

# Diversity and phylogeny of the *Helicobacter pylori* outer membrane protein-encoding gene *homC*



Instituto Nacional de Saúde  
Doutor Ricardo Jorge



Rita Cordeiro<sup>1</sup>, Armelle Ménard<sup>2,3</sup>, Stéphane Breurec<sup>4</sup>, Francis Mégraud<sup>2,3</sup>, Mónica Oleastro<sup>1</sup>

<sup>1</sup>Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal, <sup>2</sup>INSERM U853, 33076 Bordeaux, France, <sup>3</sup>Université Bordeaux 2 Victor Segalen, Laboratoire de Bactériologie, 33076 Bordeaux, France, <sup>4</sup>Institut Pasteur, Unité de Bactériologie Médicale et Environnementale, Dakar, Sénégal



## Introduction

*Helicobacter pylori* is a gram-negative gastric pathogen possessing a large set of outer membrane proteins (OMPs), which mediate important pathogen-host interactions. The *homC* gene codes for a *H. pylori* OMP and belongs to the *hom* family, together with the recently studied *homB* and *homA* genes. *homB* is implicated in bacterial adherence and in IL-8 activation. No specific function of *homC* is known yet.

## Aim

This work aims to study the genetic diversity and evolution of the *homC* gene, in a large panel of clinical and reference *H. pylori* strains, isolated from patients from different geographical origins and presenting different gastric diseases.

## Materials and Methods

### Bacterial strains:

- 26 *homC* sequences from *Hp* complete genome (NCBI).
- 182 *Hp* clinical strains isolated from patients presenting different gastric disease were used in the analysis:
  - 81 from Western countries (Portugal: 28, France: 2, Sweden: 11, Germany: 12, USA: 14, Colombia: 6, Brazil: 8) presenting non-ulcer dyspepsia and gastritis-G (n=45), peptic ulcer-U (n=33) or gastric cancer-GC (n=3);
  - 53 from East Asian countries (Japan: 27 and South Korea: 26) presenting non-ulcer dyspepsia and gastritis-G (n=25), peptic ulcer-PU (n=27) or gastric cancer-GC (n=1);
  - 48 from African countries (Burkina Faso: 8 and Senegal 40) presenting peptic ulcer-PU (n=36) or gastric cancer-GC (n=4); unknown (n=8).

### Sequence Analysis and Phylogeny of *homC*:

- The complete sequences of each gene were obtained by PCR and sequencing;
- Bioinformatic analysis was based on similarity plots and phylogenetic trees obtained with *SimPlot* Version 3.5.1 and *MEGA* (Molecular Evolutionary Genetics Analysis) 4.1 software, respectively, using the DNA sequence alignments generated by the *BioEdit Sequence Alignment Editor* (Version 7.0.1).

## Results

- All but one strain harboured a complete *homC* gene at a conserved locus.
- Phylogenetic reconstruction of *homC* revealed a geographic segregation, with three predominant clusters (Fig. 1): **Western cluster** (*hpEurope*), comprising strains from Europe and most of the strains from Columbia, USA and Brazil; **Asian/Ameridian cluster** (*hpEAsia*), including strains from Korea, Japan (*subtype hspEAsia*) and from Peru and Venezuela (*subtype hspAmerind*), and **African cluster** (*subtype hspWAfrica*) mostly comprised of strains from Burkina Faso, Senegal, Gambia and a few strains from Portugal, France, USA and Brazil.

