



Centre National de Référence des  
Campylobacters et Helicobacters

# La phylogénie appliquée à l'étude des bactéries - Le modèle Helicobacter

A. Ménard

université  
de **BORDEAUX**



**BaRITon**  
Bordeaux Research  
In Translational Oncology

# The *gyrA* gene: a pertinent target for *Helicobacteraceae* species taxonomy

A. Ménard, A. Buissonnière,  
V. Prouzet-Mauléon, F. Mégraud

Systematic and Applied Microbiology  
Volume 39, Issue 2, March 2016, Pages 77–87

# The 16S rRNA gene sequencing : A “gold standard” method

- study the evolution and phylogeny of bacteria
- as a reference in clinical microbiology
  - to identify bacteria in case of failure of the phenotypic tests
  - to identify and propose novel genus and species

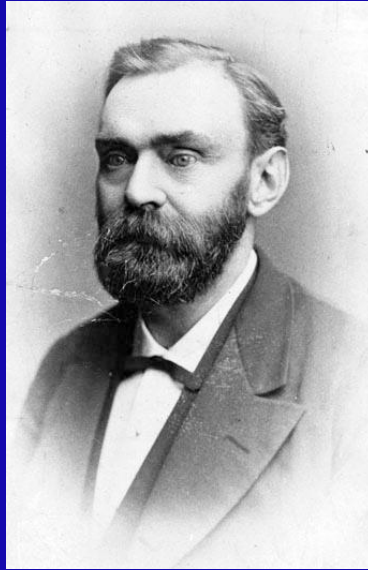
# The 16S rRNA gene sequencing in *Helicobacter* species

- discrepancies between 16S rRNA gene, phenotypic analyses and other genotypic methods
- 16S rRNA gene not highly informative in *Helicobacter* species  
(Vandamme 2000 ; Dewhirst 2005)
- ➡ could be explained by horizontal transfer of 16S rRNA gene fragments and creation of mosaic molecules with loss of phylogenetic information  
(Dewhirst 2005)

# The *Helicobacter* genus

- gastric species:
  - *H. pylori*
  - *Enterohepatic species: H. hepaticus*
- 25 species formally identified (Bergey's Manual 2005)
- numerous new proposed species
  - *H. burdigaliensis*
  - *H. muricola*
  - *H. callitrichis*
  - *H. macacae*
  - *H. westmeadii* .....

# Various mammals



humans



non-human primates

# Various mammals

Cats



Dogs



Ferrets



Musk shrews



**Non-rodent mammals**

# Rodents



Rats



Mice



Gerbils



Hamsters

# Rodents

Woodchucks



Squirrels



Meadow voles



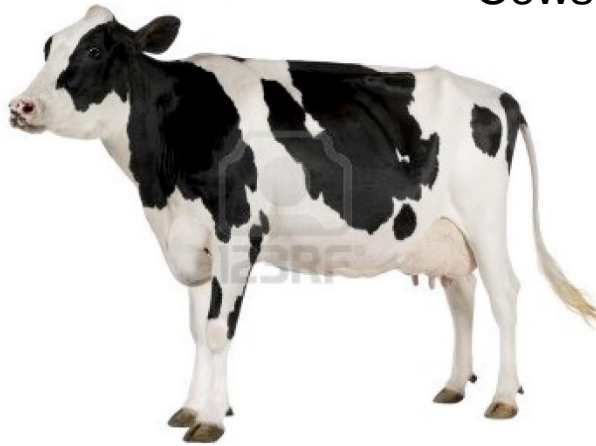
Chipmunks

# Livestock

Swine



Cows



Horses



Sheep



Goats



# Other wild animals



Deers



Raccoons



Cottontail rabbits



Stoats



Cheetahs



Opossums

# Marine mammals



Dolphins



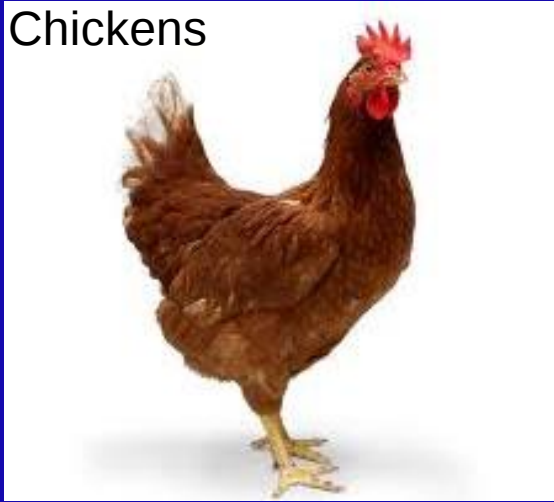
Whales



Seals

# Isolated in more than 26 bird species

Chickens



House sparrows



Terns



Gulls



Sandpipers



# Turtles



Eastern box turtles



Pancake tortoises

# Variable tropisms of the genus *Helicobacter*

- various hosts:
  - mammals
  - birds
  - reptiles...
- alternative sites within the digestive tract:
  - saliva
  - stomach
  - cecum
  - colon
  - liver ....

# Infection by *Helicobacter* species

- Various hosts infected by *Helicobacter* species
- Variable tropisms of *Helicobacter*
  - ➡ *Helicobacter* phylogeny hampered by
    - Gene adaptation
    - Gene exchange between *Helicobacter* species
  - ➡ evaluate new targets for *Helicobacter* species phylogeny

# Aims

- ➡ To investigate the *gyrA* gene as an alternative target for *Helicobacter* genus taxonomy
- ➡ To design a *gyrA* gene-based PCR to amplify different *Helicobacter* species

# Genes used for *Helicobacter* species phylogeny

- *cpn60* (*hsp60*) (Mikkonen 2004, Hill 2006)
- *rpoB-rpoC* (Zakharova 1999)

➡ no distinct separated clusters between gastric and enterohepatic *Helicobacter* species

# Genes used for *Helicobacter* species phylogeny

- *cpn60* (*hsp60*) (Mikkonen 2004, Hill 2006)
- *rpoB-rpoC* (Zakharova 1999)

➡ no distinct separated clusters between gastric and enterohepatic *Helicobacter* species

- 23S rRNA (Dewhirst 2005)
- *gyrB* (Hannula 2007)

# DNA gyrase

DNA gyrase : a type II topoisomerase



**2A 2B** → introduce negative supercoils into DNA

# DNA gyrase

## DNA gyrase : a type II topoisomerase



**2A 2B → introduce negative supercoils into DNA**

- ➡ essential in all bacteria
- ➡ attractive target for phylogeny

# DNA gyrase and bacterial phylogeny

- ***gyrB*** gene extensively used for bacterial phylogeny

  - ⇒ pertinent target for *Helicobacter* phylogeny

  - (1100 bp used – 41% of *gyrB*, Hannula 2007)

- ***gyrA*** gene reported to be an important tool for bacterial phylogeny

  - ⇒ pertinent target for *Campylobacter* and *Arcobacter* phylogeny

  - (Ménard 2005 ; Abdelbaqi 2007)

- ⇒ **To evaluate the *gyrA* gene for *Helicobacter* phylogeny**

# Materials & Methods

- 36 type *Helicobacter* type strains
- other proposed *Helicobacter* species

PCRs : - 16S rRNA gene (1,500 bp)  
- *hsp60* (*cpn60*) (600 bp)  
- *gyrA* (2,483 to 2,487 bp)  
- *gyrB* (1,500 bp)

# Methods - Bioinformatic analysis

## Bioinformatic analyses of *gyrA* and GyrA deduced protein

- Sequence alignments
  - *clustalW* (codons)
- Phylogenetic trees
  - *MEGA 5.2.2*
  - *SeaView 4.4.2*

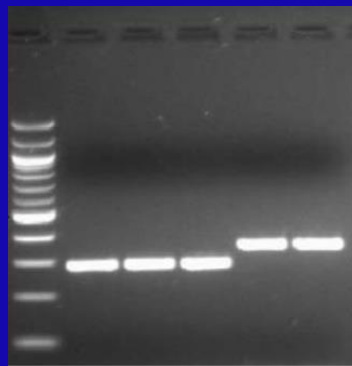
# Methods - Sequence analysis & PCR

## Sequence alignment : MultAlin

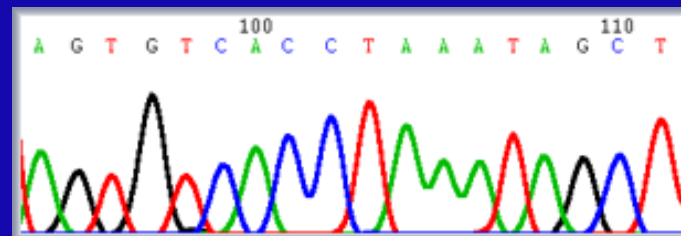
```
501                                     600
H-hepaticu AATCCCAAAATTGCTTGTAATGGTTCAAGCGGTATTGCGGTAGGTATGGCGACTTCTATTCCTCCGCATCGCATTGATGAGATTATAGATGCGCTTATT
H-cinaedi AATCCCAAAATTGCTTGTAATGGCTCAAGTGGTATTGCTGTGGGTATGGCGACTTCTATCCCGCCACACCGCATTGATGAGATTATTGACGCATTAT
H-bilis AATCCCAAAATTGCTTGTAATGGTTCTAATGGTATCGCCGTAGGTATGGCGACTTCTATCCACACATAGACTTGATGAAGTAGTCGATGCCCTAATT
H.wingham CTTACCAAACTTGCTGATTAAATGGTTCTAATGGAATTGCTGTTGGTATGGCAACAAATATCCACACATAGAACTGATGAAATCATTGACGCGCTTGTG
H-canadens TTTGCCAAATTGCTTATTAAATGGATCAATGGGATTGCTGTGGGTATGGCTACTAATATTCCTCCACATAGAATCGATGAGATTATTGATGCTTTGGTG
H-acinonyc TCTGCCTAACCTTTTAGTCAATGGCGCTAATGGGATTGCTGTAGGGATGGCCACTTCTATACCCCCCATAGGATTGATGAAATCATAGACGCTTAGTG
H-pylori TCTGCCTAACCTTTTAGTCAATGGGCTAATGGGATCGCCGTAGGATGGCGACTTCTATCCCCCTCATAGGATTGATGAAATCATAGACGCTTAGCG
H-cetorum TCTGCCTAATCTCTTAGTCAATGGGCTAATGGTATTGCAAGTAGGTATGGCTACTTCTATACCGCCGATAGAATTGATGAAATCATAGACGCTTAGTG
H-bizzozzer CTTGCCTAATCTCTCATCAATGGCTCAAGCGGATTGCTGTGGGATGGCGACCTCATTCCCCCCACCAGGATCATCGATGCGTTGGTG+
H-felis TCTGCCTAATCTCTCATCAATGGCTCAAGCGGATTGCTGTGGGATGGCGACCTCATTCACCCCATCGCCCCGATGAAATCATCGATGCGTTGAT
H-pullorum TTTGCCAAATTGCTTATTAAATGGCTCTAATGGGATTGCTGTGGGTATGGCGACAAATATCCACCCACAGAGTAGATGAAATATTGATGCTTTGGTG
Consensus ttTgCCAAAtTgcTtaT..ATGG.tCtAatGGgATtGCTGTgGGtATGGCGAcTtctATtCC.CC.CAtaGaattgaTGaaT.at.GATGC.tT.gTg

601                                     700
H-hepaticu ATGGTGATTGATAAAAAATACGCAAGATGTGGAAGACTTGCTTGATATTGTGTCAGGGACCGGATTTTCCACAGGTGGAATTATCTATGGCAAAGCAGGAA
H-cinaedi TTTGTGATTGATAACAACACGCAAAATGTTTGATGATGTGCTTGAACCTTGTGCAAGGACCGGATTTTCCACAGGTGGGATTATCTATGGCAAGATGGGAA
H-bilis TATAGAATCAATAATCCACAAGTGAAATGGAAGATATATGCGCTTAAATGAGGGACCAAGACTTCCACAGGTGGGATTATACATGGCAGACAGGGCA
H.wingham TATTTAATTGACAATCCAAAAGCAAGTTTAGATGAGATTTTGCAATTTGTGAAGGACCAAGATTTCCACAGGTGGGATTATTTATGGGAAAAGTGGA
H-canadens TATTTAATTGATACCCAAAAGCAGAAATTAATGAATTTTGCAATTTATGAAGGACCAAGATTTCCACAGGTGGAATCATCTATGGTAAAAATGGCA
H-acinonyc CATGTTTtagAAAACCCtAACGCTGAATTGAATGAGGTTTGGAAATTTATCAAGGGCCtGACtTTCCACCGGTGGGATCATCTATGGCAAGGTGGGAA
H-pylori CATGTCTTAGGAAACCCtAACGCTGAATTAGATAAAATTTGGAATTTGTCAAGGACCTGACtTTCTACTGGTGGGATCATCTATGGCAAGGCGGGTA
H-cetorum CATGTGCTAGAAAACCCtAACGCTGAATTGAATGAATCTTAGAATTTGTGAAGGGCCAGATTTCCCACTGGAGGAATCATCTATGGTAAGGTGGGTA
H-bizzozzer TTGTGCTAGATAACCCCAAAATGATCTGCTCTCTATTTTGAAGTGAATCAAGGTCCAAGATTTCCACAGGGGGGATTAATTATGGCAAGGCGGGCA
H-felis TATGTGCTAGAAAACCCCAAGCTGAATTGATCGACCTTTGAGAGATCGTGCAGGGGCCAGATTTCCACAGGGGGCGTGGTCTATGGCAAGGCAGGGA
H-pullorum TATTTGATTGATAATCCCAAGCAGAAATTAACGAATTTGCAATTTGTGAAGGTCCAAGATTTCCACAGGTGGTATTAATCTATGGTAAAAATGGAA
Consensus tatgtgaT.gAtAAccc.aagctgAatT..a.ga.aTttTggaatTtgTtgAaGGaCaGAtTTCCaACAGtGGaTtaTctATGGcAaa...GG.A
```

Primer design & PCR



→ Sequencing



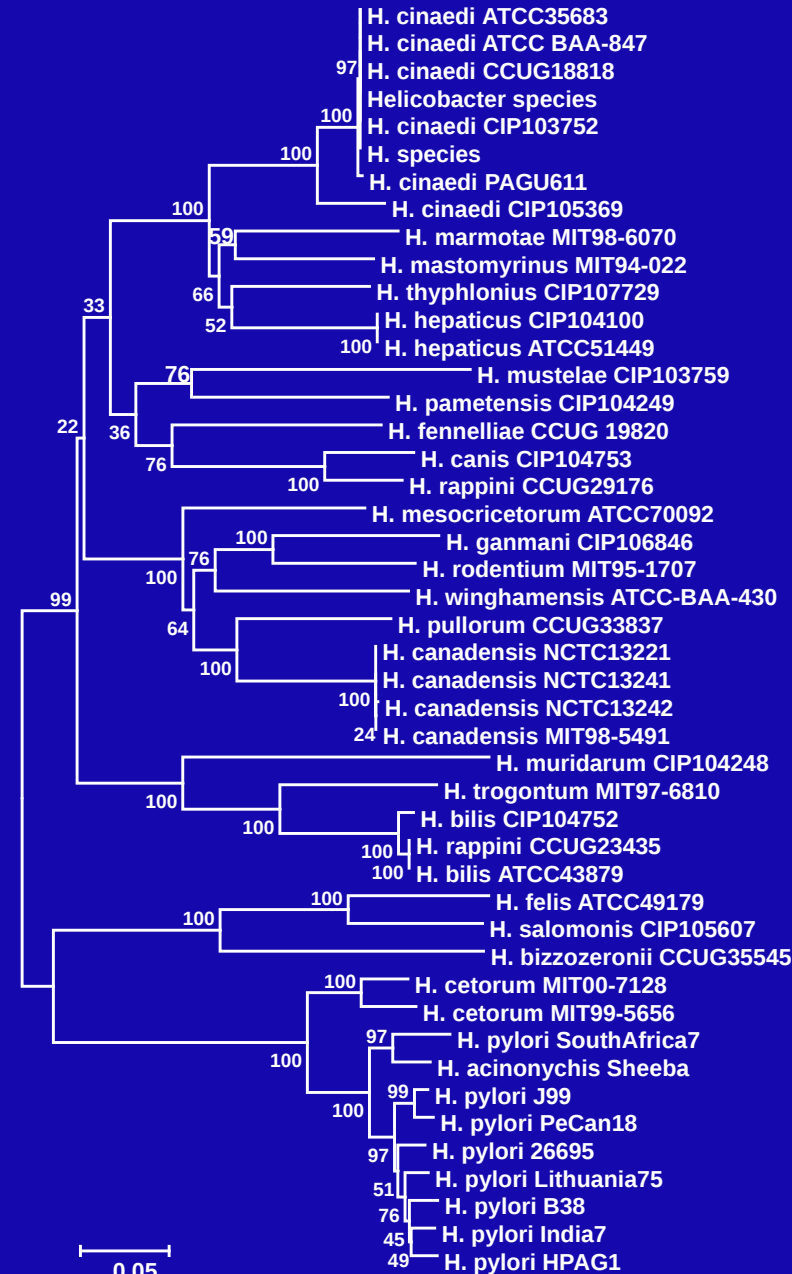
# Results

- All PCR generated amplicons : 81 *gyrA* sequences
- Others PCR:
  - 16S rRNA gene
  - *hsp60*
  - *gyrB*
- 85-100% of the *gyrA* gene: 2,100 to 2,103 pb  
    ➡ 827 aa GyrA
- Alignment: *gyrA* gene is highly divergent among *Helicobacter* species

# Results – Phylogenetic tree of *gyrA*

46 *gyrA* sequences

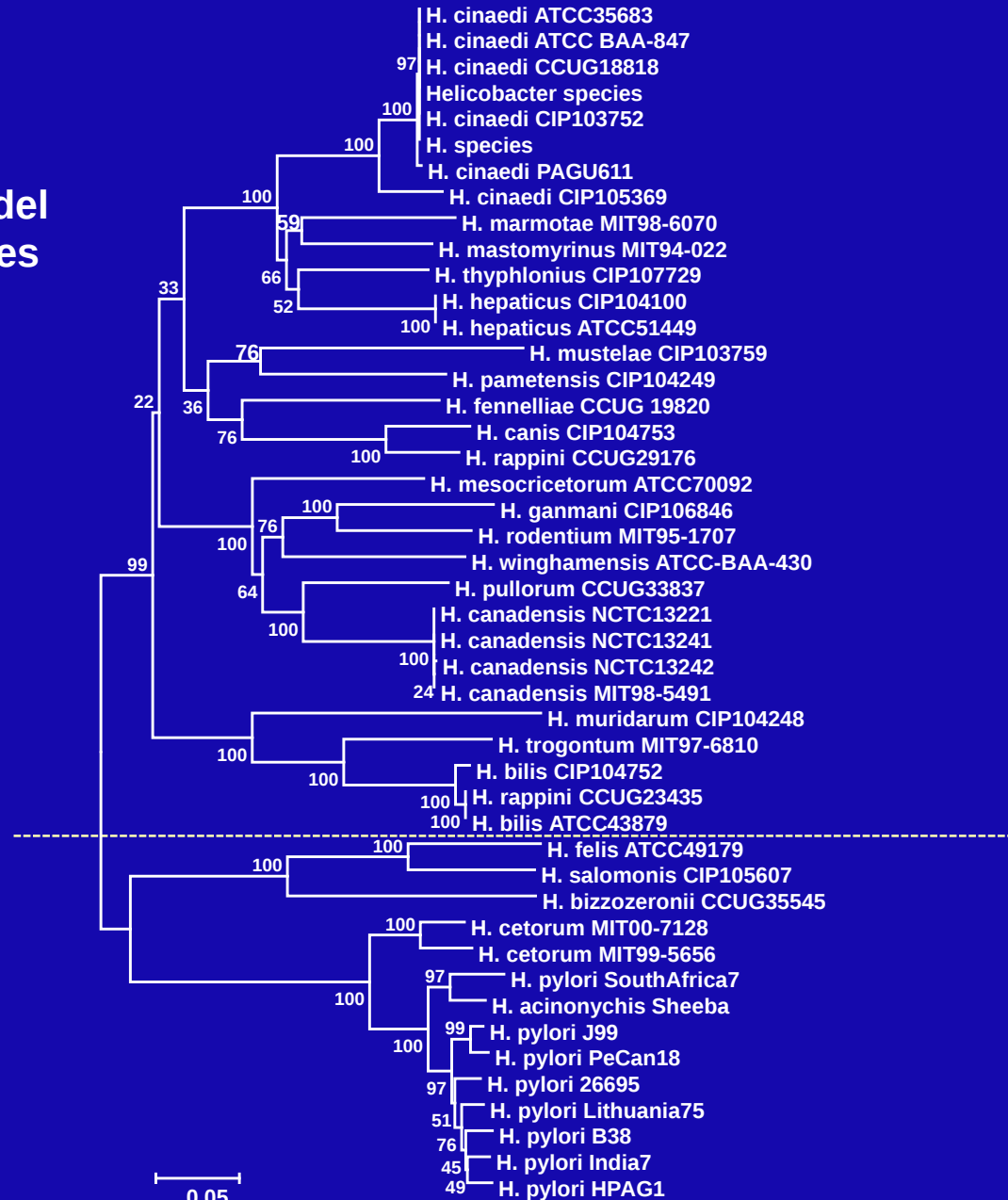
Neighbour-Joining  
Kimura 2-parameter model  
Bootstrap 1,000 replicates



# Results – Phylogenetic tree of *gyrA*

46 *gyrA* sequences

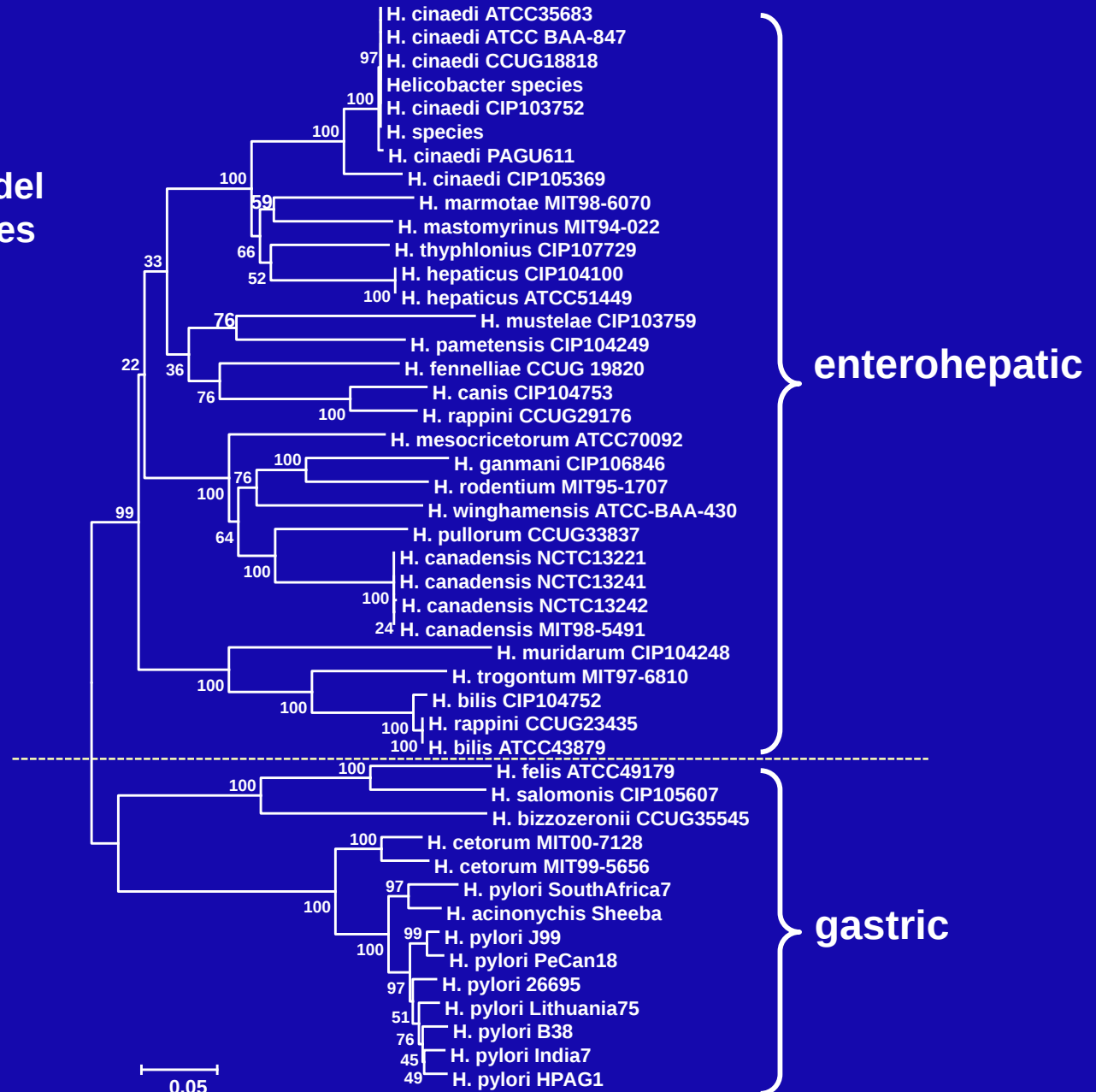
Neighbour-Joining  
Kimura 2-parameter model  
Bootstrap 1,000 replicates



