmission ..... 23, 25, 26, 27, 28, 29  
ulcères ..... 14, 15, 16, 19  
uréase ..... 53, 55  
VacA.....34, 36, 38, 39, 40, 41, 42, 51,  
55, 67, 68, 94, 109, 110, 136, 143, 145,  
173  
variation de phase ..... 80, 81, 82, 86, 90  
Vietnam.....104, 105, 106, 107, 110, 126,  
128, 130, 132, 136, 139, 140, 141, 142,  
144, 145  
virus JC ..... 117  
virus T-lymphotropique humain ..... 117  
*Wolinella succinogenes* ..... 67, 83  
Warren R..... 14  
Western CagA..... 46, 47, 49, 110, 113, 139  
zone de plasticité..... 76

## **Résumé**

La bactérie *Helicobacter pylori* est associée à des pathologies gastro-duodénales sévères mais est également un marqueur de migrations humaines. Nous avons documenté les migrations humaines en Océanie et en Asie du sud-est et nous avons recherché un lien entre les facteurs de virulence et la pathologie gastro-duodénale au Sénégal et au Cambodge.

## **Mots clés**

*Helicobacter pylori*, population génétique, migrations humaines, Taiwan, Océanie, Inde, Asie du sud-est, cancer gastrique, *cagA*, EPIYA, *vacA*, introgression, Sénégal, Cambodge

## **Laboratoire de rattachement**

Institut Pasteur, Unité de Pathogenèse de *Helicobacter*, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

## **Laboratoire associé**

Institut Pasteur, Unité de Bactériologie Médicale et Environnementale, 36 avenue Pasteur, Dakar, Sénégal

PÔLE : MICROBIOLOGIE / THERAPEUTIQUES ANTIINFECTIEUSES

UNIVERSITÉ PARIS-SUD 11  
UFR «FACULTÉ DE PHARMACIE DE CHATENAY-MALABRY »  
5, rue Jean Baptiste Clément  
92296 CHÂTENAY-MALABRY Cedex