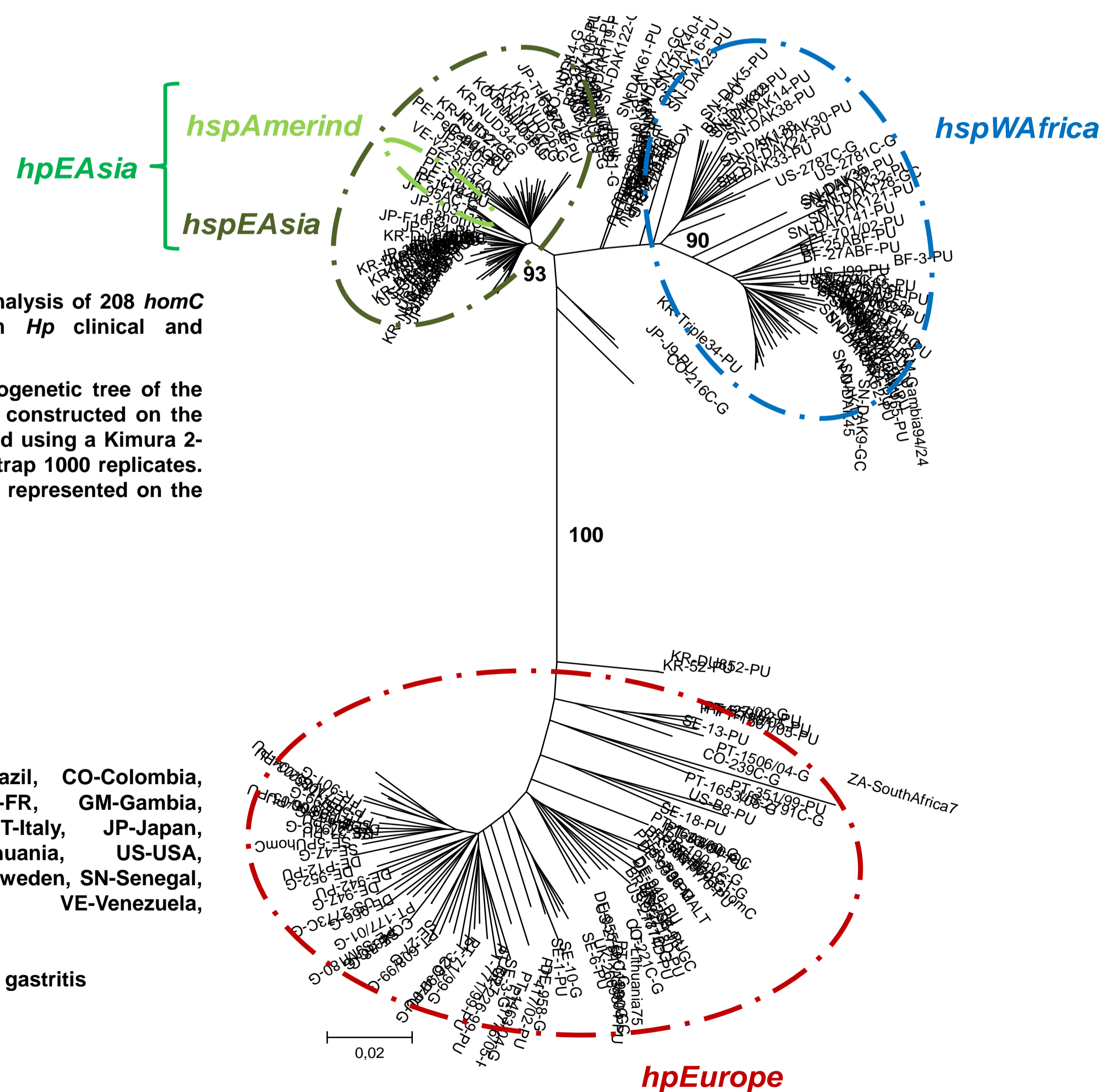


FIGURE 1: Phylogenetic analysis of 208 *homC* sequences obtained from *Hp* clinical and reference strains.

The Neighbor-Joining phylogenetic tree of the nucleotide alignments was constructed on the basis of distances estimated using a Kimura 2-parameter model and bootstrap 1000 replicates. The branch length index is represented on the tree.

### Abbreviations used:

BF-Burkina Faso, BR-Brazil, CO-Colombia, DE-Germany, FR-France-FR, GM-Gambia, HR-Croatia, IN-India, IT-Italy, JP-Japan, KR-SouthKorea, LT-Lithuania, US-USA, PE-Peru, PT-Portugal, SE-Sweden, SN-Senegal, UK-UnitedKingdom, VE-Venezuela, ZA-SouthAfrica.

G: non-ulcer dyspepsia and gastritis  
PU: peptic ulcer  
GC: gastric cancer

## Results

- A similarity plot analysis suggests a conserved profile of gene segmentation, where three segments were defined (FIGURE 2):

- segment 1 (5' end extremity):** sequences are separated according to the geographical origin of the strain in two groups: **East Asian/Ameridian and African group** (level of similarity ~40%);
- segment 2 (middle region):** highly polymorphic region (level of similarity ~40%), in which 8 allelic variants were identified (AI-AVIII);
- segment 3 (3' end extremity):** more conserved region (level of similarity ~90%).