# Results – Phylogenetic tree of *gyrA*

46 *gyrA* sequences

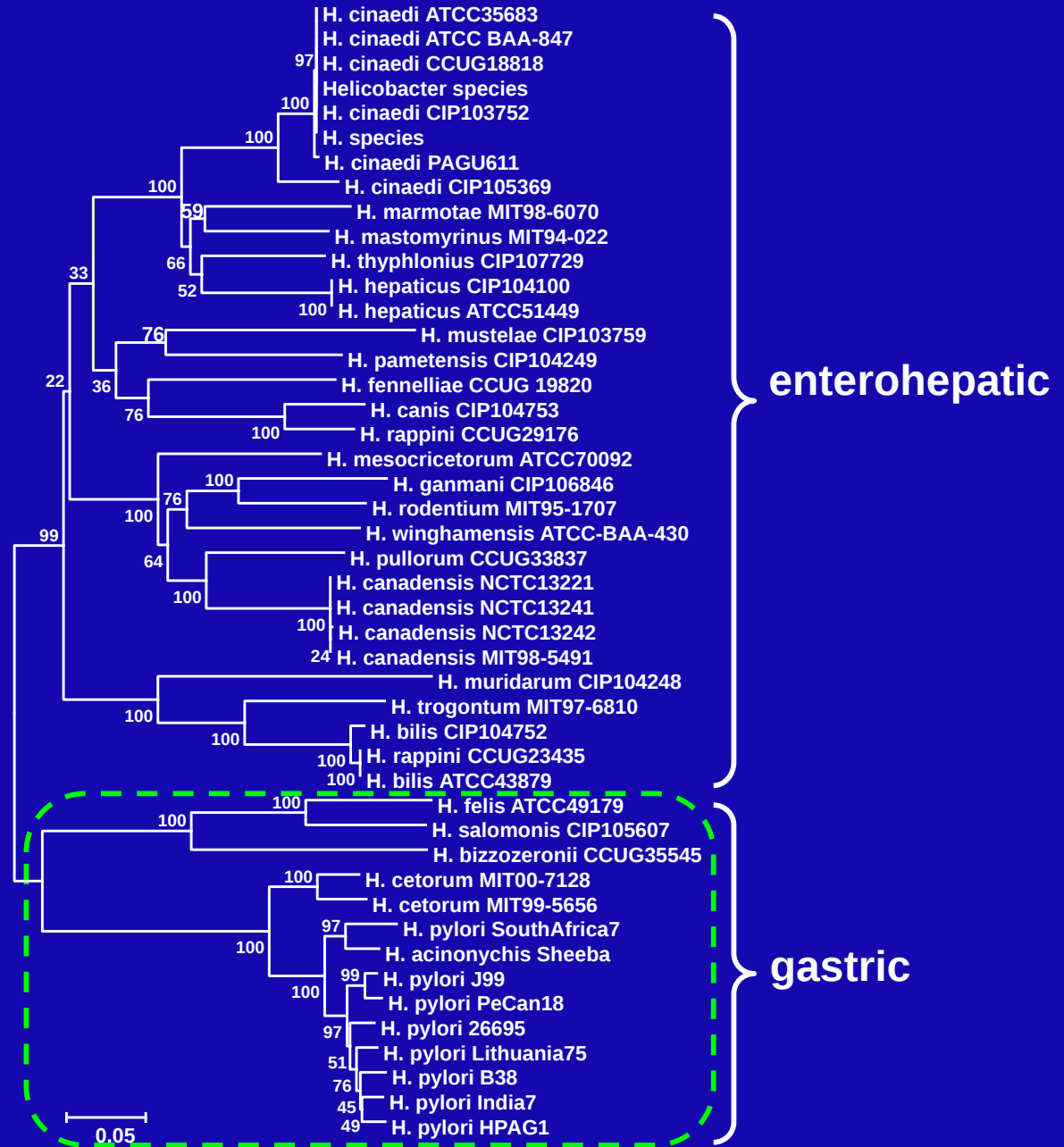
Neighbour-Joining  
Kimura 2-parameter model  
Bootstrap 1,000 replicates



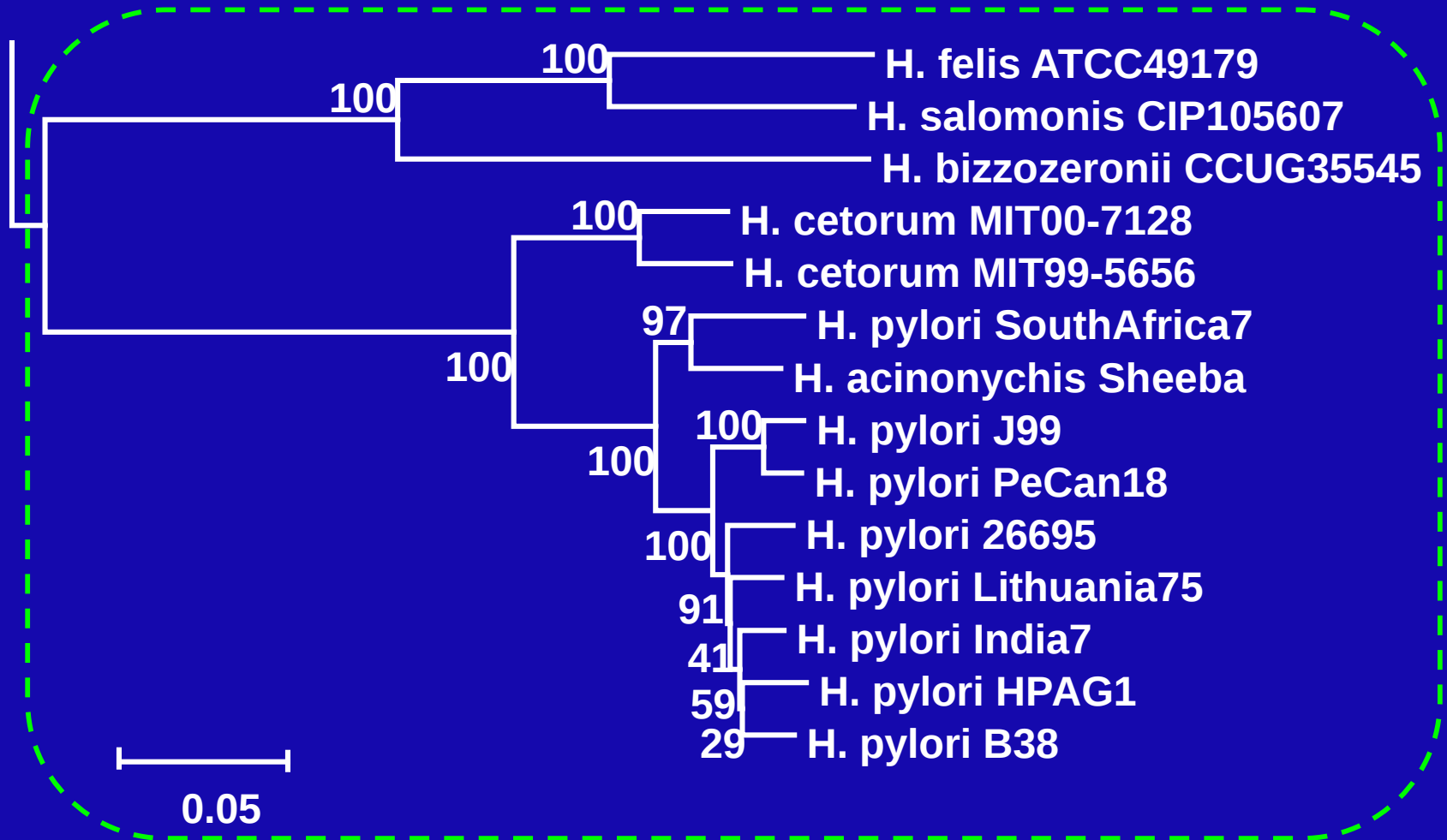
# Results – Phylogenetic tree of *gyrA*

46 *gyrA* sequences

Neighbour-Joining  
Kimura 2-parameter model  
Bootstrap 1,000 replicates

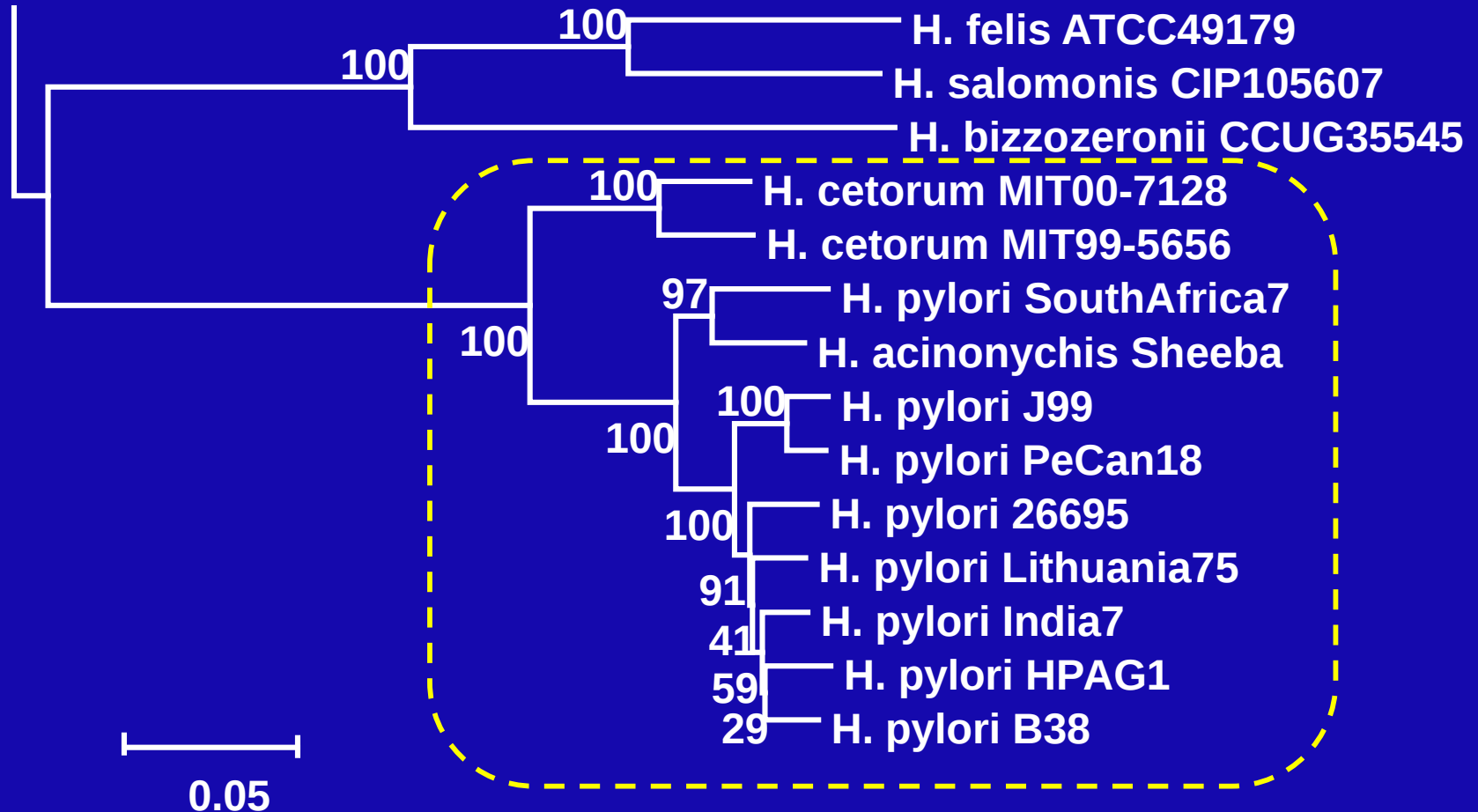


# Results - Phylogenetic tree of *gyrA* Gastric *Helicobacter* species



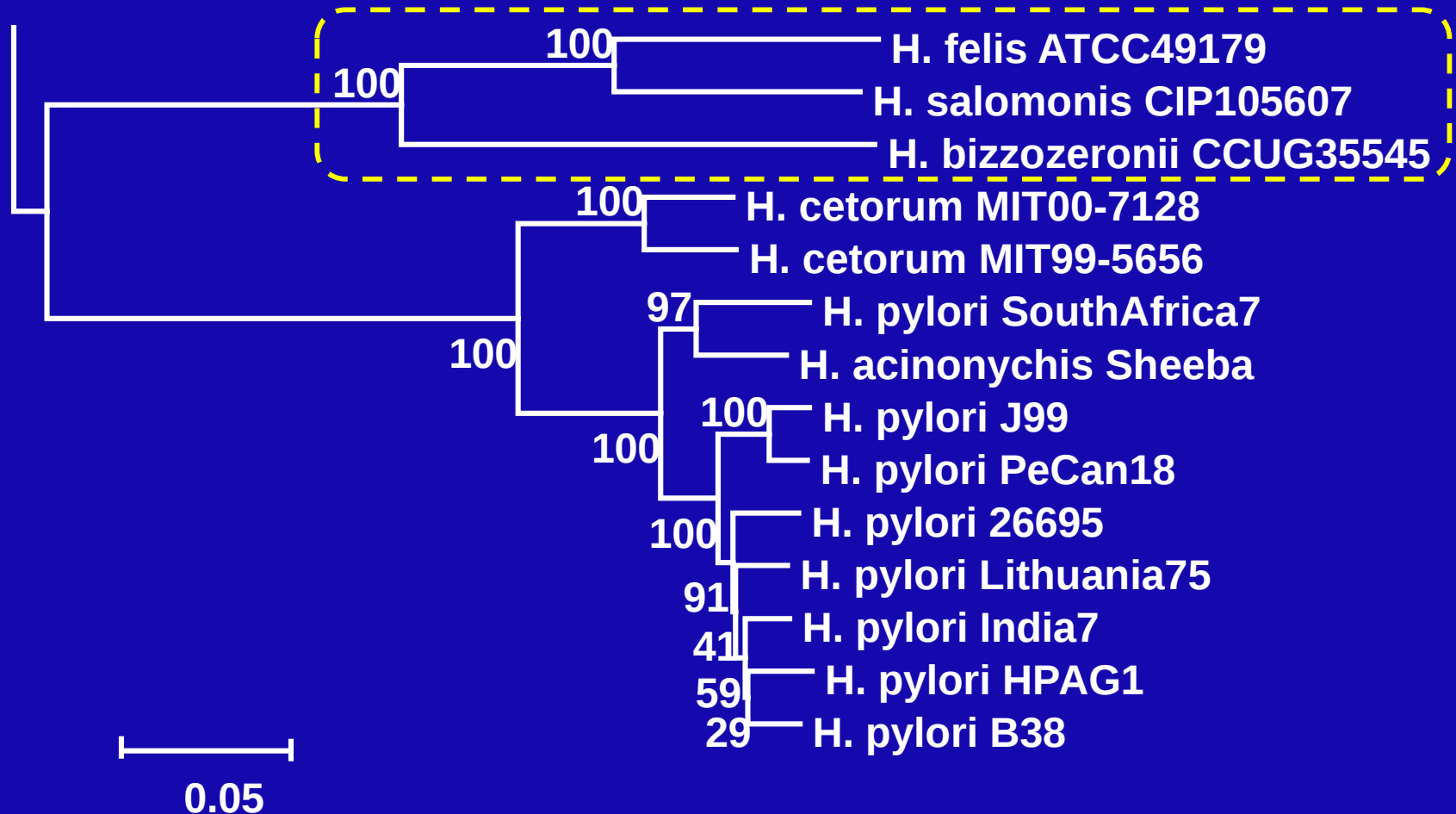
# Results - Phylogenetic tree of *gyrA*

## Gastric *Helicobacter* species



# Results - Phylogenetic tree of *gyrA*

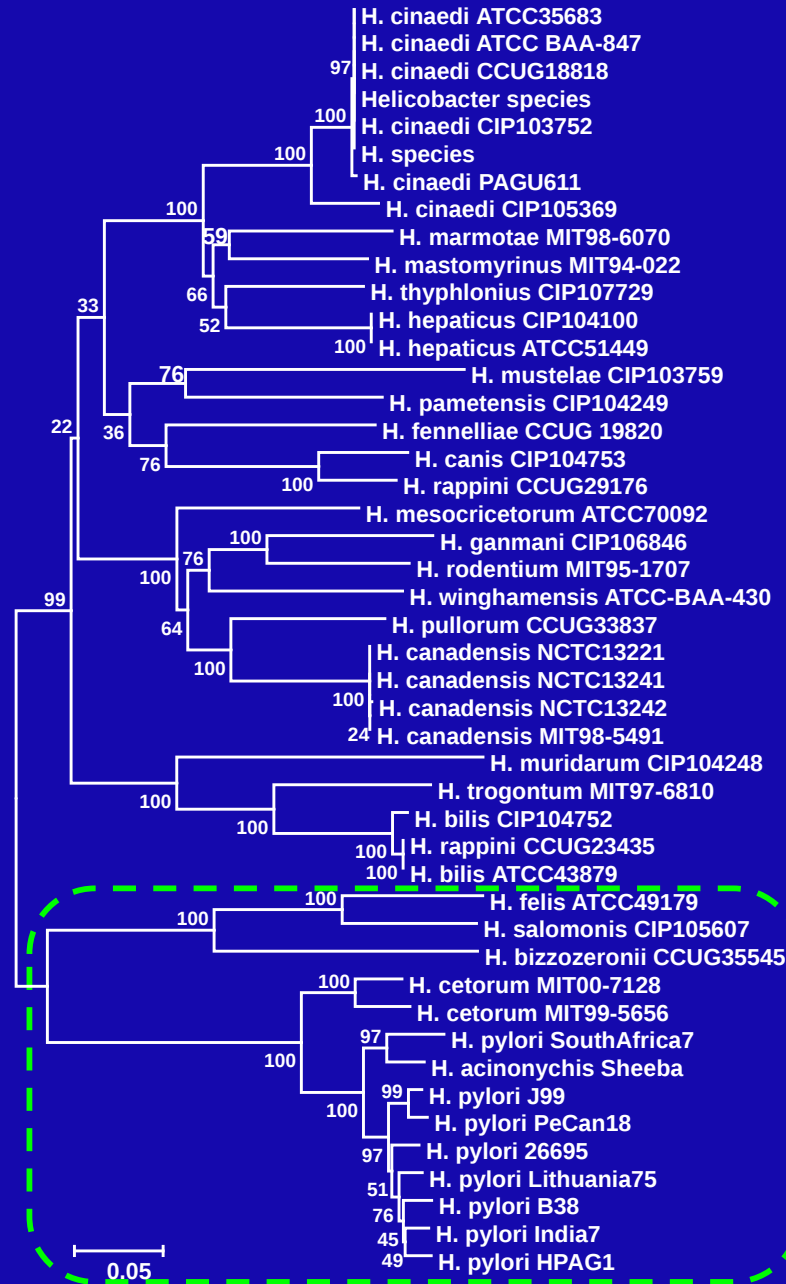
## Gastric *Helicobacter* species



# Results - Phylogenetic tree of *gyrA*

46 *gyrA* sequences

Neighbour-Joining  
Kimura 2-parameter model  
Bootstrap 1,000 replicates

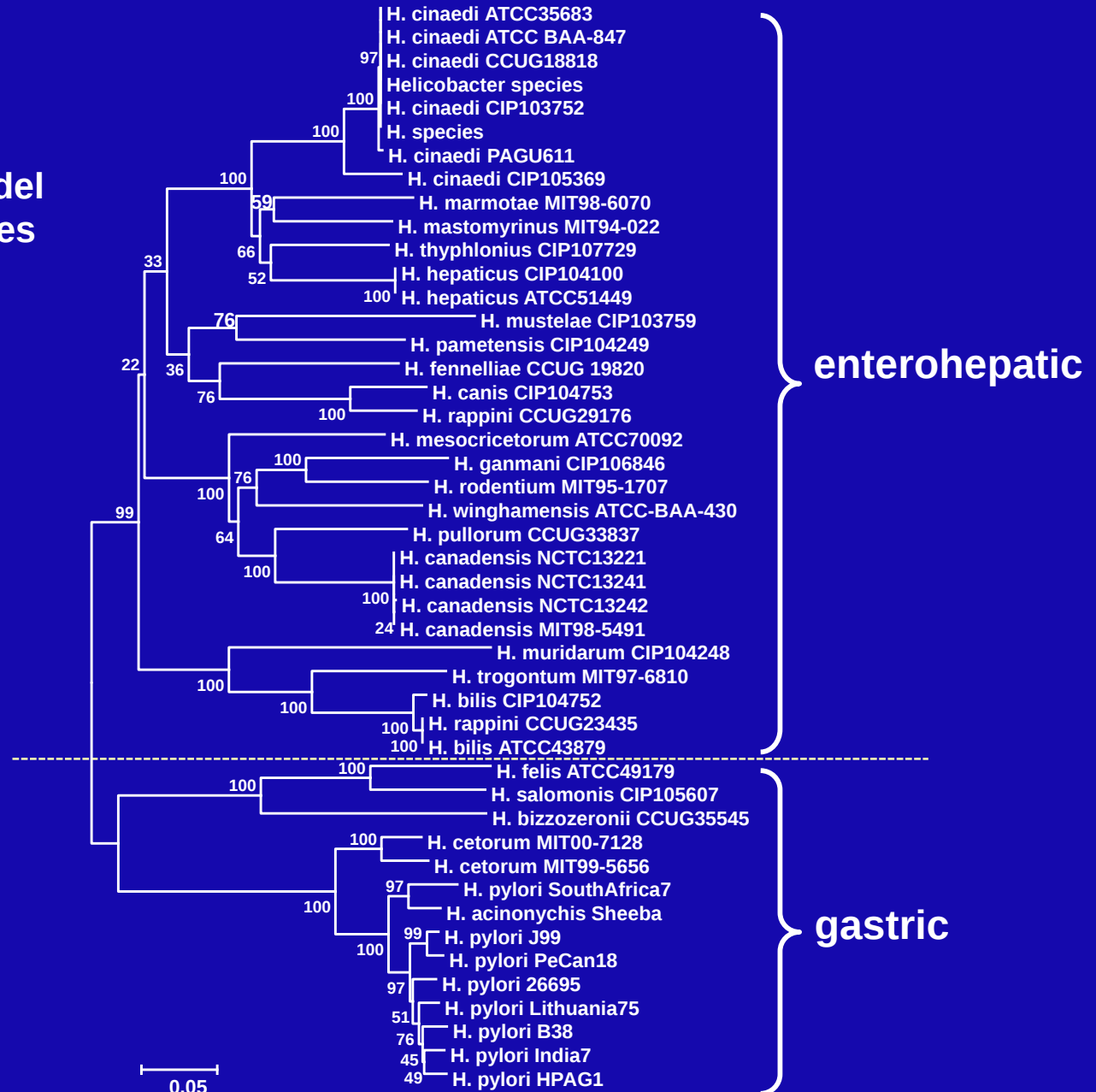


Similar for 23S rDNA  
and *gyrB* phylogenies

# Results – Phylogenetic tree of *gyrA*

46 *gyrA* sequences

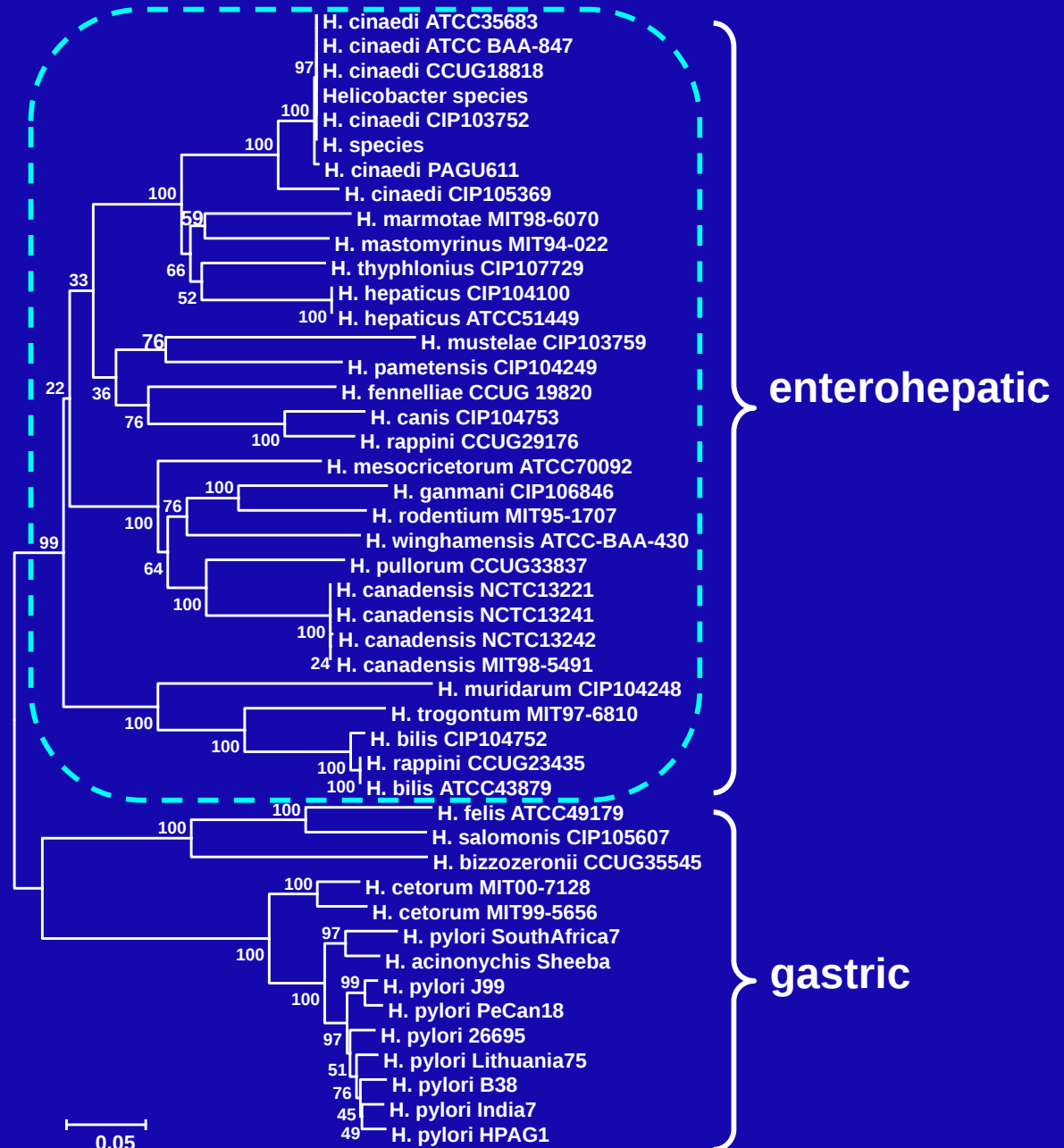
Neighbour-Joining  
Kimura 2-parameter model  
Bootstrap 1,000 replicates



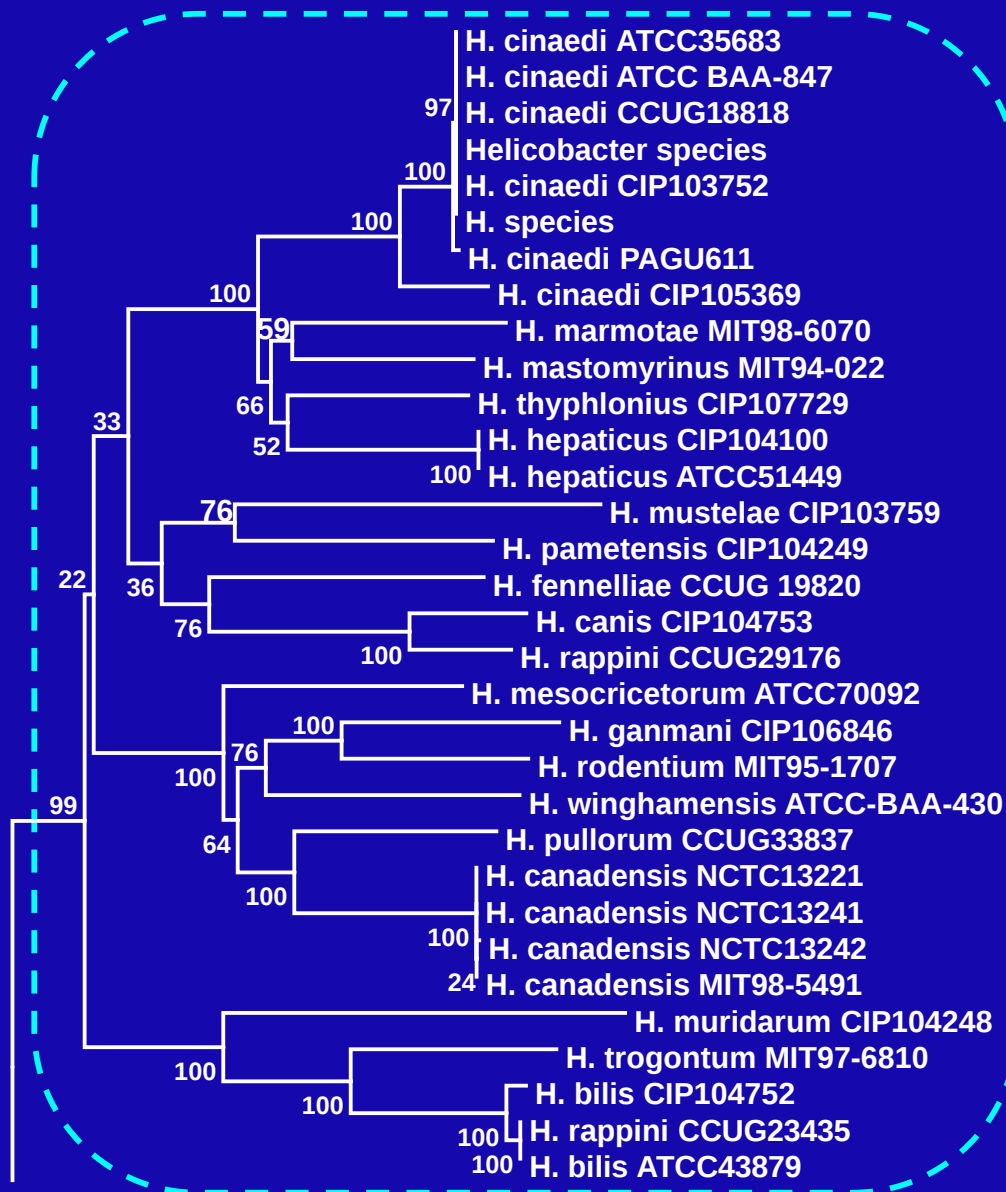
# Results – Phylogenetic tree of *gyrA*

46 *gyrA* sequences

Neighbour-Joining  
Kimura 2-parameter model  
Bootstrap 1,000 replicates

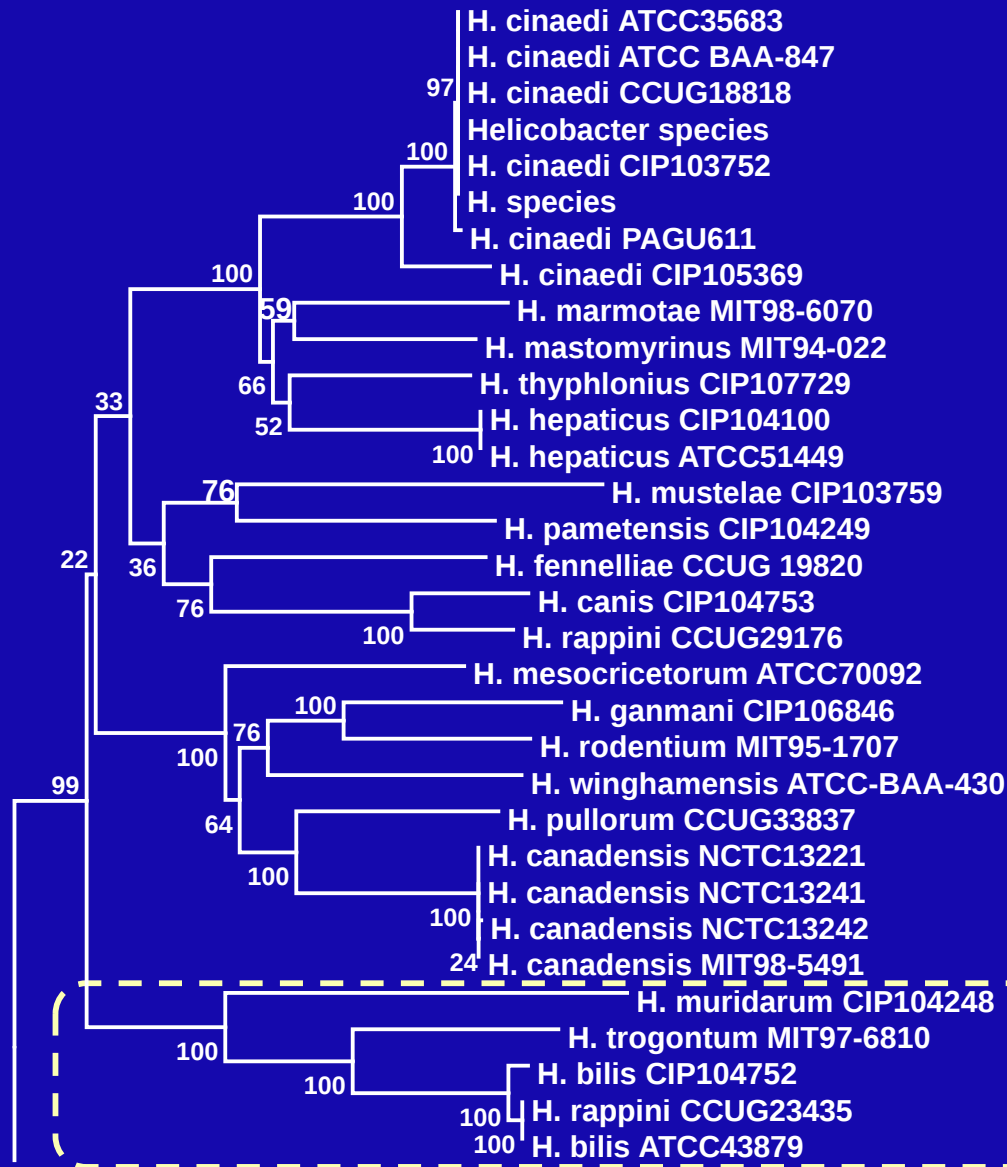


# Results - Phylogenetic tree of *gyrA* Enterohepatic *Helicobacter* species



# Results - Phylogenetic tree of *gyrA*

## Enterohepatic *Helicobacter* species

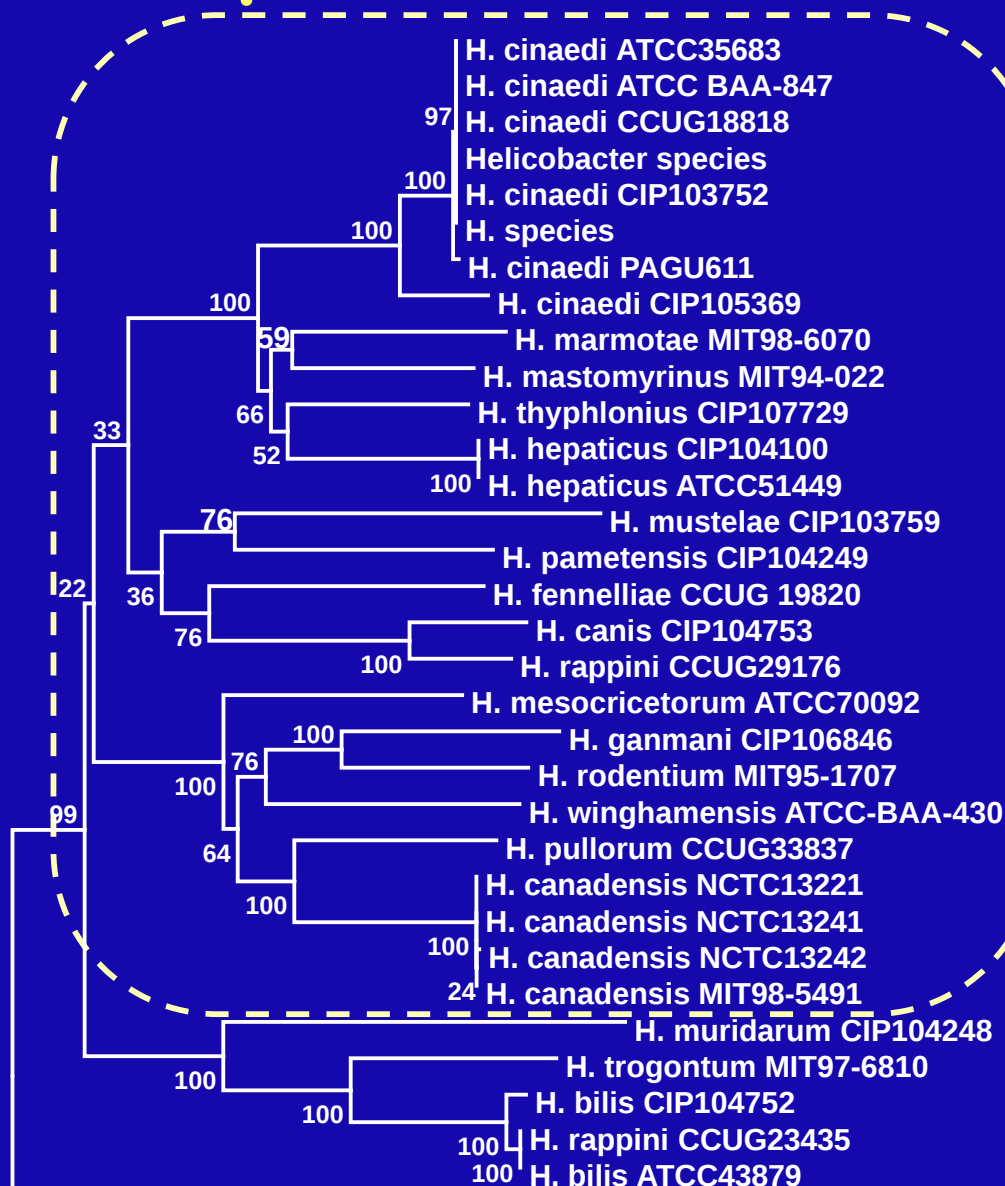


### EHH branch no.1

- similar to 23S *rDNA* phylogeny
- identical to *gyrB* phylogeny

# Results - Phylogenetic tree of *gyrA*

## Enterohepatic *Helicobacter* species

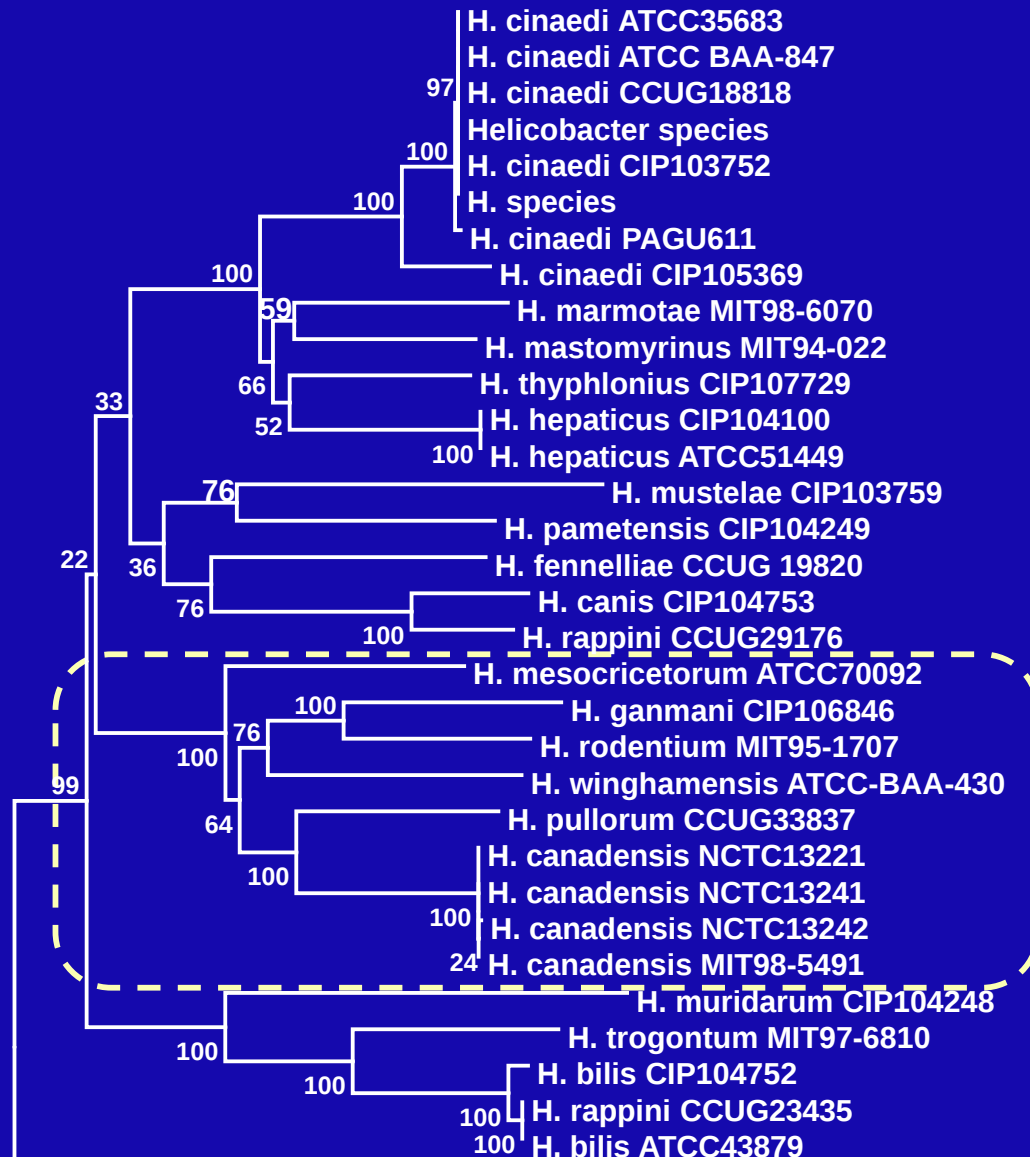


EHH branch no.2

- Difficult to compare with *gyrB*

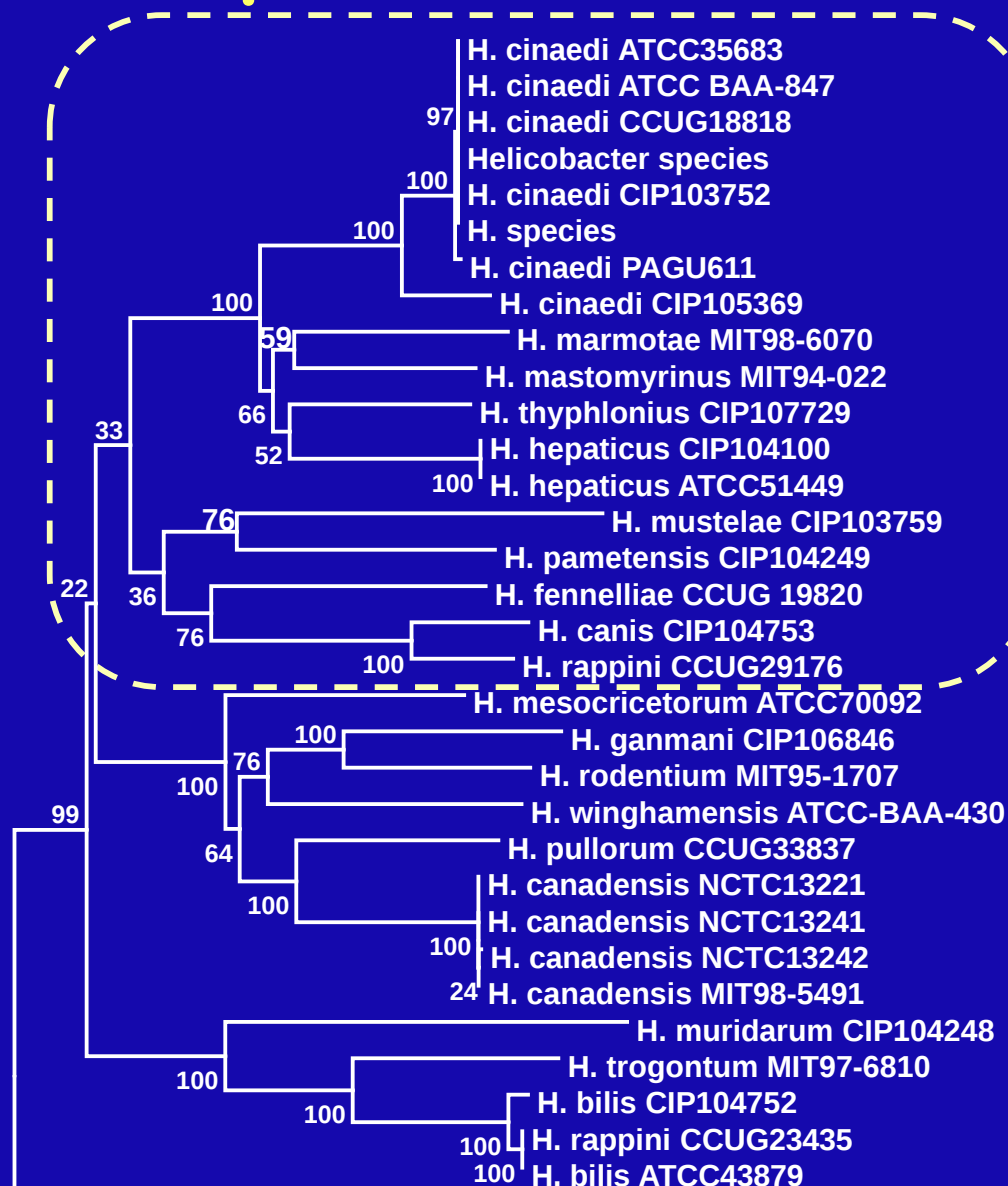
# Results - Phylogenetic tree of *gyrA*

## Enterohepatic *Helicobacter* species



Unsheated *Helicobacter* species

# Results - Phylogenetic tree of *gyrA* Enterohepatic *Helicobacter* species



Lower bootstrap values

# PCR design for *Helicobacter* species identification



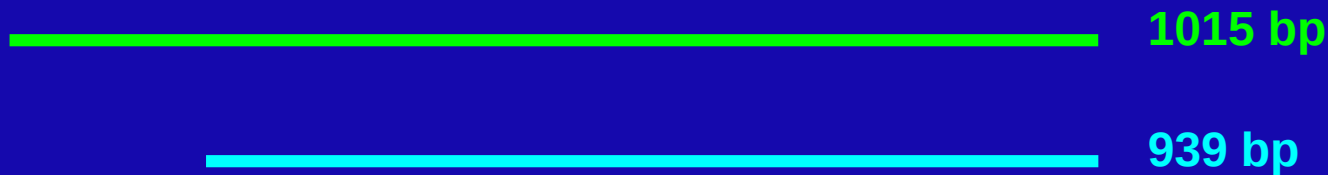
QRDR: Quinolones Resistance Determining Region

# PCR design for *Helicobacter* species identification

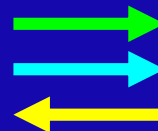


QRDR: Quinolones Resistance Determining Region

# PCR design for *Helicobacter* species identification



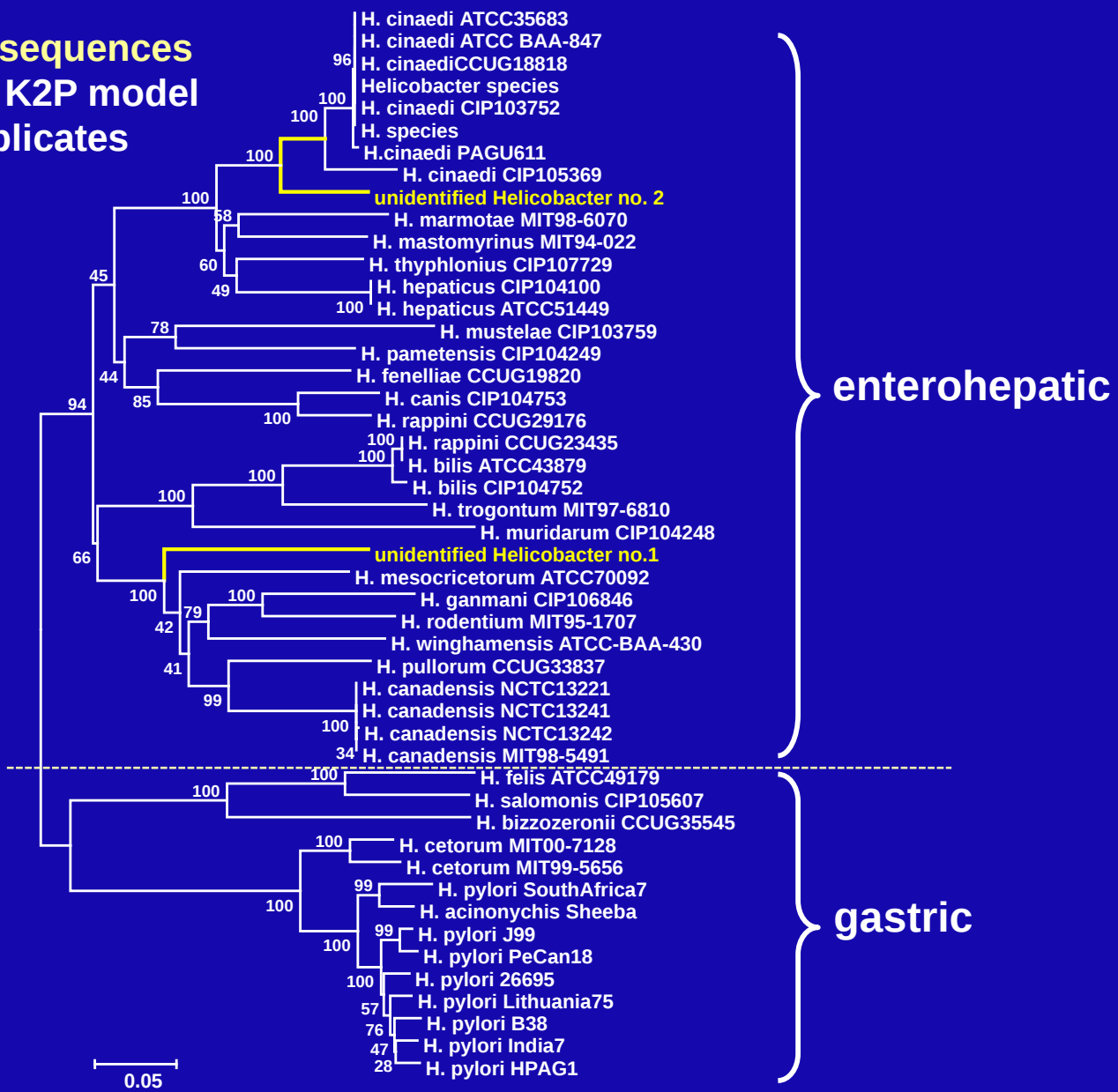
Sequencing



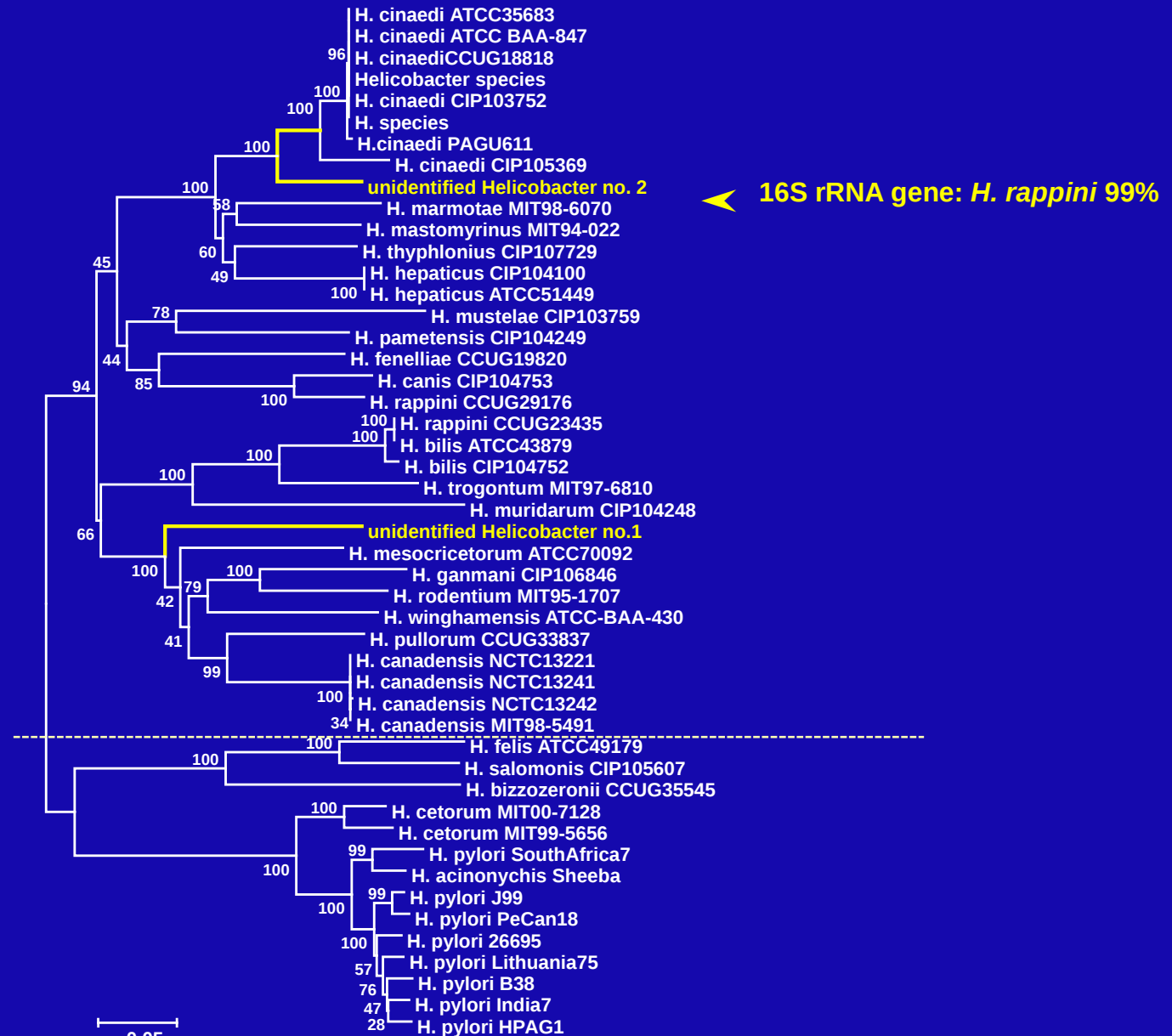
QRDR: Quinolones Resistance Determining Region

# Results - Identification of clinical isolates

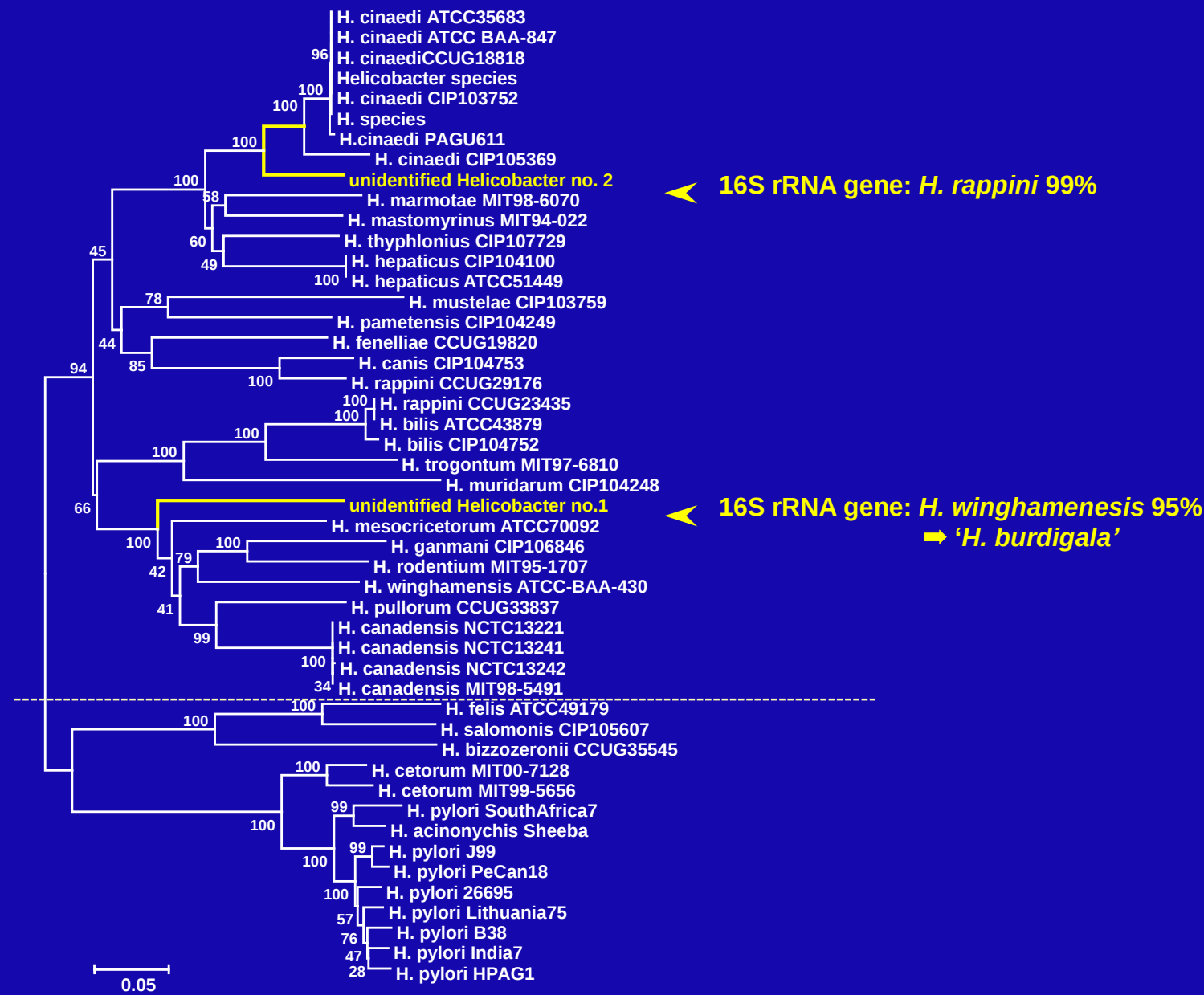
New tree : 49 *gyrA* sequences  
Neighbour-Joining K2P model  
Bootstrap 1,000 replicates



# Results - Identification of clinical isolates



# Results - Identification of clinical isolates



# Conclusion

- *gyrA* gene is a pertinent target for *Helicobacter* phylogeny
- A *gyrA*-based PCR easy to implement was developed
  - ➡ New tool for *Helicobacter* species phylogeny



# High worldwide conservation of a *Helicobacter pylori* outer membrane protein gene, *homD*

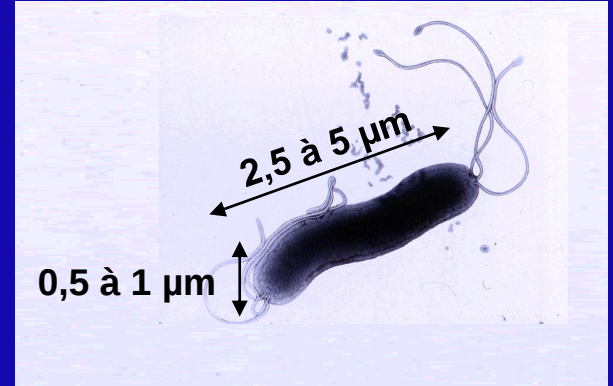


INSTITUTO NACIONAL DE SAÚDE  
Dr. Ricardo Jorge



Faculdade de Ciências, Universidade de Lisboa

# *Helicobacter pylori*



- Gram négatif
- Microaérophile
- Bactérie de forme hélicoidale et motile
- Isolée pour la première fois en 1984
- Habitat : muqueuse gastrique de l'homme
- Uréase : urée  $\rightarrow$  ammoniacque +  $\text{CO}_2$  (pH7)



**Survie au milieu acide de l'estomac**

# Infection chronique à *H. pylori*

- Prévalence de l'infection :
  - 50% de la population mondiale
  - 25% de la population adulte française



**Contamination au cours de la petite enfance**

# *Moyens de lutte contre l'infection à *Helicobacter pylori**

- **Aucun vaccin disponible**
- **Traitement : 2 antibiotiques + IPP**



**Augmentation de la résistance aux antibiotiques**

**→ Développer de nouvelles cibles thérapeutiques préventives et curatives**

# *Infection à H. pylori et pathologies*

Muqueuse gastrique normale

*H. pylori* — — → ↓

Gastrite chronique

100%

Asymptomatique

AGE

PREVALENCE

# Infection à *H. pylori* et pathologies

Muqueuse gastrique normale

*H. pylori* — — → ↓

Gastrite chronique

100%

Ulcère duodénal



Ulcère gastrique



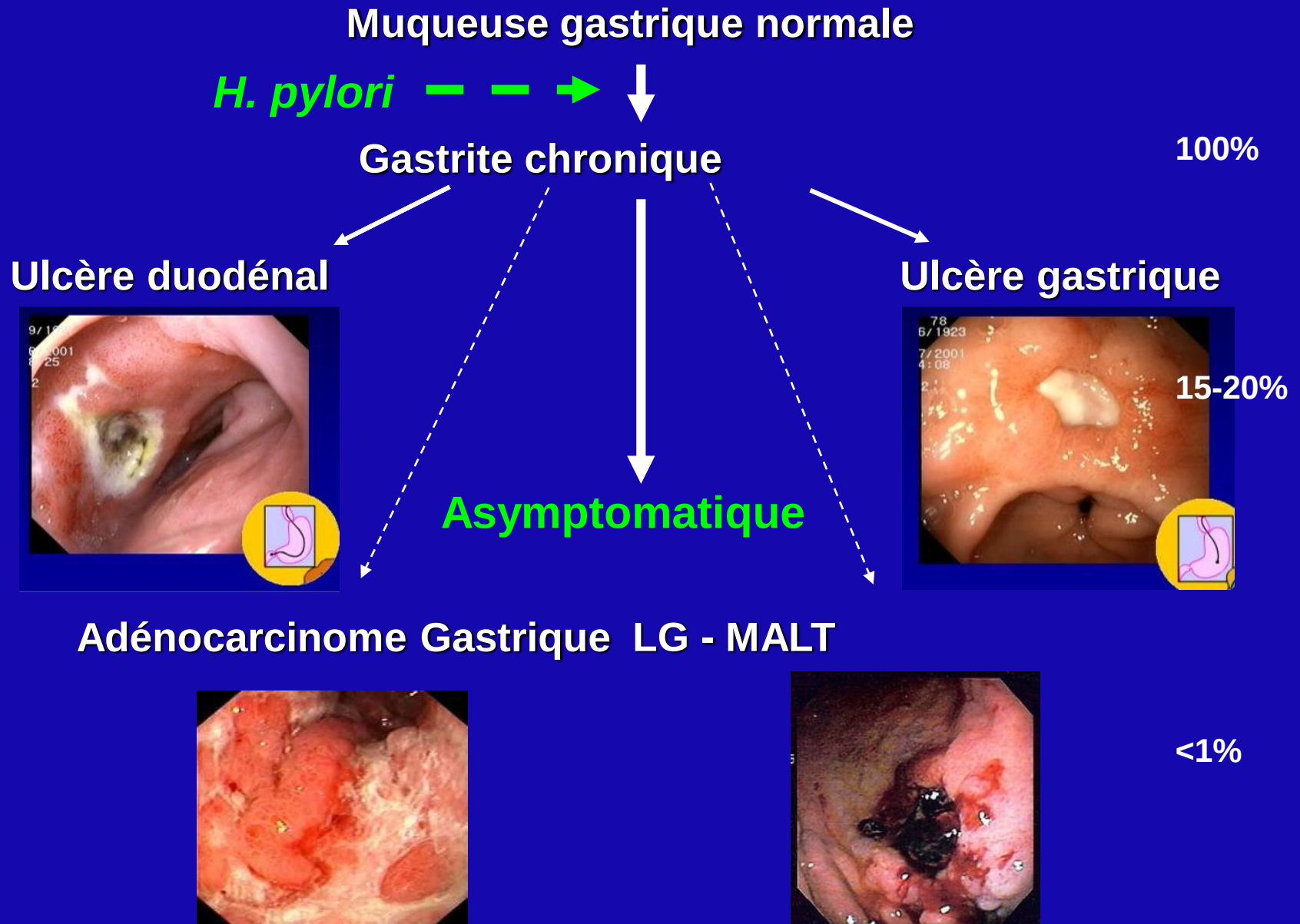
15-20%

Asymptomatique

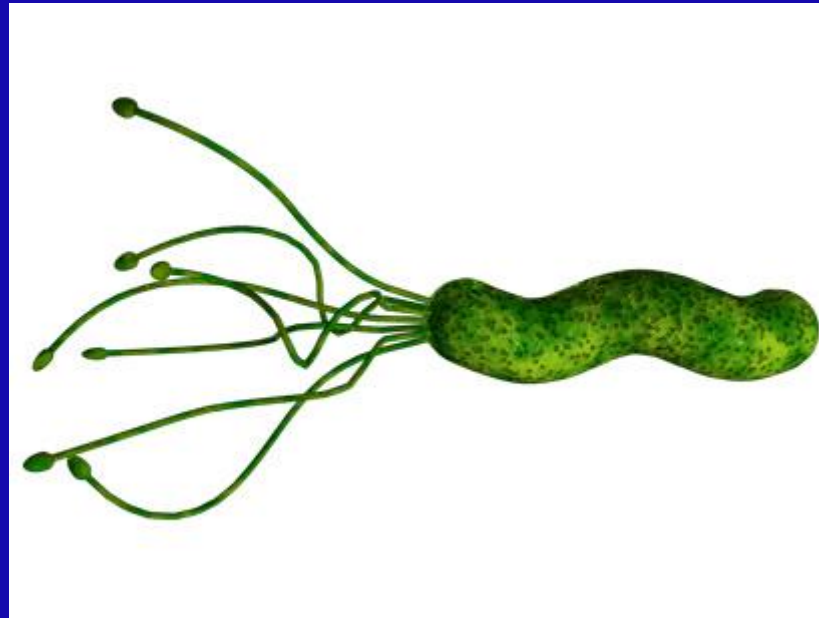
AGE

PREVALENCE

# Infection à *H. pylori* et pathologies



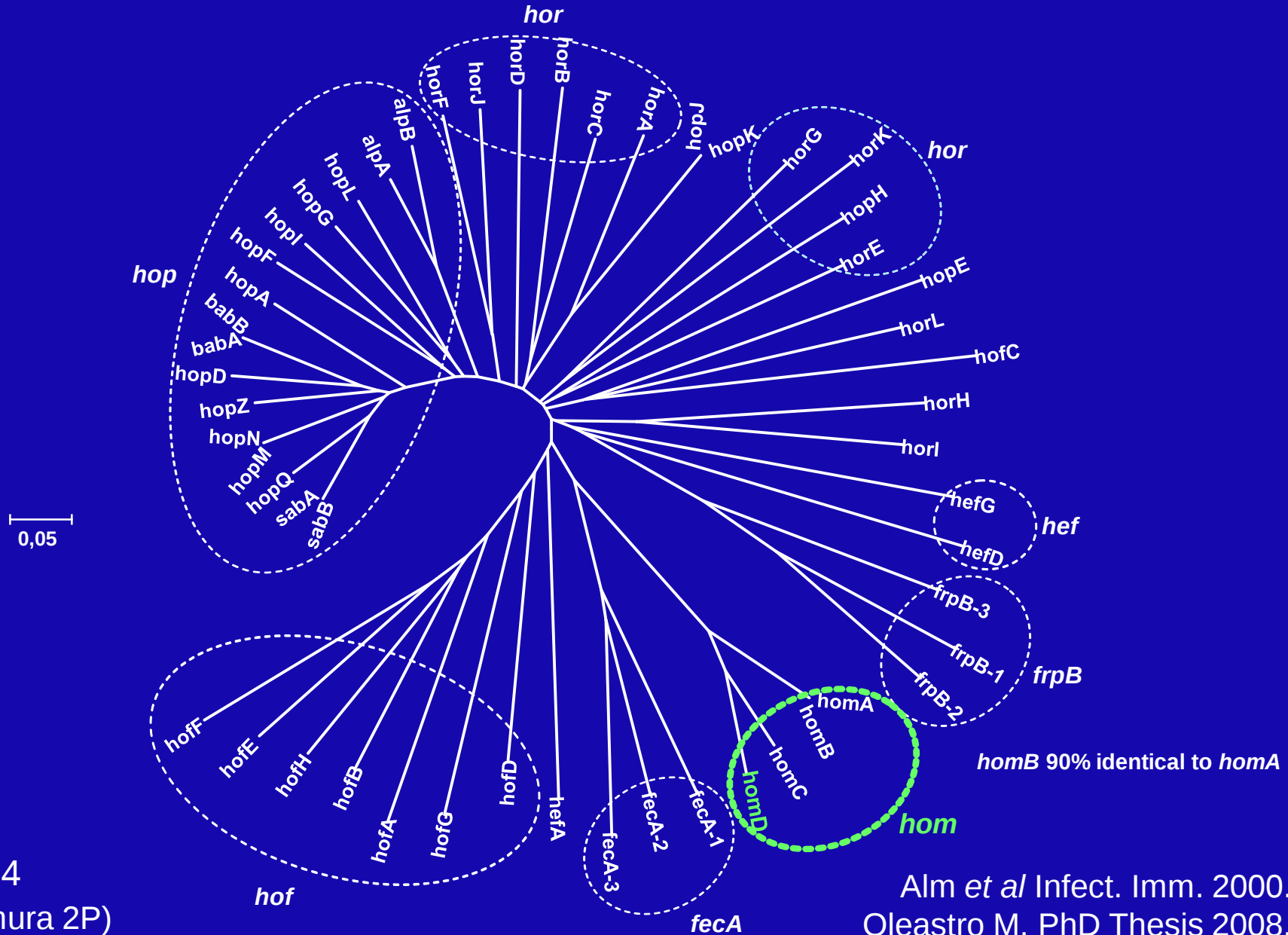
**1994 : Infection à *H. pylori* classée carcinogène de type I  
(niveau maximal) par le Centre International de  
Recherche contre le Cancer (agence de l' OMS)**





**La découverte du rôle de *H. pylori* dans la maladie ulcéreuse par Robin Warren and Barry J. Marshall a été récompensée en 2005 par le prix Nobel de médecine**

# Outer membrane protein in strain J99



# Aims

To investigate:

- ⇒ the diversity and phylogeny of *homD* on a large panel of *H. pylori* strains (patients from different geographical origins with different gastric diseases)
- ⇒ the expression and antigenicity of HomD

# Aims

To investigate:

- ⇒ the diversity and phylogeny of *homD* on a large panel of *H. pylori* strains (patients from different geographical origins with different gastric diseases)
- ⇒ the expression and antigenicity of HomD

# Materials

## Sequences and bacterial strains:

- 72 *homD* sequences from *H. pylori* complete genome (NCBI)
- 523 *H. pylori* clinical strains isolated from patients presenting different gastric diseases:
  - Gastritis
  - Peptic ulcer disease
  - Gastric carcinoma
  - MALT lymphoma

# Materials - origin of the strains (n=523)

---

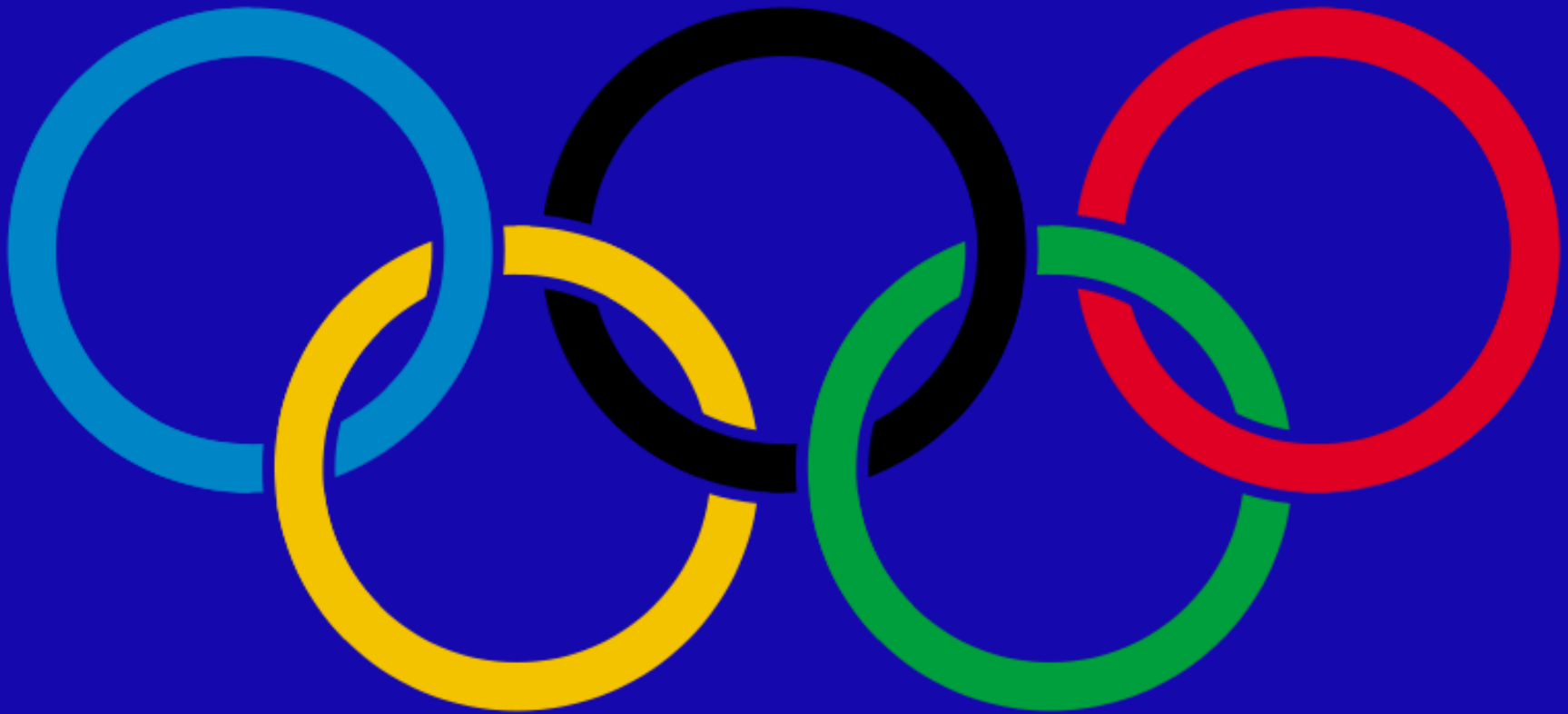
Origin\*

No. of strains

Origin

No. of strains

---



# Materials - origin of the strains (n=523)

Origin\*

No. of strains

Origin

No. of strains

## Oceania

n=24



|                  |    |
|------------------|----|
| Australia        | 12 |
| Indonesia        | 1  |
| New Caledonia    | 10 |
| Papua New Guinea | 1  |

(3 Aborigènes & 11 Maori)

## Asia

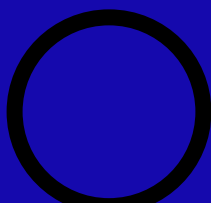
n=137



|             |    |
|-------------|----|
| China       | 37 |
| India       | 17 |
| Japan       | 42 |
| Malaysia    | 14 |
| Singapore   | 7  |
| South Korea | 15 |
| Taiwan      | 5  |

## Africa

n=144



|              |    |
|--------------|----|
| Benin        | 3  |
| Burkina Faso | 11 |
| Egypt        | 3  |
| Gambia       | 1  |
| Senegal      | 71 |
| South Africa | 55 |

# Materials - origin of the strains (n=523)

Origin\*

No. of strains

Origin

No. of strains

## America

n=121

Argentina

2

Brazil

7

Canada

9

Colombia

25

El Salvador

1

Mexico

3

Peru

10

USA

63

Venezuela

1

## Europe

n=100

Croatia

13

France

26

Germany

12

Italy/Spain

2

Lithuania

1

Portugal

22

Russia

4

Sweden

14

United Kingdom

5

8 Amerindians & 2 Inuvialuits

Ötzi

the Tyrolean 5300-year-old Iceman

Similaun Man, the Man from Hauslabjoch



# Methods

## Sequence analysis:

- Downloading
  - complete genomes
  - whole genome shotgun contigs
- PCR (amplification of *homD* locus with external primers) & sequencing

## Bioinformatic analyses of *homD* and HomD deduced protein:

- Similarity plots
  - *BioEdit Sequence Alignment Editor* (Version 7.0.1)
  - *SimPlot* (Version 3.5.1)
- Phylogenetic trees
  - *clustalW*
  - *MEGA 4.1*
- Hydrophilicity, antigenic index and surface probability
  - *Protean software in the DNASTAR package* (Version 5.0)

# Results

## Sequence analysis:

- All (but two strains) harboured a complete *homD* gene at a conserved locus
- *homD* gene varies from 2,223 to 2,271 nt (740 to 756 residue corresponding protein)
- *homD* does not seem to be regulated by a slipped-strand mispairing mechanism (“on/off” regulation)

## Bioinformatic analyses of *homD* and HomD deduced protein:

- High similarity at nucleotide (94.8%) and amino acid (95.7%) levels
- Frequency of synonymous substitutions ( $K_s$ ) =  $0.167 \pm 0.01$ 
  - > frequency of non-synonymous substitutions ( $K_a$ ) =  $0.023 \pm 0.003$ 
    - ⇒ conservation at the protein level ( $K_a/K_s < 1$ , purifying selection  $P_{Z\text{-Test}} < 0.001$ )  
 $K_a/K_s = 0.624$

## Results - Variable region (5' extremity)

| nt position | 172                            | 219                                |
|-------------|--------------------------------|------------------------------------|
| ZA-SouthAf  | AAACCCCAACAAAGGAACAGCCCAAACCC  | AAACCCAACCCAGCCACCCCGCAAAGCACCTAT  |
| NC-NC34     | CAACTCAAGAAAAACCCAAACC         | AAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| NC-NC35     | CAACTCAAGAAAAACCCAAACC         | AAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| NC-NC51     | CAACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAGCACTTAT  |
| NC-NC52     | CAACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAGCACTTAT  |
| NC-NC55     | CAACTCAAGAAAAACCCAAACC         | AAAACCCA-----TTACCCCTCAAAGCACTTAT  |
| PE-PeCan4   | CCACTCAAGAAAAACCCAAACC         | AAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| PE-Cuz20    | CAACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| PE_Amerind  | CAACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| PE-Sat464   | CAACTCAAGAAAAACCCAAACC         | CAAGCCCA-----TTACCCCTCAAAGCACCTAT  |
| VE_ameridi  | CAACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| NO-35A      | CAACTCAAGAAAAACCCAAACC         | AAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| JP-F30-DU   | CAACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| JP-F57_GC   | CAACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| JP-F16_Ga   | CAACTCAAGAAAAACCCAAACC         | AAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| KR-51_DU    | CCACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAACACCTAT  |
| SN-DAK42    | CCACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| SN-DAK72    | CCACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| SN-DAK8     | CCACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| SN-DAK60    | CCACTCAAGAAAAACCCAAACC         | AAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| SN-DAK73    | CCACTCAAGAAAAACCCAAACC         | AAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| SN-DAK65    | CCACTCAAGAAAAACCCAAACC         | CAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| SN-DAK97    | CCACTCAAGAAAAACCCAAACC         | AAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| US-J99_DU   | CCACTCAAGAAAAACCCAAACC         | AAAACCCA-----TTACCCCTCAAAGCACCTAT  |
| SN-DAK3     | CCACTCAAGAAAAACCCAAACC         | CA-----TTACCCCTCAAAGCACCTAT        |
| SN-DAK35    | CCACTCAAGAAAAACCCAAACC         | CA-----TTACCCCTCAAAGCACCTAT        |
| SN-DAK24    | CCACTCAAGAAAAACCCAAACC         | TAAGCCCAACCCATTACCCCTCAAAGCACCTAT  |
| GM-Gambia9  | CCACTCAAGAAAAACCCAAACC         | CAAACCCAAGGCCATTACCCCTCAAAGCACCTAT |
| SN-DAK16    | CCACTCAAGAAAAACCCAAACC         | CAAACCTAAGGCCATTACCCCTCAAAGCACCTAT |
| SN-DAK19    | CCACTCAAGAAAAACCCAAACC         | CAAACCCAACCCATTACCCCTCAAAGCACCTAT  |
| FR-B25_ML   | CCACTCAAGAAAAACCCAAACC         | AAAGCCAACCCATTACCCCTCAAAGCACCTAT   |
| UK-26695_G  | CAACTCAAGAAAAACCTAAACCTAAACCAA | GCCAAACCCATTACCCCTCAAAGCACCTAT     |
| FR-B38_ML   | CCACTCAAGAAAAACCCAAACC         | ACCCAACCCATTACCCCTCAAAGCACCTAT     |
| SE-HPAG1_G  | CCACTCAAGAAAAACCTAAACCTAAACCAA | GCCAAACCCATTACCCCTCAAAGCACCTAT     |
| US-B8_GU    | CAACTCAAGAAAAACCCAAACC         | GCCAAACCCATTACCCCTCAAAGCACCTAT     |
| LT-Lithuan  | CCACTCAAGAAAAACCCAAACC         | GCCAAACCCATTACCCCTCAAAGCACCTAT     |
| DE-P12_DU   | CCACTCAAGAAAAACCCAAACC         | GCCAAACCCATTACCCCTCAAAGCACCTAT     |
| KR-52       | CAACTCAAGAAAAACCCAAACC         | ACCAAACCCATTACCCCTCAAAGCACCTAT     |
| JP-F32_GC   | CCACTCAAGAAAAACCTAAACCTAAACCAA | ACCAAACCCATTACCCCTCAAAGCACCTAT     |

# Results - Variable number of KP repeats (2 to 9)

| aa position | 1              | 58                | 75                          | 100                                               |
|-------------|----------------|-------------------|-----------------------------|---------------------------------------------------|
| ZA-SouthAfr | MLKLASKTICLSLI | SSFTALEAYPKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAQT   | PNKEQPKPKPKP-----ATPQSTYGKYYISQSTILKNATELFAE      |
| NC-NC34     | MLKLVSKTICLSLI | SSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE           |
| NC-NC35     | MLKLVSKTICLSLI | SSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE           |
| NC-NC51     | MLKLVSKTICLSLI | SSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE         |
| NC-NC52     | MLKLVSKTICLSLI | SSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQNEQTIATTQ   | EKKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE         |
| NC-NC55     | MLKLVSKTICLSLI | SSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE         |
| PE-PeCan4   | MLKLASKTICLSLI | SSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE       |
| IN-India7   | MLKLASKTICLSLI | GSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE       |
| PE-Cuz20    | MLKLASKTICLSLI | GSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTKEQTIATTQ  | EKKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE       |
| PE-Sat464   | MLKLASKTICLSLI | GSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTKEQTIATTQ  | EKKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE       |
| VE_ameridi  | MLKLASKTICLSLI | GSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTKEQTIATTQ  | EKKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE       |
| PE_Amerind  | MLKLASKTICLSLI | GSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTKEQTIATTQ  | EKKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE           |
| KR-51_DU    | MLKLVSKTICLSLI | SSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKP-----TPQNTYGKYYISQSTILKNATELFAE         |
| NO-35A      | MLKLVSKTICLSLI | SSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE       |
| JP-F30_DU   | MLKLVSKTICLSLI | SSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP--KP-----TPQSTYGKYYISQSTILKNATELFAE |
| JP-F57_GC   | MLKLASKTICLSLI | SSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE           |
| JP-F16_Ga   | MLKLVSKTICLSLI | SSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE           |
| SN-DAK42    | MLKLASKTICLSLI | SSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE     |
| SN-DAK60    | MLKLASKTICLSLI | SSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTKEQTIATTQ  | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE     |
| SN-DAK65    | MLKLASKTICLSLI | SSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE     |
| SN-DAK8     | MLKLASKTICLSLI | SSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE     |
| SN-DAK97    | MLKLASKTICLSLI | SSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE     |
| US-J99_DU   | MLKLASKTICLSLI | SSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE     |
| SN-DAK73    | MIKLASKTICLSLI | ISFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE     |
| SN-DAK3     | MLKLASKTICLSLI | SSFTAVEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE       |
| SN-DAK35    | MLKLASKTICLSLI | SSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE       |
| SN-DAK72    | MLKLASKTICLSLI | SSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKPKP----TPQSTYGKYYISQSTILKNATELFAE    |
| GM-Gambia9  | MLKLASKTICLSLI | SSFTAAEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKPKP----TPQSTYGKYYISQSTILKNATELFAE    |
| SN-DAK24    | MLKLASKTICLSLI | ISFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKPKP----TPQSTYGKYYISQSTILKNATELFAE    |
| FR-B25_ML   | MLKLASKTICLSLI | SSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKPKP----TPQSTYGKYYISQSTILKNATELFAE    |
| SN-DAK16    | MLKLASKTICLSLI | SSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE   |
| SN-DAK19    | MLKLASKTICLSLI | SSFTAVEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKPKP----TPQSTYGKYYISQSTILKNATELFAE    |
| UK-26695_G  | MLKLASKTICLSLI | GSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTKEQTIATTQ  | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTVLKNATELFAE     |
| SE-HPAG1_G  | MLKLASKTICLSLI | GSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTVLKNATELFAE     |
| FR-B38_ML   | MLKLASKTICLSLI | GSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE     |
| LT-Lithuan  | MLKLASKTICLSLI | GSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTVLKNATELFAE     |
| US-B8_GU    | MLKLASKTICLSLI | GSFTALEAYQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTVLKNATELFAE     |
| DE-P12_DU   | MLKLASKTICLSLI | GSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE     |
| PE-SJM180   | MLKLASKTICLSLI | GSFTAVEAFQKHQKDGF | FFIEAGFETGLLQGTQTQEQTIAATTQ | EKKPKPKPKPKPKPKPKP-----TPQSTYGKYYISQSTILKNATELFAE |

2 to 9 KP repeats

**KP repeats**

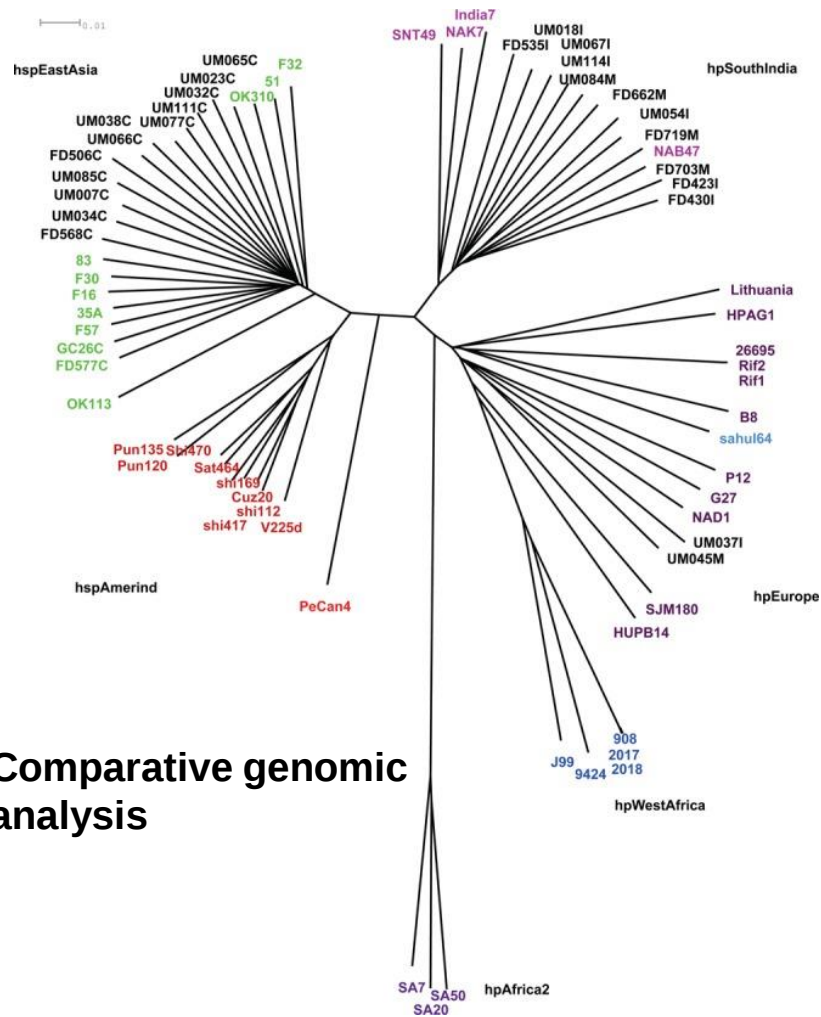
**Hydrophilicity (2.71)**

**Antigenic Index (6.82)**

**Surface probability (2.95)**

K P K P K P K P K P K P K P

# Results - Phylogenetic tree of genome vs HomD

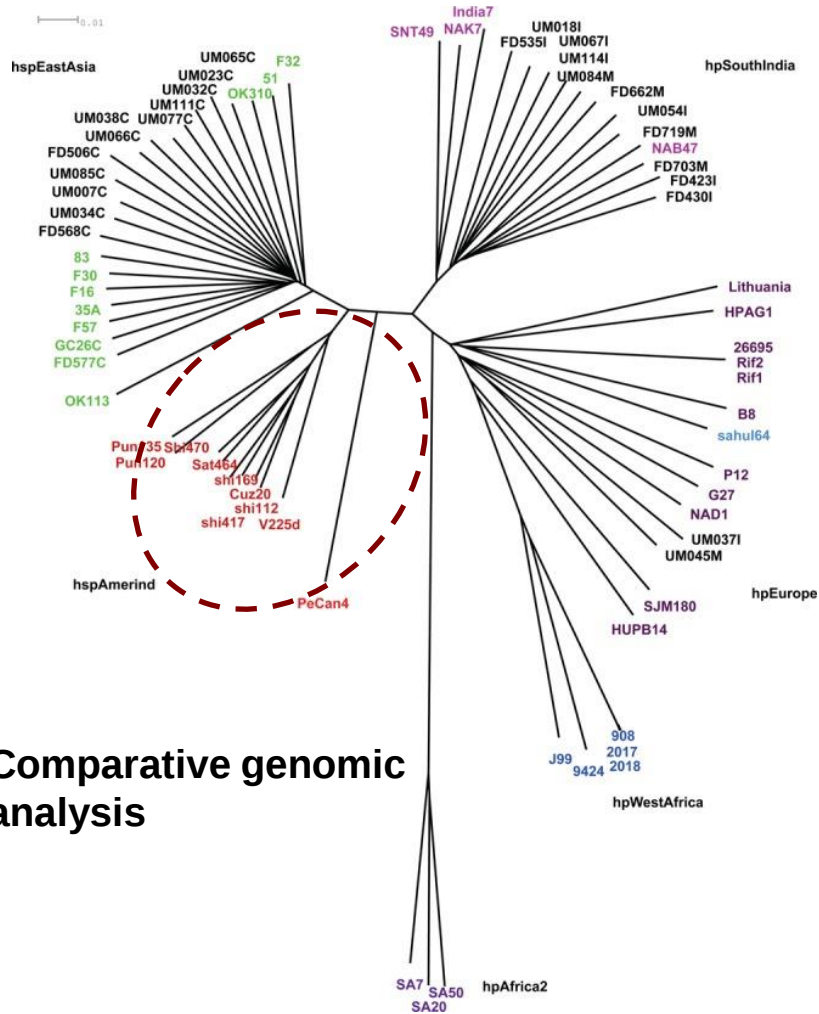


Complete *H. pylori* genomes  
(no contigs)

Comparative genomic  
analysis

Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from  
Malaysia identifies

# Results - Phylogenetic tree of genome vs HomD

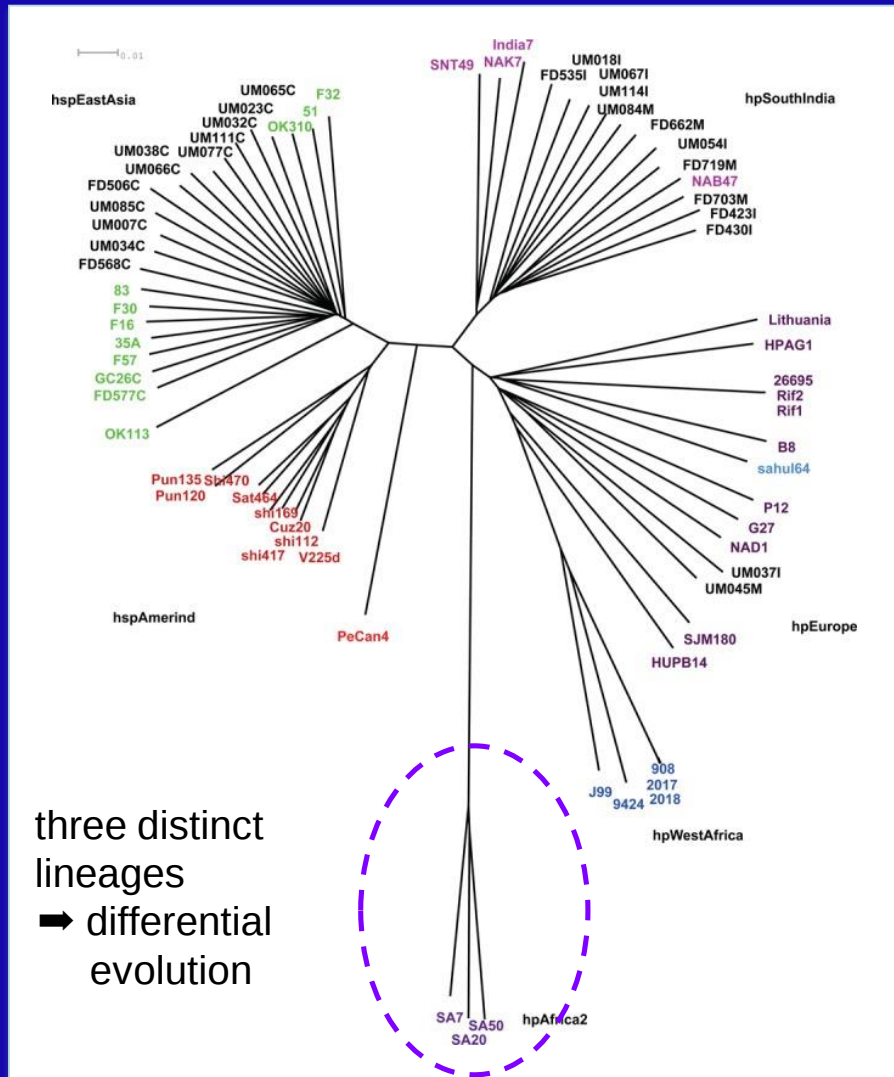


Complete *H. pylori* genomes  
(no contigs)

Comparative genomic  
analysis

Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from  
Malaysia identifies

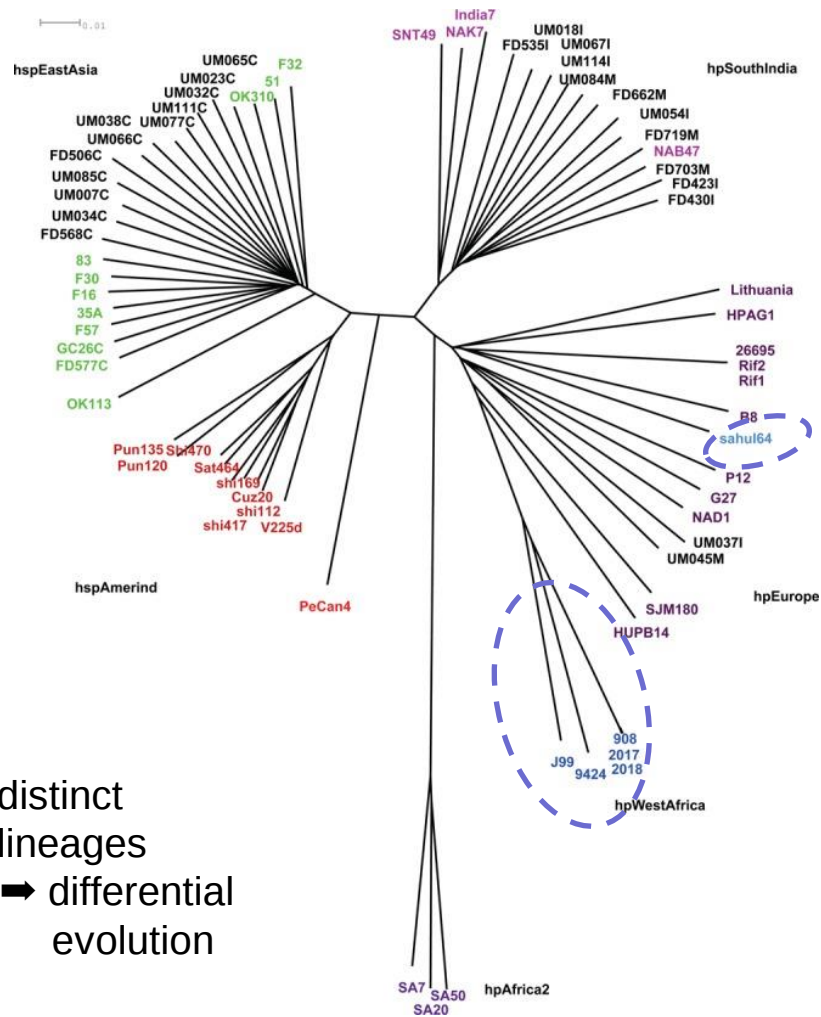
# Results - Phylogenetic tree of genome vs HomD



Complete *H. pylori* genomes  
(no contigs)

Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from  
Malaysia identifies

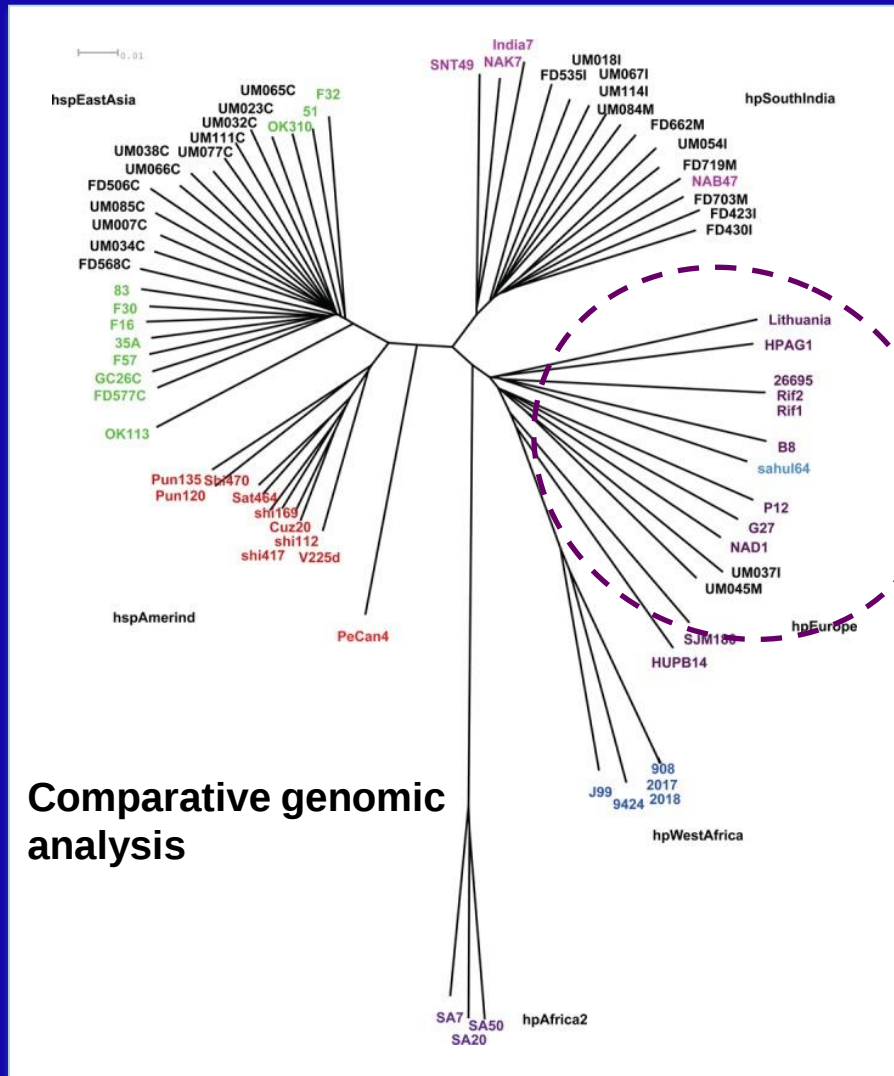
# Results - Phylogenetic tree of genome vs HomD



Complete *H. pylori* genomes  
(no contigs)

Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from  
Malaysia identifies

# Results - Phylogenetic tree of genome vs HomD

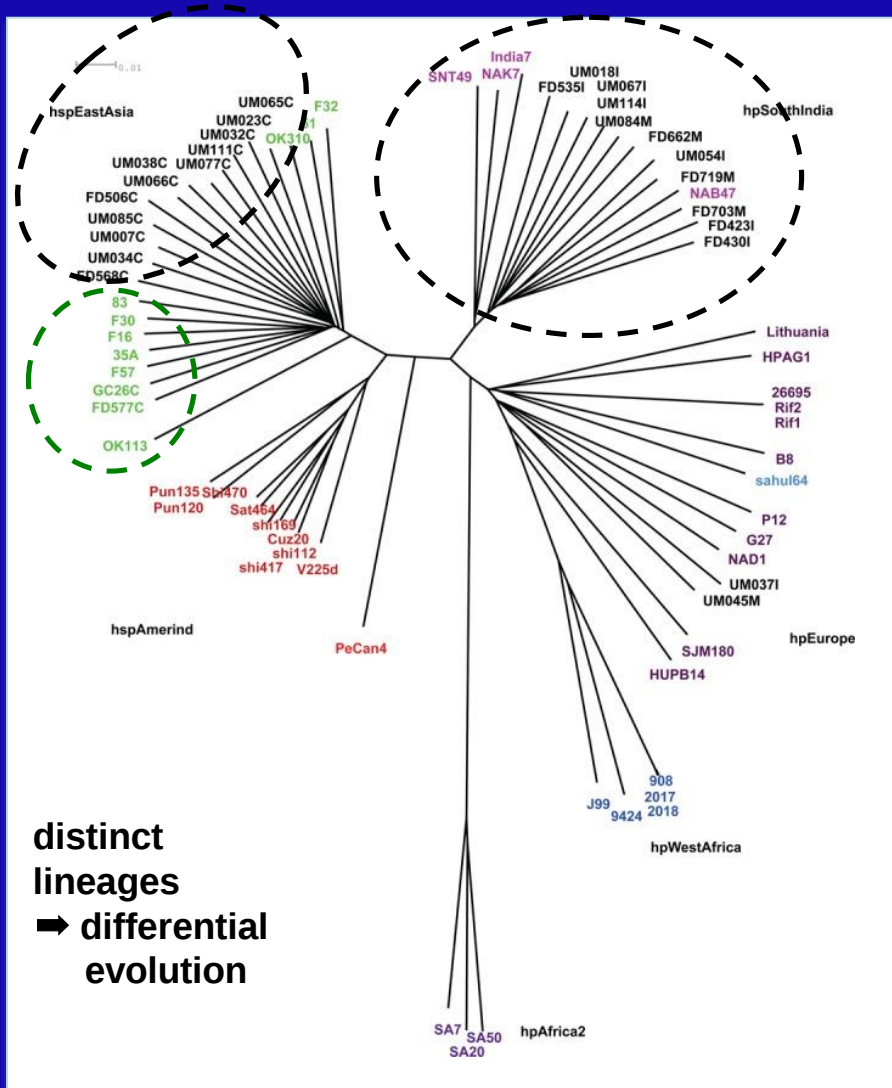


Complete *H. pylori* genomes  
(no contigs)

Comparative genomic  
analysis

Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from  
Malaysia identifies

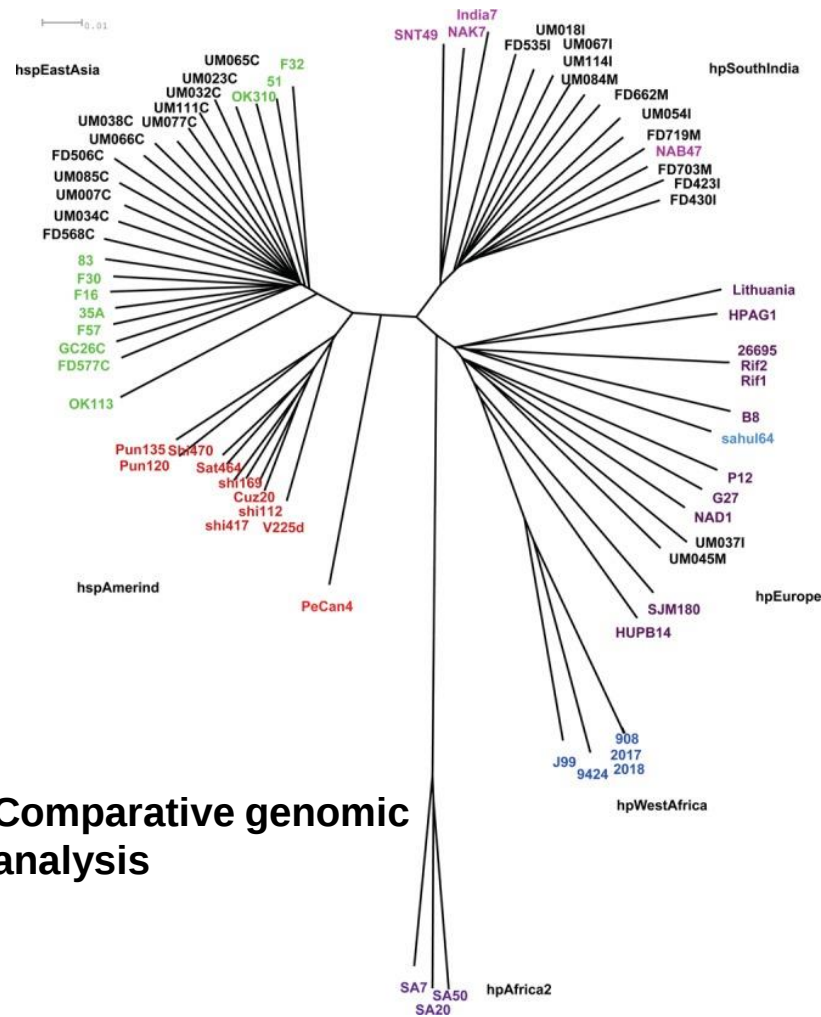
# Results - Phylogenetic tree of genome vs HomD



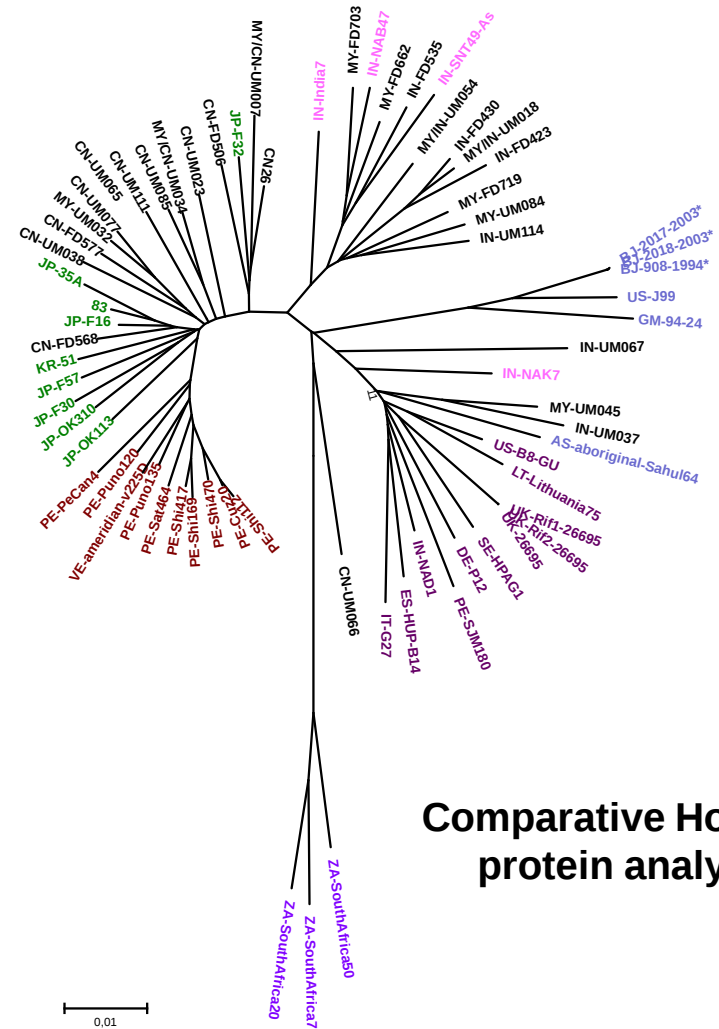
Complete *H. pylori* genomes  
(no contigs)

Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from  
Malaysia identifies

## Results - Phylogenetic tree of genome vs HomD



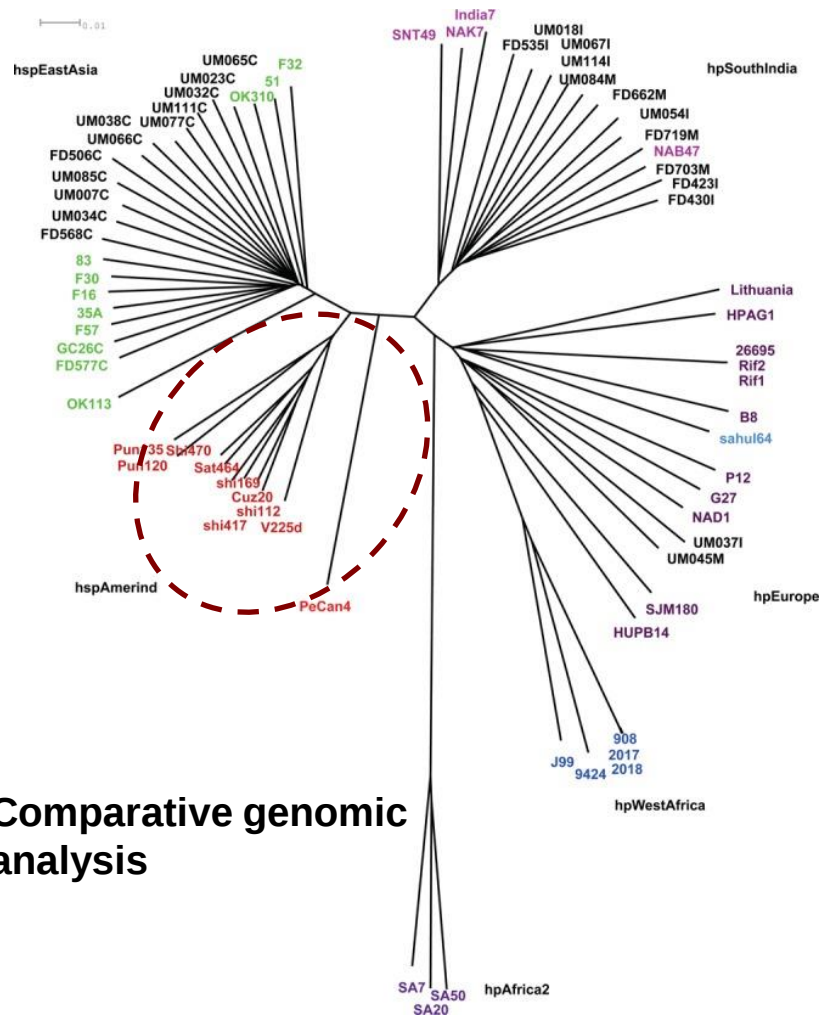
## Comparative genomic analysis



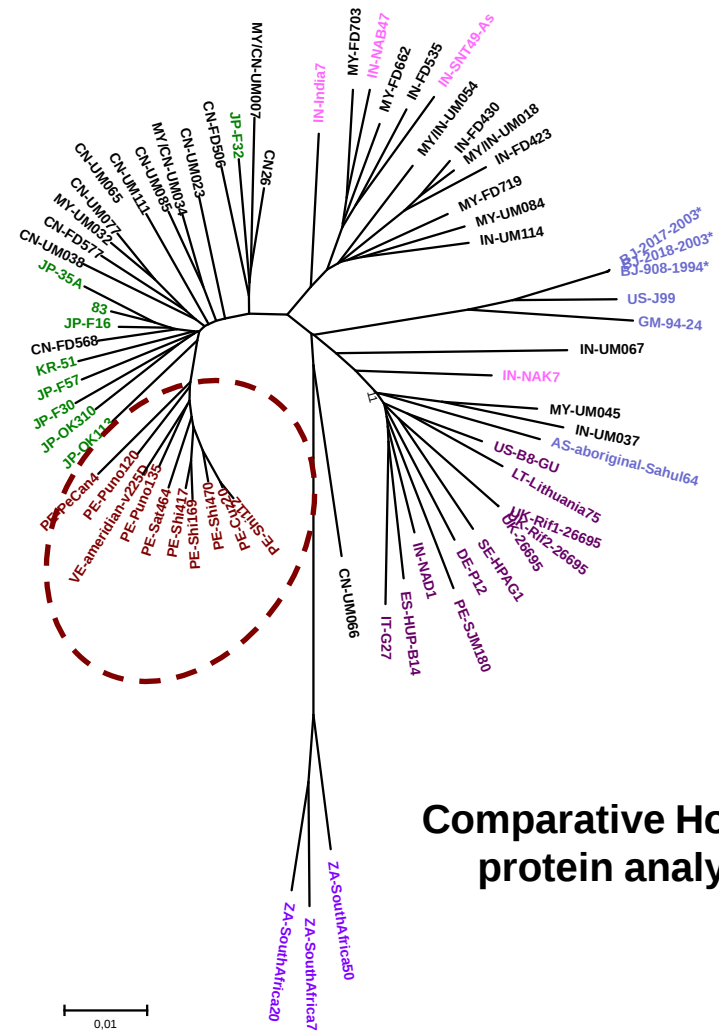
## Comparative HomD protein analysis

Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from  
Malaysia identifies

# Results - Phylogenetic tree of genome vs HomD



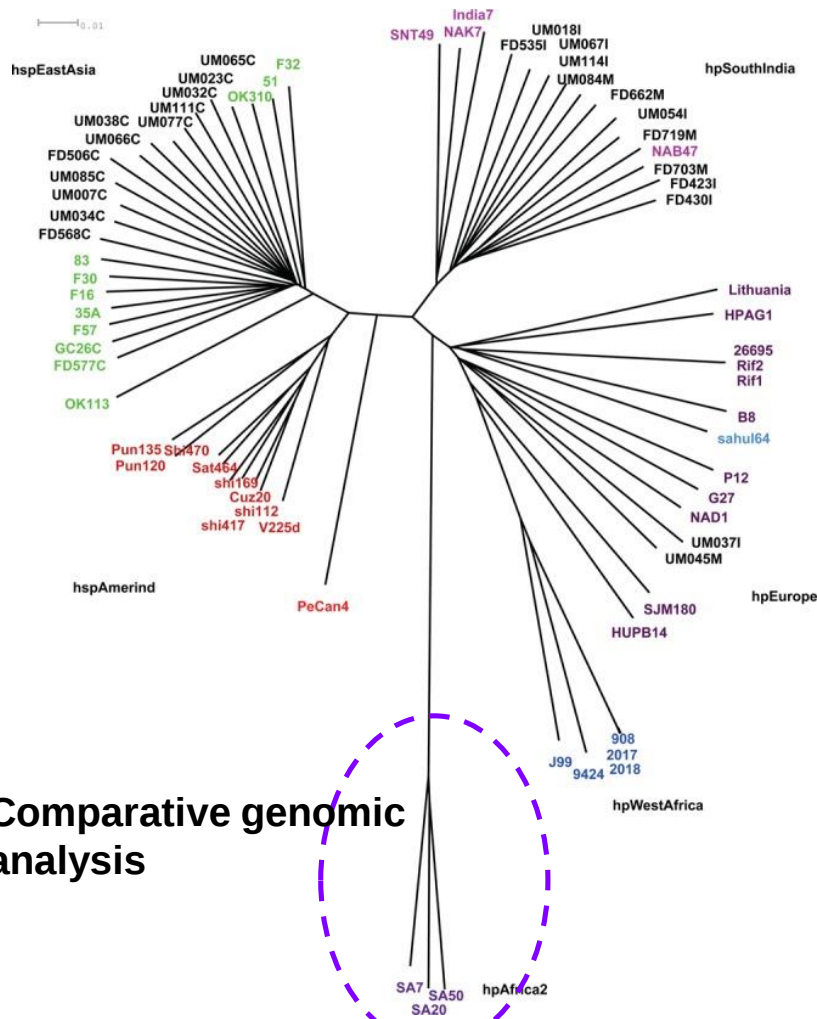
**Comparative genomic analysis**



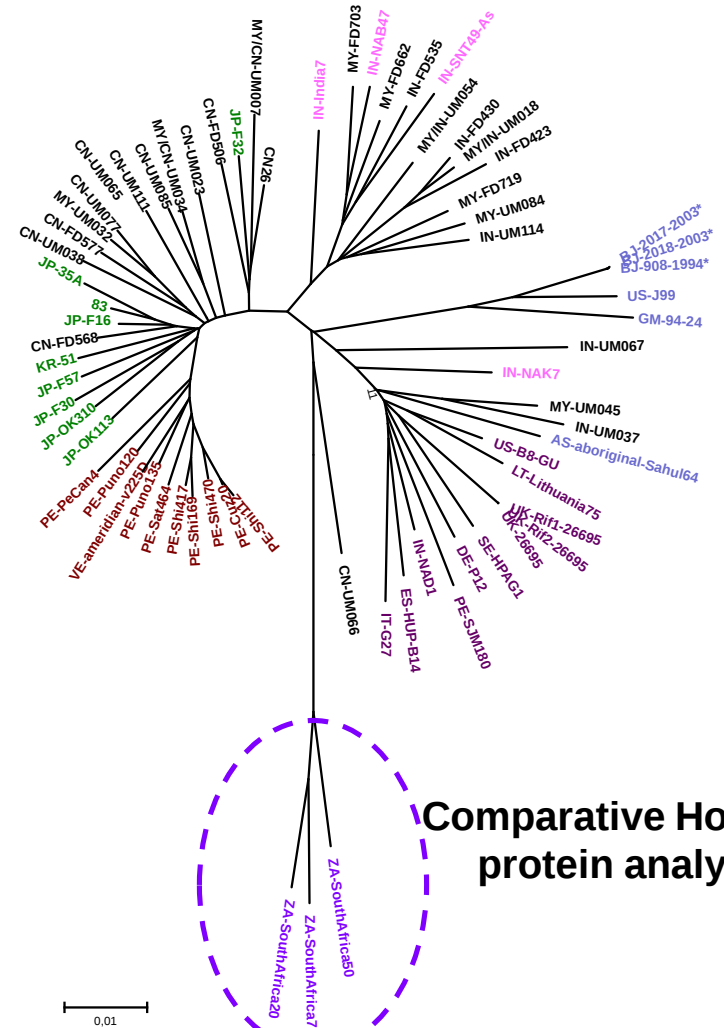
**Comparative HomD protein analysis**

Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from Malaysia identifies

# Results - Phylogenetic tree of genome vs HomD



Comparative genomic analysis



Comparative HomD protein analysis

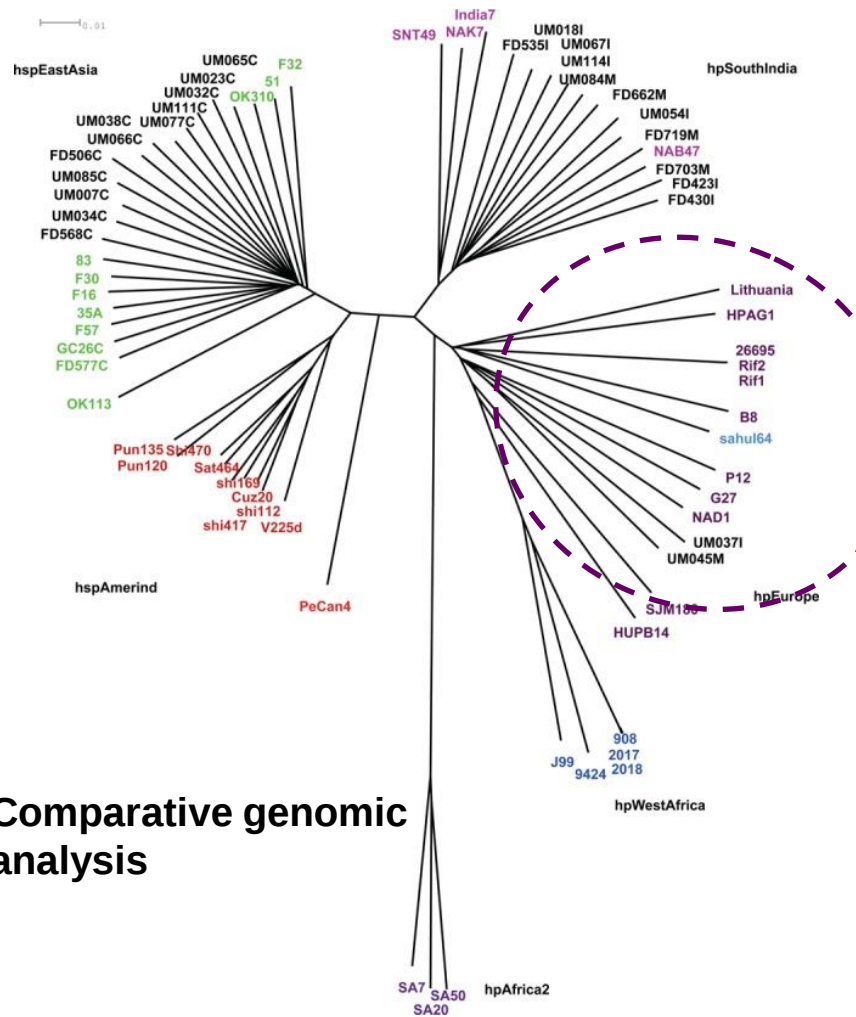
Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from Malaysia identifies

## Results - Phylogenetic tree of genome vs HomD

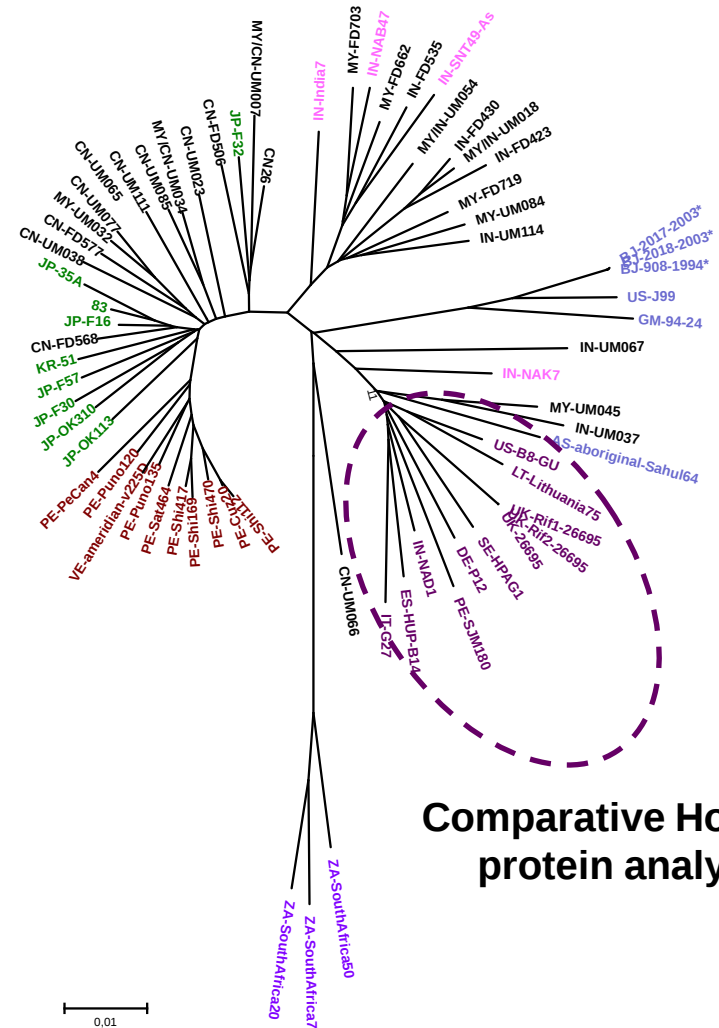


Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from  
Malaysia identifies

## Results - Phylogenetic tree of genome vs HomD



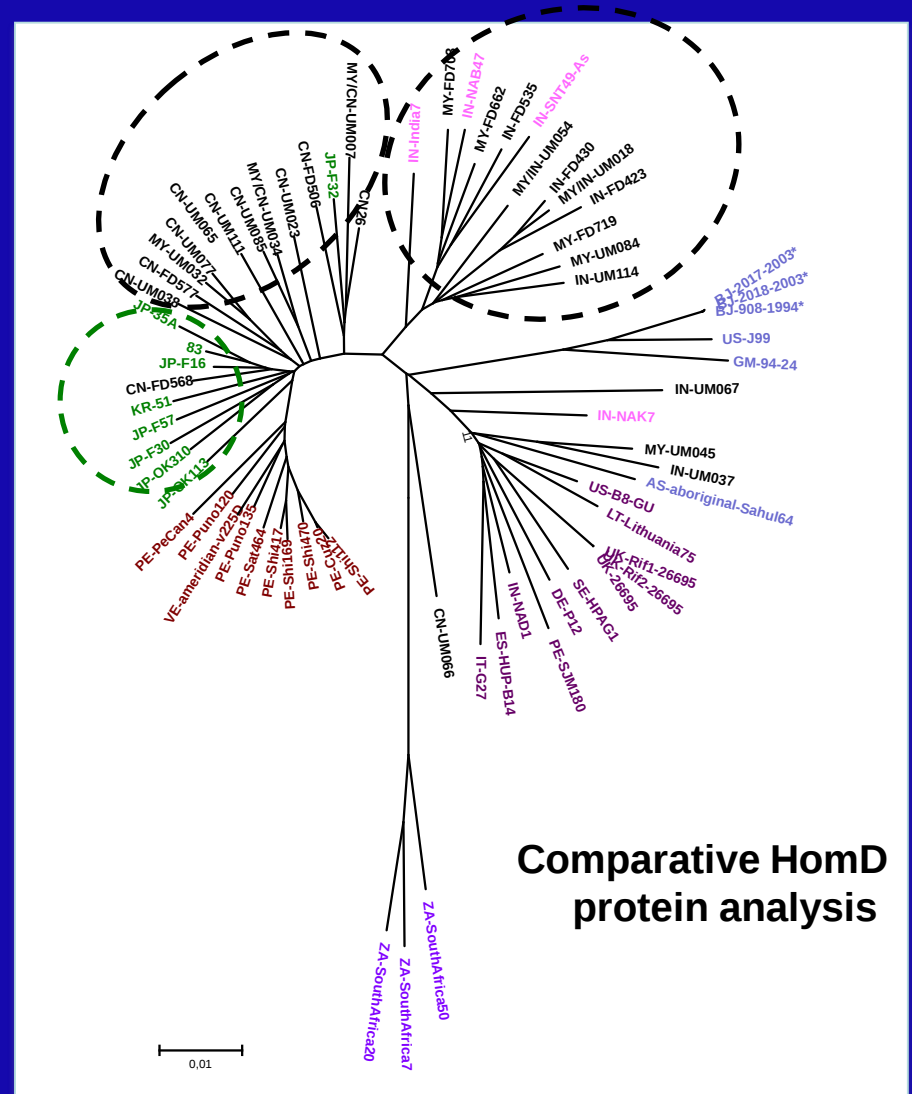
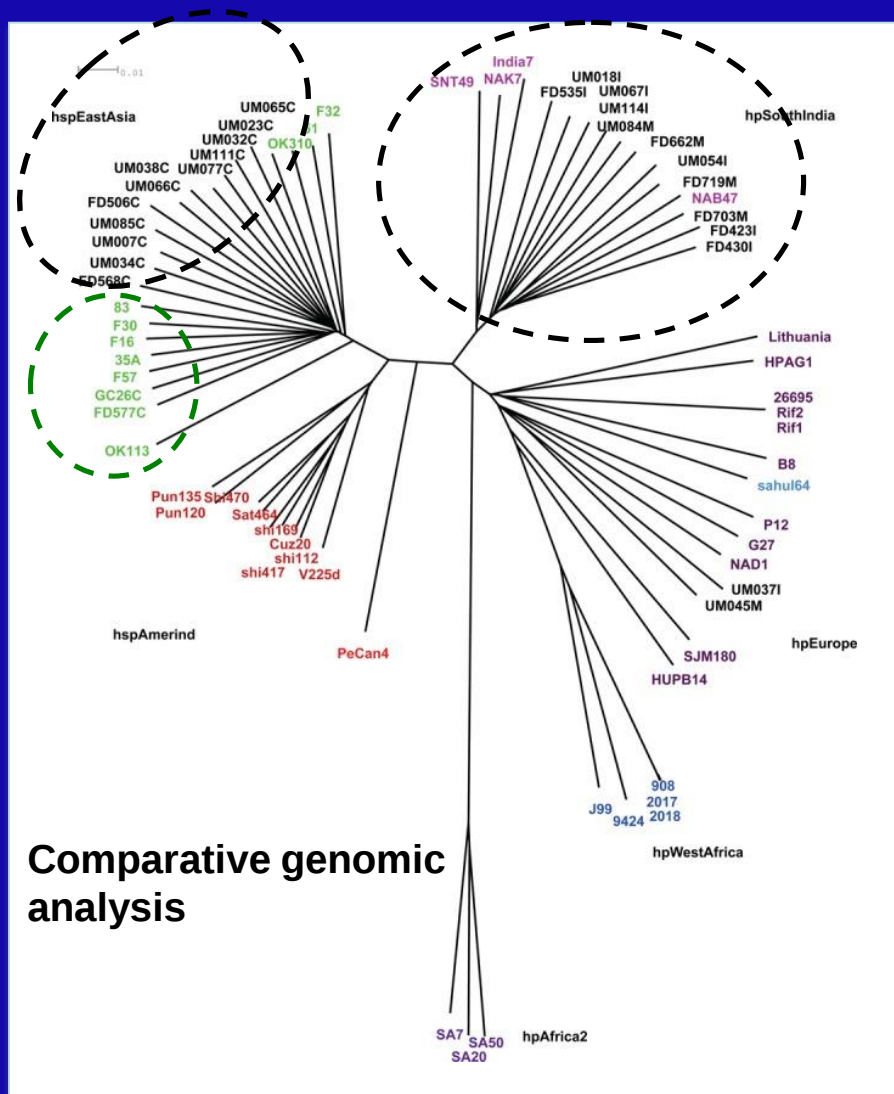
## Comparative genomic analysis



## Comparative HomD protein analysis

Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from  
Malaysia identifies

## Results - Phylogenetic tree of genome vs HomD



Kumar et al. 2015 Nucleic Acids Res.  
Comparative genomic analysis of *Helicobacter pylori* from  
Malaysia identifies

# Results - Corrected MEGA alignment

| DNA Sequences                                |            | Translated Protein Sequences |  |
|----------------------------------------------|------------|------------------------------|--|
| Species/Abbrv                                | Group Name |                              |  |
| 1. Ot11-1ceman                               |            |                              |  |
| 2. 2A-SA164AC/SA165AC/SA171AC/SA172A/SA302AC |            |                              |  |
| 3. 2A-SA222AC                                |            |                              |  |
| 4. 2A-SA252AC                                |            |                              |  |
| 5. 2A-SA224AC                                |            |                              |  |
| 6. 2A-SA29C/SA168AC                          |            |                              |  |
| 7. 2A-SA226AC                                |            |                              |  |
| 8. 2A-SA301AC                                |            |                              |  |
| 9. 2A-SA161AC/SA214AC/SA215C/SA216AC         |            |                              |  |
| 10. 2A-SA30AC                                |            |                              |  |
| 11. 2A-SA31C/SA158AC/SA163AC/210AC/300AC     |            |                              |  |
| 12. 2A-SA29A                                 |            |                              |  |
| 13. 2A-SA194AC                               |            |                              |  |
| 14. 2A-SA213AC                               |            |                              |  |
| 15. 2A-SA45AC                                |            |                              |  |
| 16. 2A-SA46AC                                |            |                              |  |
| 17. 2A-SA47AC                                |            |                              |  |
| 18. 2A-SA169C                                |            |                              |  |
| 19. 2A-SA170AC                               |            |                              |  |
| 20. 2A-SA173AC                               |            |                              |  |
| 21. 2A-SA303C                                |            |                              |  |
| 22. 2A-SA166A                                |            |                              |  |
| 23. 2A-SA172C                                |            |                              |  |
| 24. 2A-SA160A                                |            |                              |  |
| 25. 2A-SA160C                                |            |                              |  |
| 26. 2A-SA162AC                               |            |                              |  |
| 27. 2A-SA253AC                               |            |                              |  |
| 28. 2A-SA227AC                               |            |                              |  |
| 29. 2A-SA251AC                               |            |                              |  |
| 30. 2A-SA144AC                               |            |                              |  |
| 31. 2A-SA155AC                               |            |                              |  |
| 32. 2A-SA144AC(2)                            |            |                              |  |
| 33. 2A-SA146AC                               |            |                              |  |
| 34. 2A-SA156AC                               |            |                              |  |
| 35. 2A-SA157AC                               |            |                              |  |
| 36. 2A-SA35AC                                |            |                              |  |
| 37. 2A-SA36C                                 |            |                              |  |
| 38. 2A-SA34A                                 |            |                              |  |
| 39. 2A-SA34C                                 |            |                              |  |
| 40. 2A-SA40AC                                |            |                              |  |
| 41. 2A-SA37AC                                |            |                              |  |
| 42. 2A-SA221AC                               |            |                              |  |
| 43. 2A-SA220AC                               |            |                              |  |

| DNA Sequences                                |            | Translated Protein Sequences |  |
|----------------------------------------------|------------|------------------------------|--|
| Species/Abbrv                                | Group Name |                              |  |
| 1. Ot11-1ceman                               |            |                              |  |
| 2. 2A-SA164AC/SA165AC/SA171AC/SA172A/SA302AC |            |                              |  |
| 3. 2A-SA222AC                                |            |                              |  |
| 4. 2A-SA252AC                                |            |                              |  |
| 5. 2A-SA224AC                                |            |                              |  |
| 6. 2A-SA29C/SA168AC                          |            |                              |  |
| 7. 2A-SA226AC                                |            |                              |  |
| 8. 2A-SA301AC                                |            |                              |  |
| 9. 2A-SA161AC/SA214AC/SA215C/SA216AC         |            |                              |  |
| 10. 2A-SA30AC                                |            |                              |  |
| 11. 2A-SA31C/SA158AC/SA163AC/210AC/300AC     |            |                              |  |
| 12. 2A-SA29A                                 |            |                              |  |
| 13. 2A-SA194AC                               |            |                              |  |
| 14. 2A-SA213AC                               |            |                              |  |
| 15. 2A-SA45AC                                |            |                              |  |
| 16. 2A-SA46AC                                |            |                              |  |
| 17. 2A-SA47AC                                |            |                              |  |
| 18. 2A-SA169C                                |            |                              |  |
| 19. 2A-SA170AC                               |            |                              |  |
| 20. 2A-SA173AC                               |            |                              |  |
| 21. 2A-SA303C                                |            |                              |  |
| 22. 2A-SA166A                                |            |                              |  |
| 23. 2A-SA172C                                |            |                              |  |
| 24. 2A-SA160A                                |            |                              |  |
| 25. 2A-SA160C                                |            |                              |  |
| 26. 2A-SA162AC                               |            |                              |  |
| 27. 2A-SA253AC                               |            |                              |  |
| 28. 2A-SA227AC                               |            |                              |  |
| 29. 2A-SA251AC                               |            |                              |  |
| 30. 2A-SA144AC                               |            |                              |  |
| 31. 2A-SA155AC                               |            |                              |  |
| 32. 2A-SA144AC(2)                            |            |                              |  |
| 33. 2A-SA146AC                               |            |                              |  |
| 34. 2A-SA156AC                               |            |                              |  |
| 35. 2A-SA157AC                               |            |                              |  |
| 36. 2A-SA35AC                                |            |                              |  |
| 37. 2A-SA36C                                 |            |                              |  |
| 38. 2A-SA34A                                 |            |                              |  |
| 39. 2A-SA34C                                 |            |                              |  |
| 40. 2A-SA40AC                                |            |                              |  |
| 41. 2A-SA37AC                                |            |                              |  |
| 42. 2A-SA221AC                               |            |                              |  |
| 43. 2A-SA220AC                               |            |                              |  |

## Results - Corrected MEGA alignment

Data Edit Search Alignment Web Sequencer Display Help



DNA Sequences Translated Protein Sequences

[illegible]

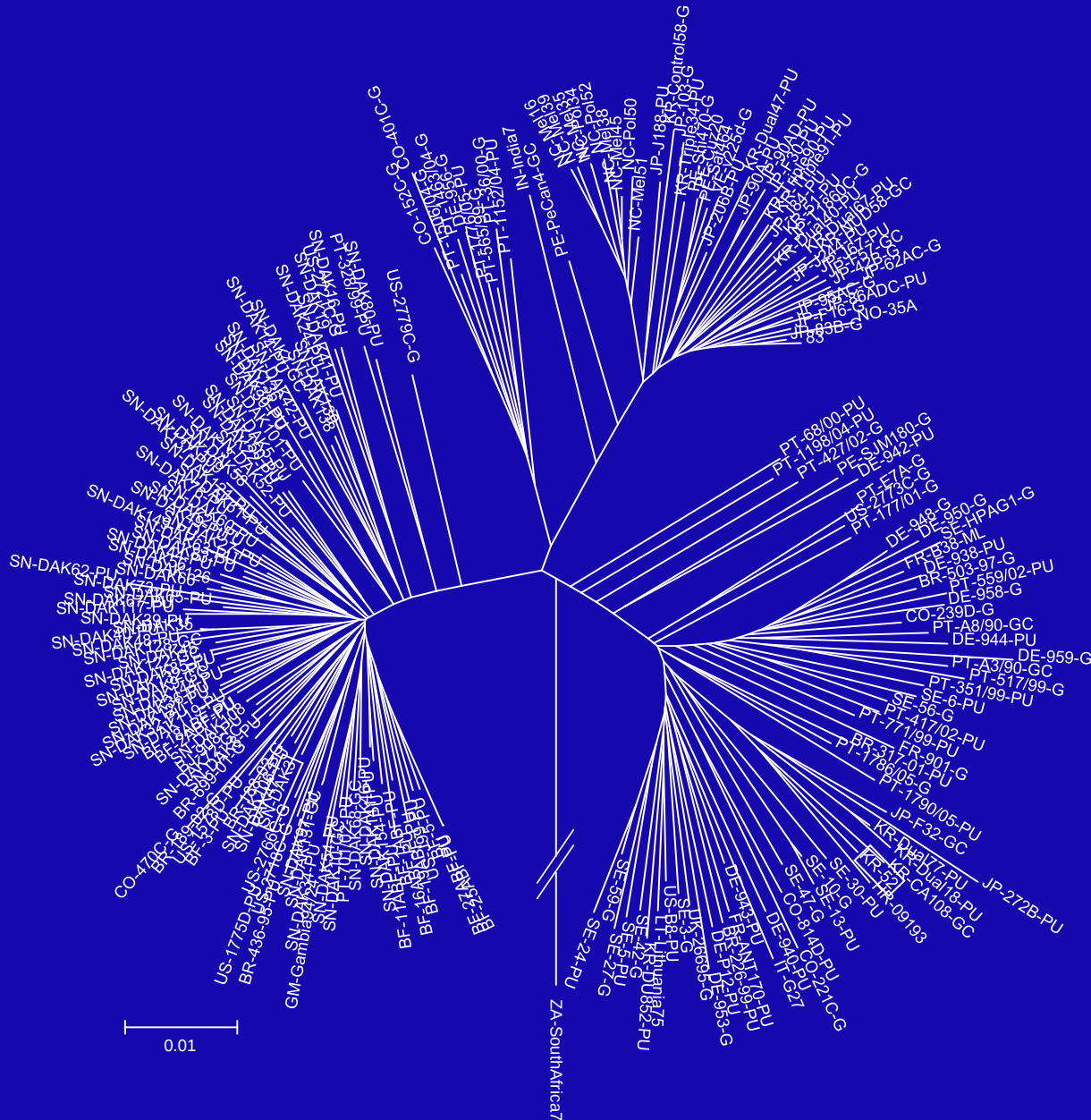
# Results - Corrected MEGA alignment

| DNA Sequences                                |            | Translated Protein Sequences |    |
|----------------------------------------------|------------|------------------------------|----|
| Species/Abbrv                                | Group Name |                              |    |
| 1. Otzi-iceman                               |            | ML                           | LA |
| 2. ZA-SA164AC/SA165AC/SA171AC/SA172A/SA302AC |            | ML                           | LA |
| 3. ZA-SA222AC                                |            | ML                           | LA |
| 4. ZA-SA252AC                                |            | ML                           | LA |
| 5. ZA-SA224AC                                |            | ML                           | LA |
| 6. ZA-SA29C/SA168AC                          |            | ML                           | LA |
| 7. ZA-SA226AC                                |            | ML                           | LA |
| 8. ZA-SA301AC                                |            | ML                           | LV |
| 9. ZA-SA161AC/SA214AC/SA215C/SA216AC         |            | ML                           | LA |
| 10. ZA-SA30AC                                |            | ML                           | LA |
| 11. ZA-SA31C/SA158AC/SA163AC/210AC/300AC     |            | ML                           | LA |
| 12. ZA-SA29A                                 |            | ML                           | LA |
| 13. ZA-SA194AC                               |            | ML                           | LA |
| 14. ZA-SA213AC                               |            | ML                           | LA |
| 15. ZA-SA45AC                                |            | ML                           | LA |
| 16. ZA-SA46AC                                |            | ML                           | LA |
| 17. ZA-SA47AC                                |            | ML                           | LA |
| 18. ZA-SA169C                                |            | ML                           | LA |
| 19. ZA-SA170AC                               |            | ML                           | LV |
| 20. ZA-SA173AC                               |            | ML                           | LA |
| 21. ZA-SA303C                                |            | ML                           | LA |
| 22. ZA-SA166A                                |            | ML                           | LA |
| 23. ZA-SA172C                                |            | ML                           | LA |
| 24. ZA-SA160A                                |            | ML                           | LA |
| 25. ZA-SA160C                                |            | ML                           | LA |
| 26. ZA-SA162AC                               |            | ML                           | LA |
| 27. ZA-SA253AC                               |            | ML                           | LA |
| 28. ZA-SA227AC                               |            | ML                           | LA |
| 29. ZA-SA251AC                               |            | ML                           | LA |
| 30. ZA-SA144AC                               |            | ML                           | LA |
| 31. ZA-SA155AC                               |            | ML                           | LA |
| 32. ZA-SA144AC (2)                           |            | ML                           | LA |
| 33. ZA-SA146AC                               |            | ML                           | LA |
| 34. ZA-SA156AC                               |            | VL                           | LA |
| 35. ZA-SA157AC                               |            | ML                           | LV |
| 36. ZA-SA35AC                                |            | ML                           | LA |
| 37. ZA-SA36C                                 |            | ML                           | LA |
| 38. ZA-SA34A                                 |            | ML                           | LA |
| 39. ZA-SA34C                                 |            | ML                           | LA |
| 40. ZA-SA40AC                                |            | ML                           | LA |
| 41. ZA-SA37AC                                |            | ML                           | LA |
| 42. ZA-SA221AC                               |            | ML                           | LA |
| 43. ZA-SA220AC                               |            | ML                           | LA |

## Results - Phylogenetic tree of *homD*

## 211 *homD* sequences

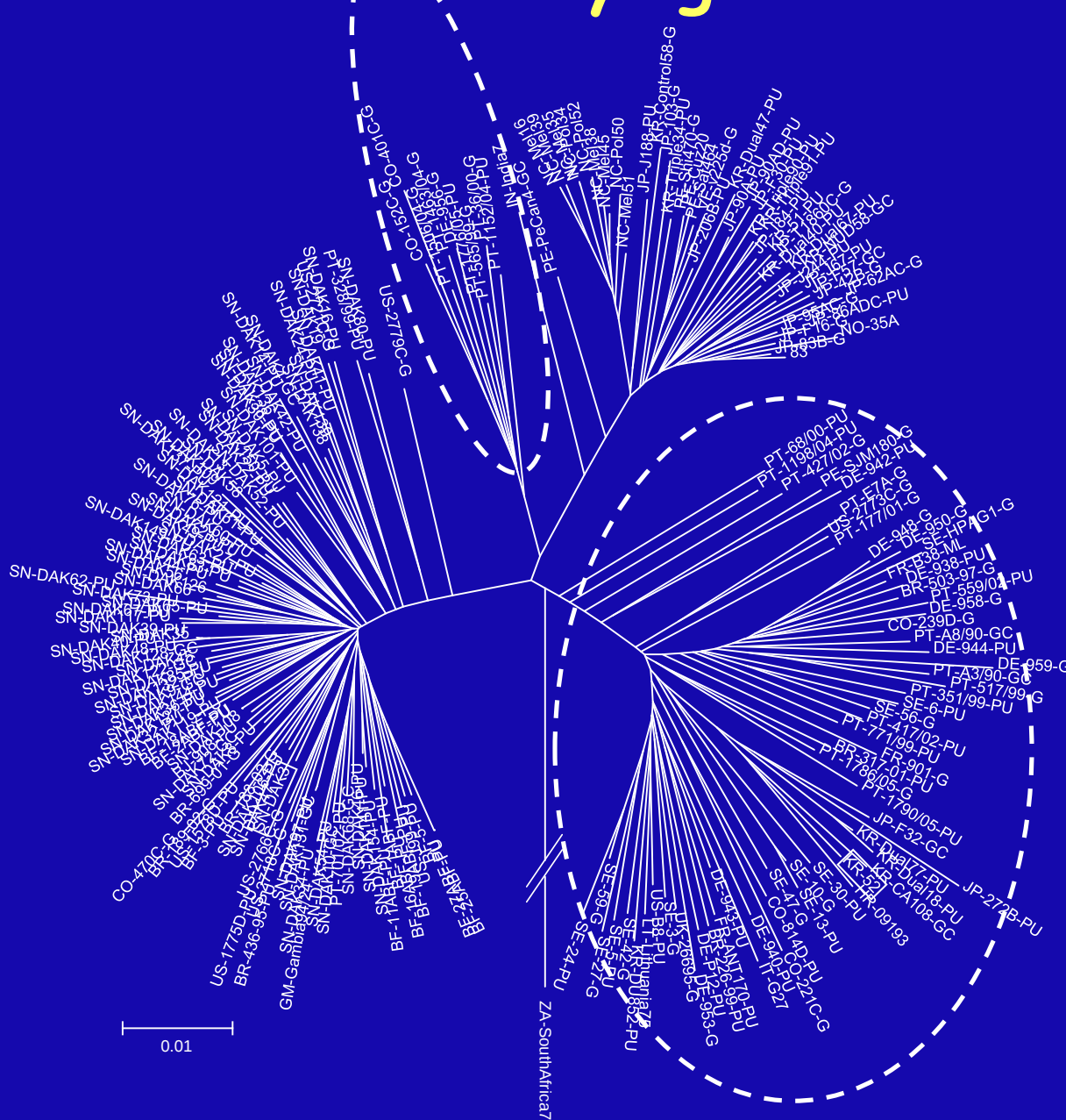
Neighbour-Joining  
Kimura 2-parameter model  
Bootstrap 1000 replicates



## Results – Phylogenetic tree of *homD*

## 211 *homD* sequences

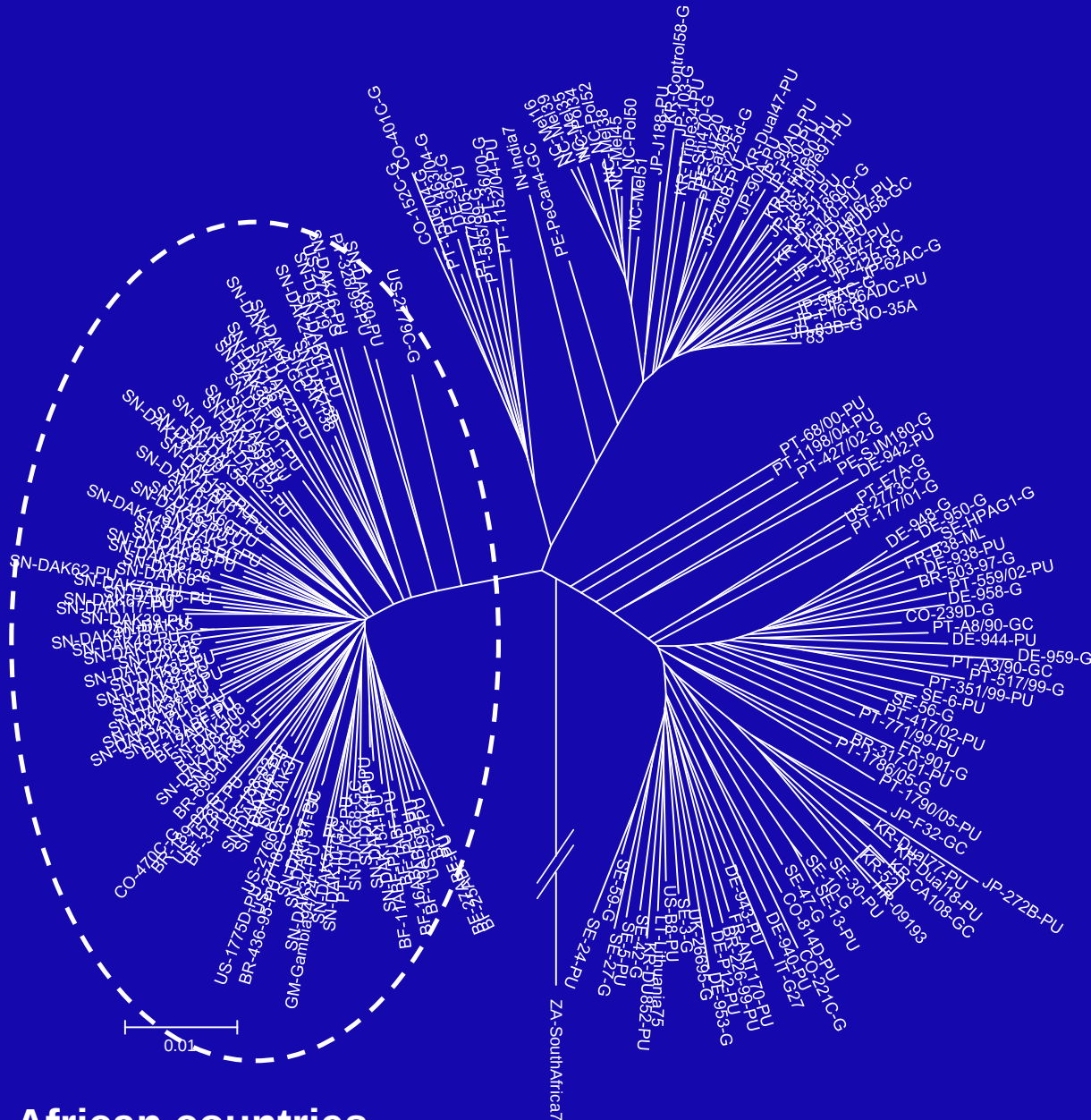
Neighbour-Joining  
Kimura 2-parameter model  
Bootstrap 1000 replicates



## Results - Phylogenetic tree of *homD*

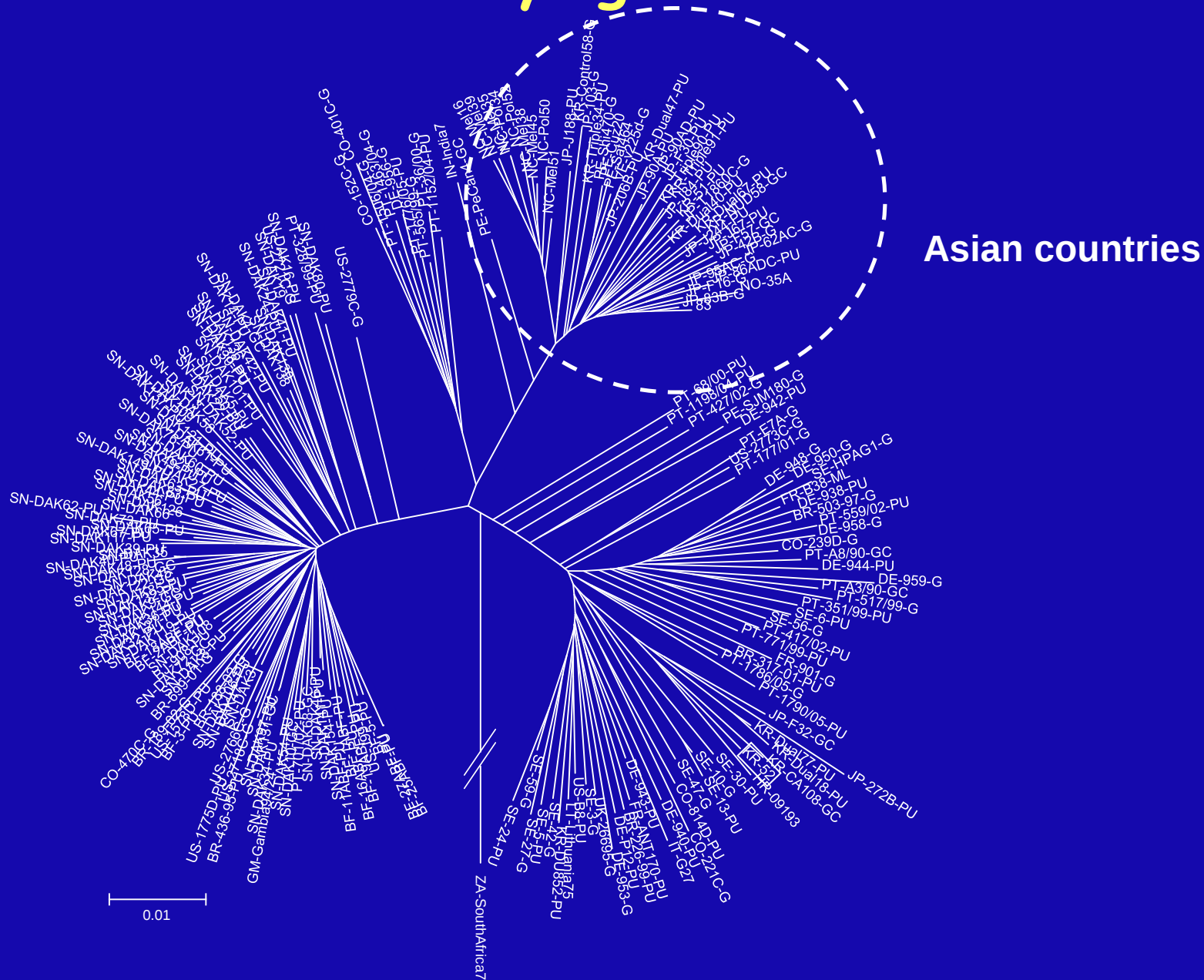
## 211 *homD* sequences

Neighbour-Joining  
Kimura 2-parameter model  
Bootstrap 1000 replicates



## African countries

## Results - Phylogenetic tree of *homD*

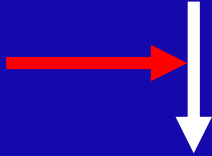


# Results - Phylogenetic tree of *homD*

Human stomach



*H. pylori*



Chronic gastritis

Adenocarcinoma



Ulcers

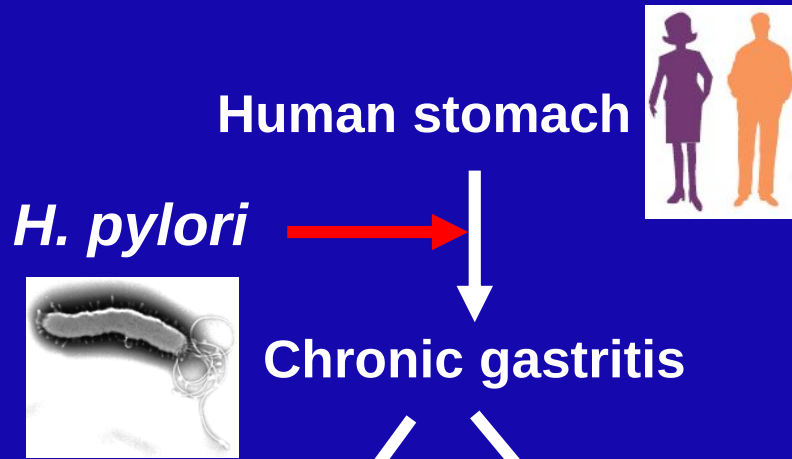


Duodenal



Gastric

# Results - Phylogenetic tree of *homD*



Adenocarcinoma



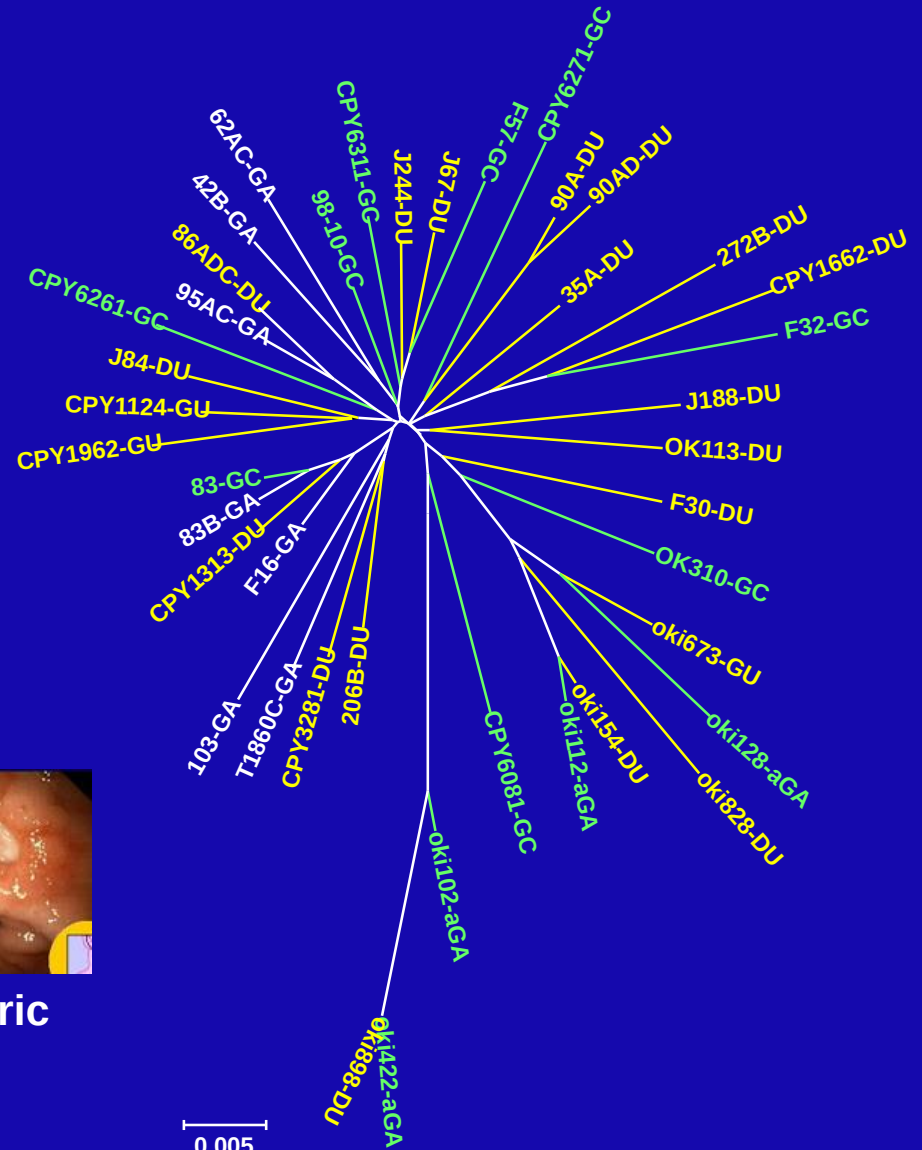
Ulcers



Duodenal



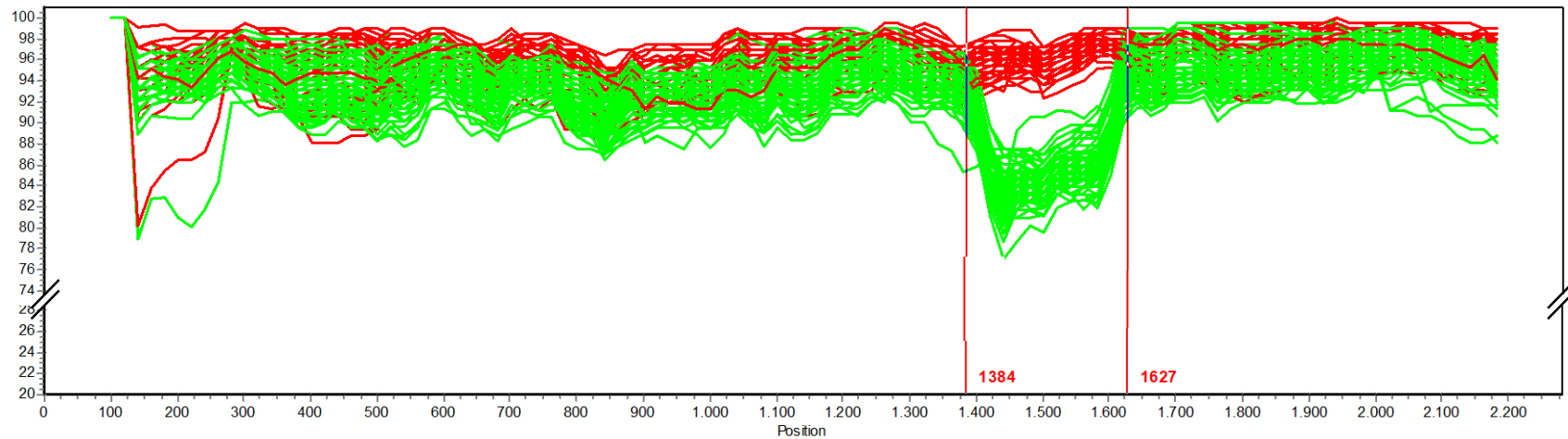
Gastric



Japanese strains

# Results - Similarity plot analysis of *homD*

211 *homD* sequences

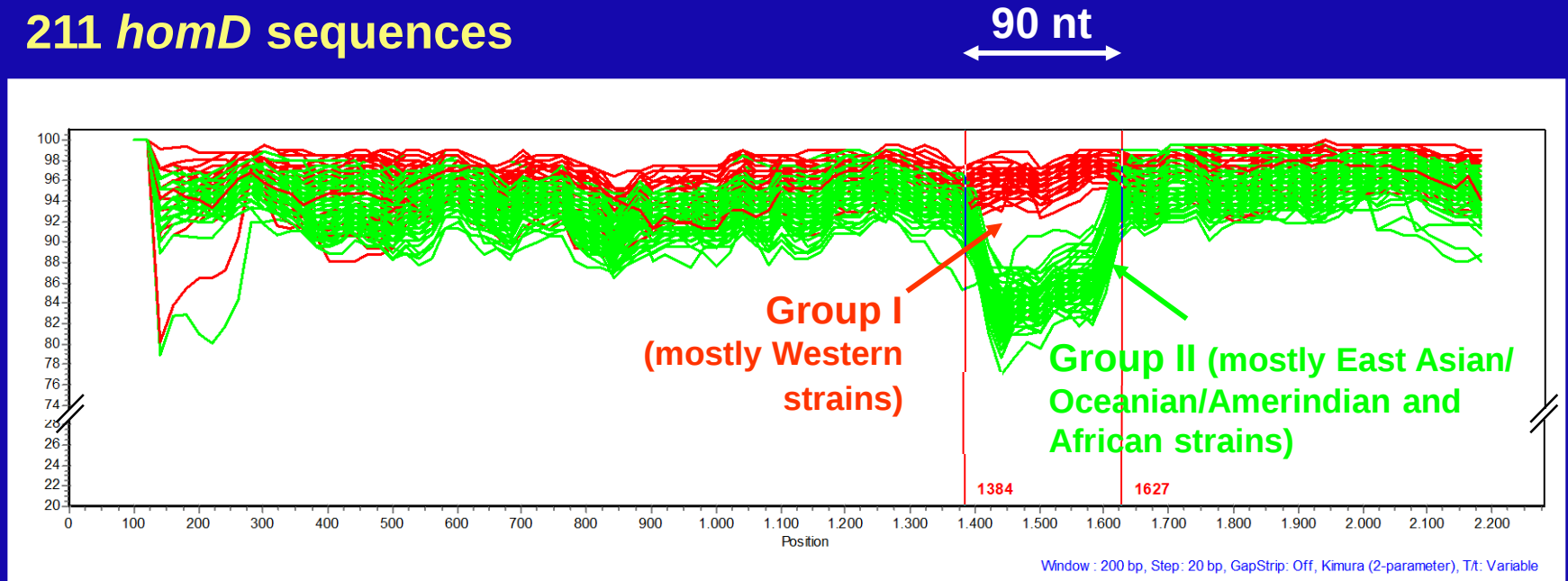


Window : 200 bp, Step : 20 bp, GapStrip: Off, Kimura (2-parameter), T/t: Variable

Kimura 2-parameter - 200-bp window - 20-bp step without GapStrip

# Results - Similarity plot analysis of *homD*

211 *homD* sequences

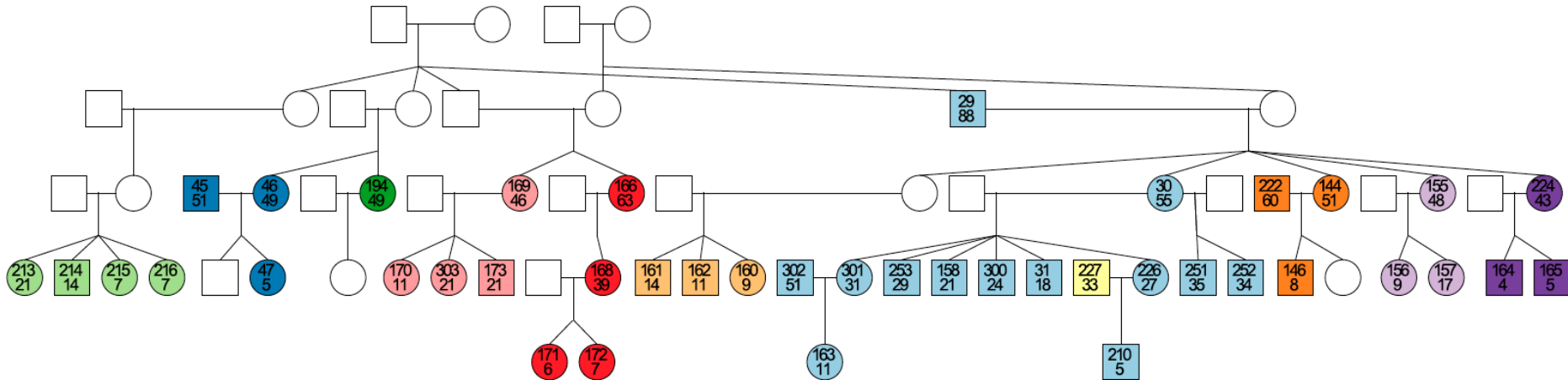


Kimura 2-parameter - 200-bp window - 20-bp step without GapStrip

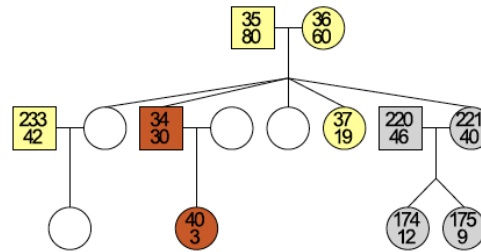
⇒ 90-nt region differing between Western strains and the other strains

# Results - HomD evolution in African families

## Family 1



## Family 2



- Increased KP repeats for a single strain between fundus and antrum
- Prediction of a single crossing over between the strains (3' end)

# Results - *homD* adaptation/evolution

- Transmission of *H. pylori* infection in couples (Australia)
  - Animal infection with *H. pylori*
    - mice
    - gerbil
    - cat
    - monkey
  - Induction of antibiotic resistance (26695 strain)
    - Metronidazole
    - Clarithromycin
    - Rifampicin
- ⇒ High HomD protein conservation

# Aims

## To investigate:

- ⇒ the diversity and phylogeny of *homD* on a large panel of *H. pylori* strains (patients from different geographical origins with different gastric diseases)
- ⇒ the expression and antigenicity of HomD

# Materials & methods

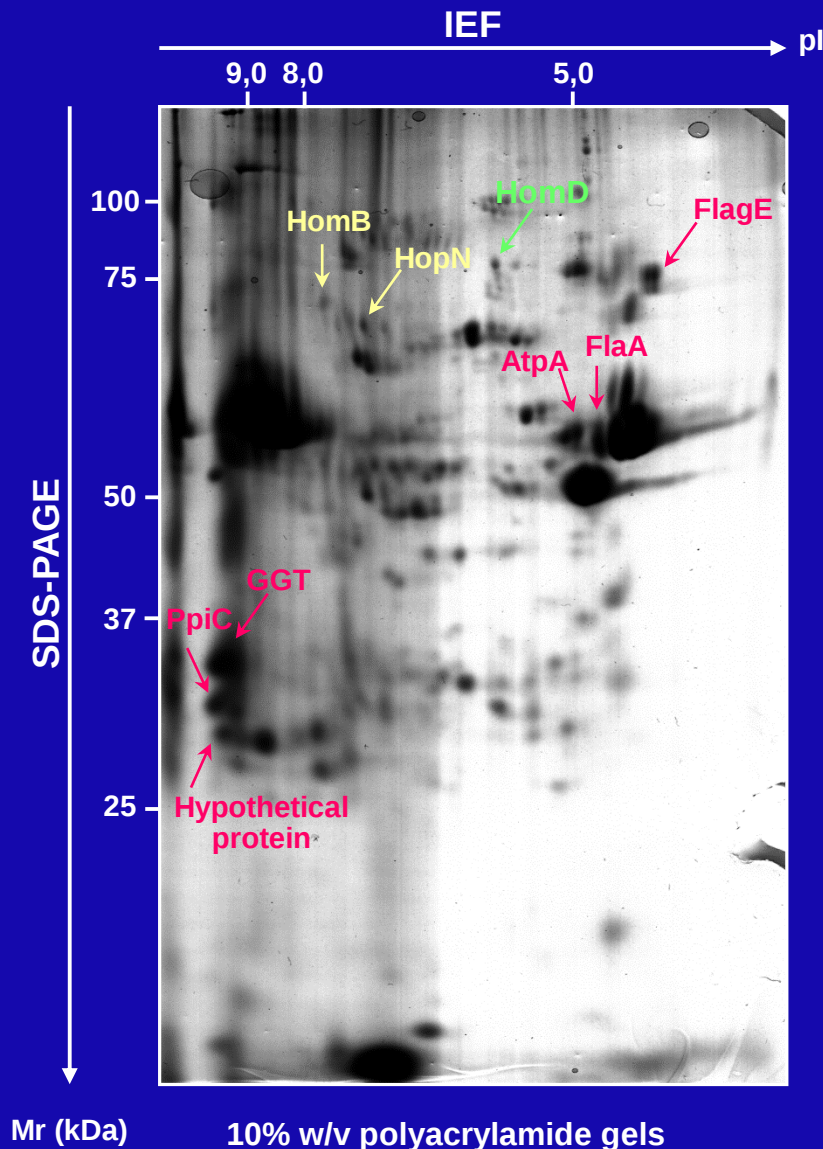
## Analysis of *H. pylori* membrane

- Two-dimensional gel electrophoresis of proteins from sarcosyl insoluble membrane preparation
- Protein peptide analysis by liquid chromatography tandem mass spectrometry (LC/MS/MS)

## Production of a recombinant HomD protein

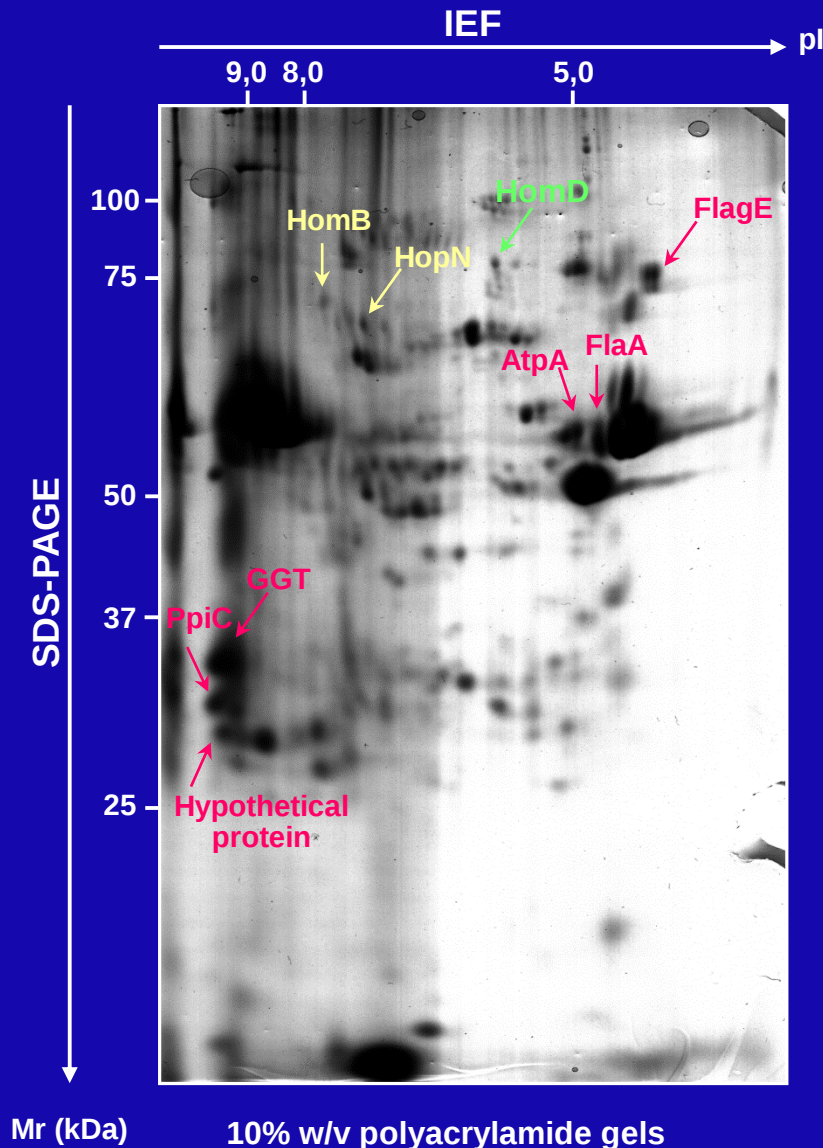
- *homD* from *H. pylori* strain 26695 (HP1453)
- Fusion to glutathion s-transferase (GST) in vector pGEX-4T-3 (GE Healthcare)
- Production in *Escherichia coli* (strain BL21)
- Western blot analysis of the crude extract (*E. coli* BL21)

# Analysis of *H. pylori* membrane protein



- Two-dimensional gel electrophoresis of proteins (IEF & SDS-PAGE)
- Coomassie blue staining
- Protein peptide analysis by LC/MS/MS

# Analysis of *H. pylori* membrane protein



**HomD (HP1453/JHP1346)**

- 2 peptides: 2.55 (2+) and 1.89 (2+)
- coverage = 3.1%

- Two-dimensional gel electrophoresis of proteins (IEF & SDS-PAGE)
- Coomassie blue staining
- Protein peptide analysis by LC/MS/MS

# Materials & methods

## Analysis of *H. pylori* membrane

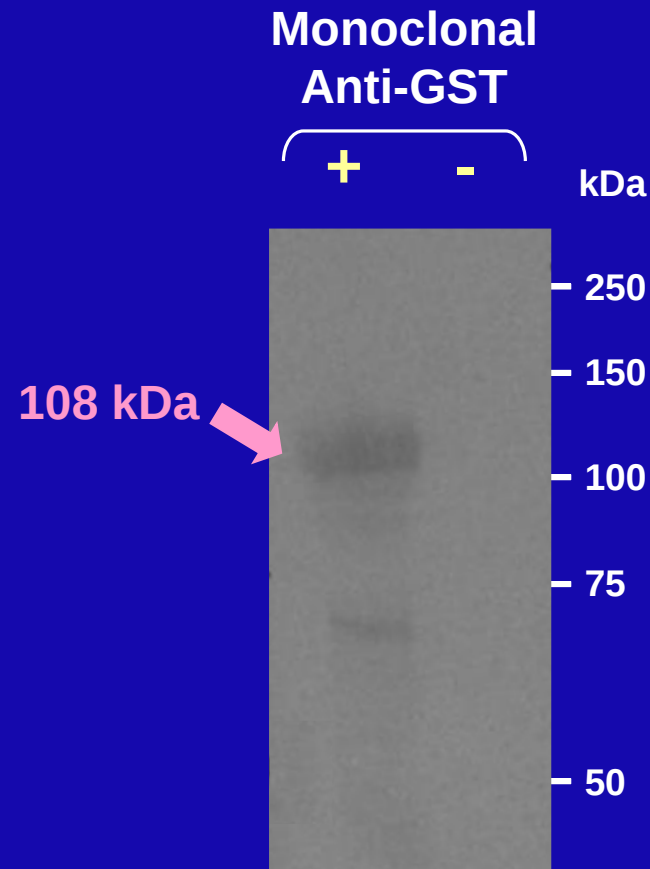
- Two-dimensional gel electrophoresis of proteins from sarcosyl insoluble membrane preparation
- Protein peptide analysis by liquid chromatography tandem mass spectrometry (LC/MS/MS)

## Production of a recombinant HomD protein

- *homD* from *H. pylori* strain 26695 (HP1453)
- Fusion to glutathion s-transferase (GST) in vector pGEX-4T-3 (GE Healthcare)
- Production in *Escherichia coli* (strain BL21)
- Western blot analysis of the crude extract (*E. coli* BL21)

# Analysis of recombinant GST-HomD

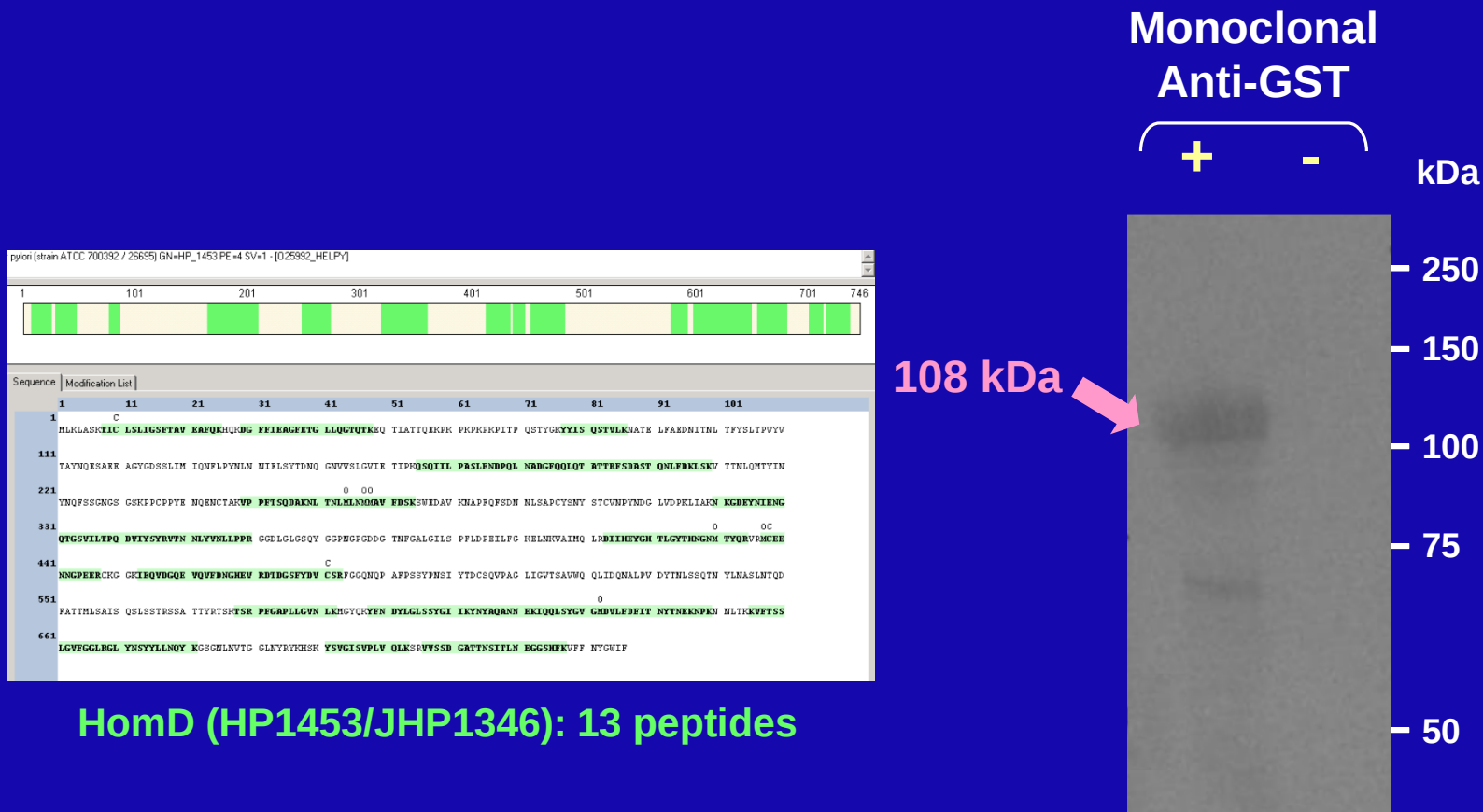
- Western blot analysis with monoclonal anti-GST antibody



⇒ Presence of a 108 kDa expected GST-tagged protein in the IPTG induced cells

# Analysis of recombinant GST-HomD

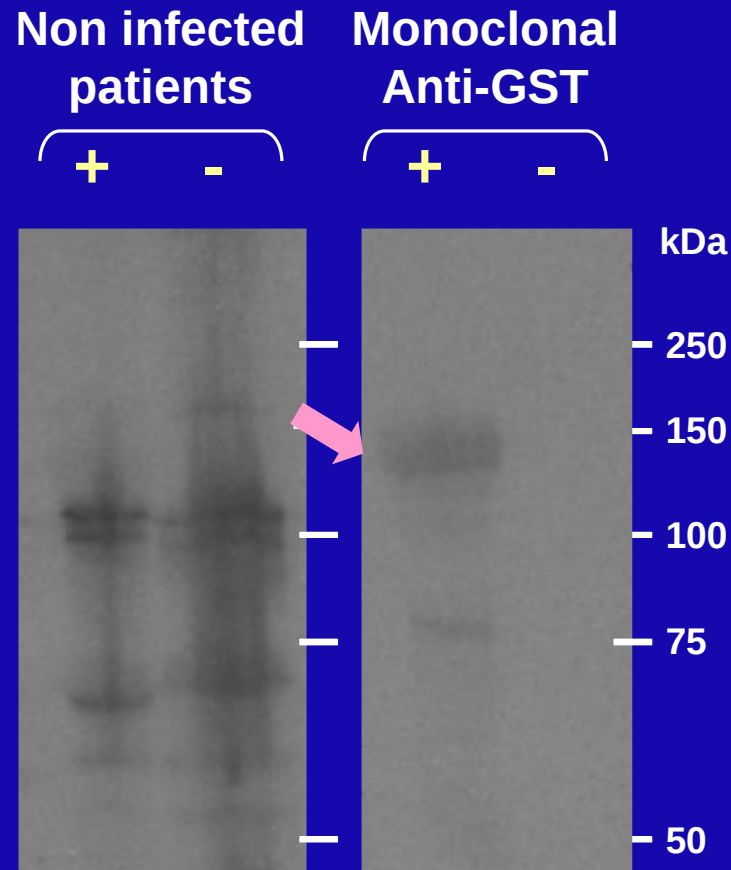
- Western blot analysis with monoclonal anti-GST antibody



⇒ Presence of a 108 kDa expected HomD-GST-tagged protein in the IPTG induced cells

# Analysis of recombinant GST-HomD

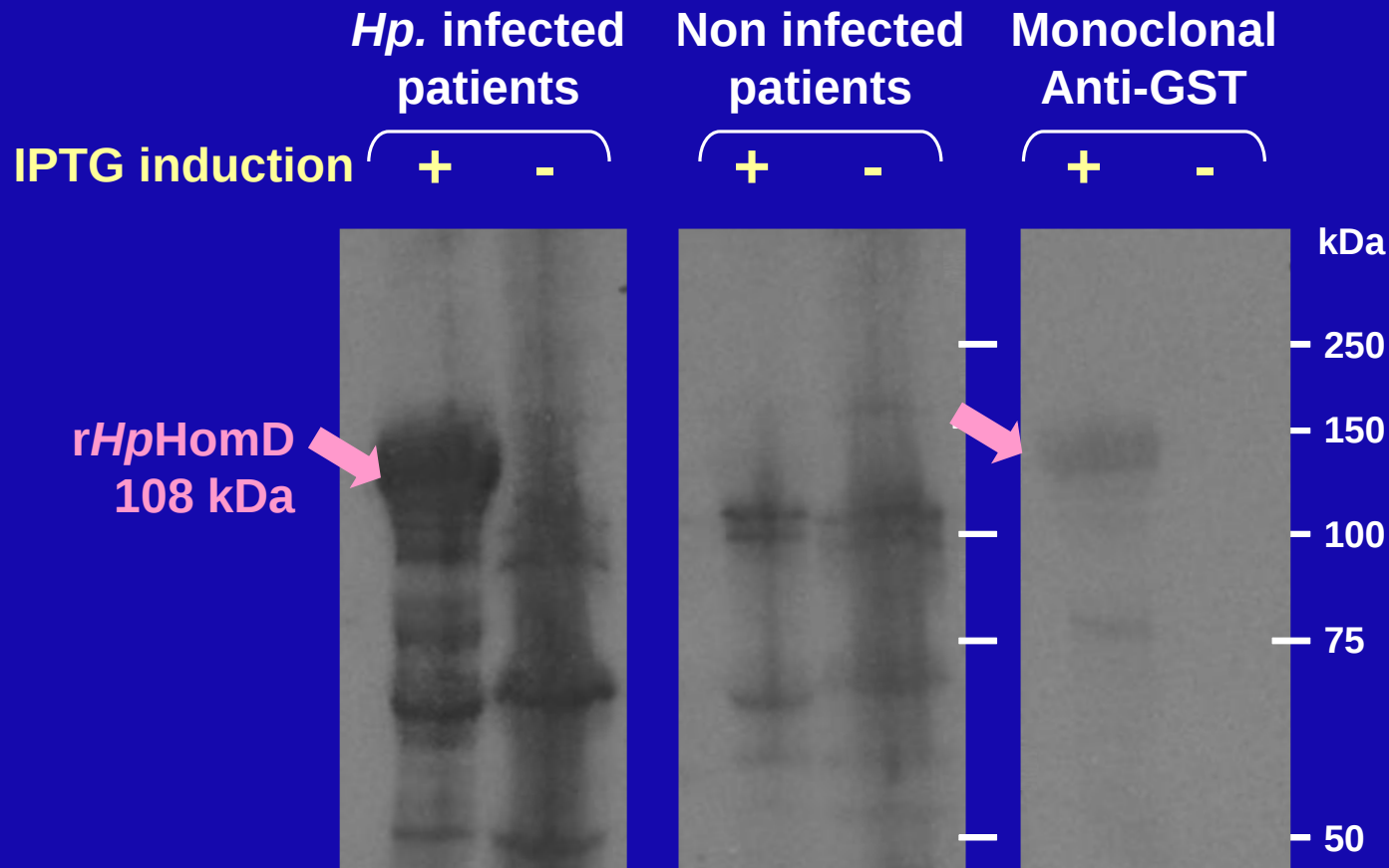
- Western blot analysis with monoclonal anti-GST antibody



⇒ Absence of the 108 kDa HomD-GST-tagged protein

# Analysis of recombinant GST-HomD

- Western blot analysis with monoclonal anti-GST antibody



- ⇒ Presence of the expected GST-HomD protein
- ⇒ HomD protein is antigenic in humans

✓ Construction of several *homD* knock-out strains



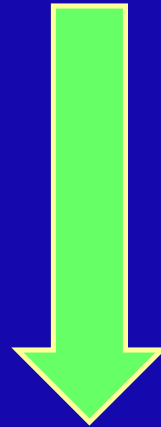
- ✓ **Analysis : wild type versus *homD* knock-out strains**
  - ⇒ **No effect on Hp adherence**
  - ⇒ **No effect IL-8 secretion**

# Conclusions

- ✓ The *H. pylori* OMP coding gene *homD* displays a high worldwide conservation, in contrast to the high polymorphism observed in other *H. pylori* OMPs
- ✓ *homD* does not seem to be “on/off” regulated
- ✓ The KP-rich repeat motif seems to be highly antigenic. This motif is rare among bacterial proteins
- ✓ The HomD protein is predicted to be exposed to the bacterial surface
- ✓ HomD is present in the membrane of *H. pylori* and is able to induce an immune response in humans

# Conclusions

**HomD fulfils important criteria for a vaccine candidate, such as antigenicity and high conservation among clinical isolates**



**NEW VACCINE TARGET**