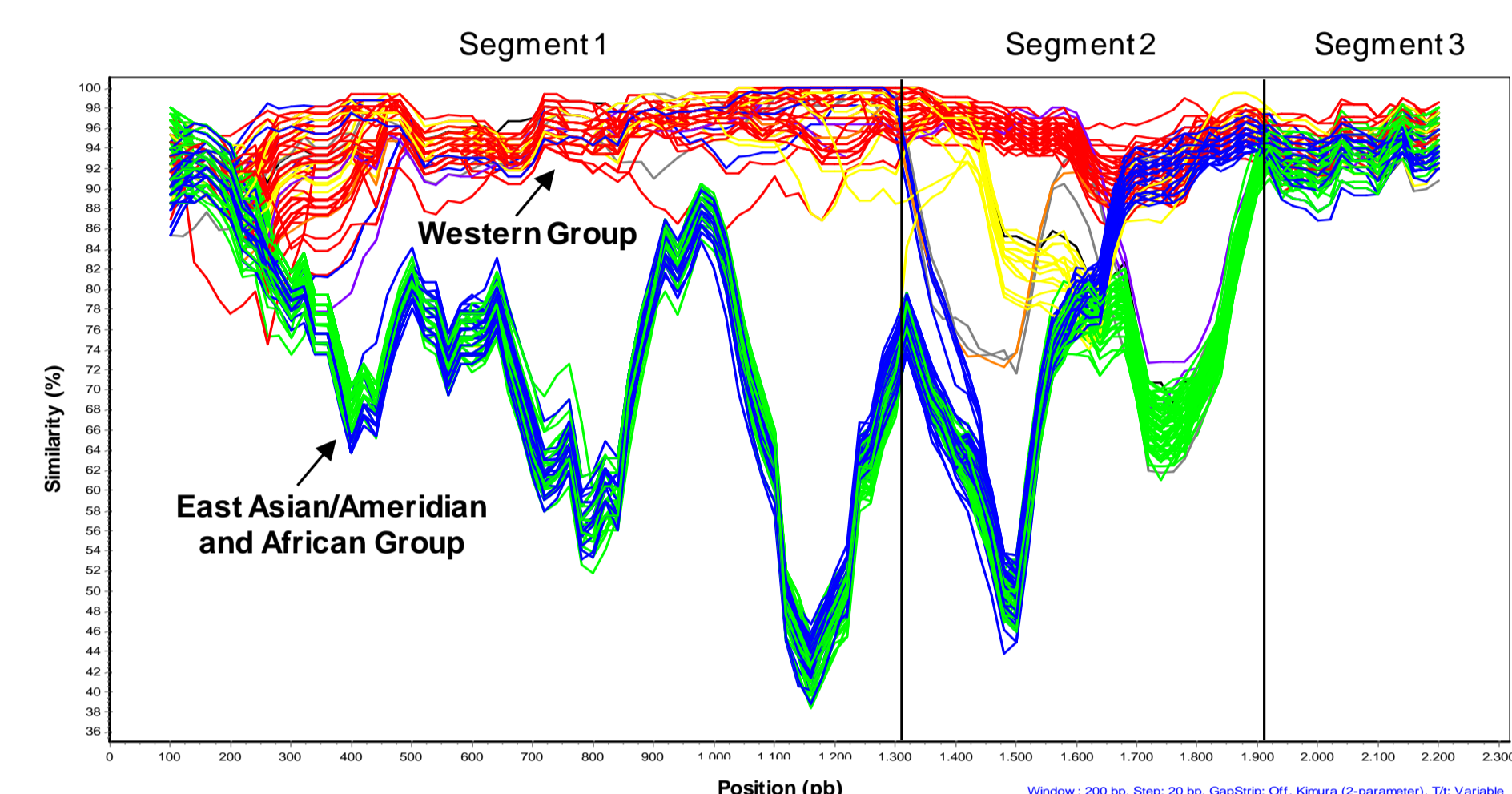


FIGURE 2: Similarity plot analysis of 208 *homC* sequences representing the eight allelic variants (AI-red, AII-blue, AIII-yellow, AIV-green, AV-purple, AVI-grey, AVII-orange and AVIII-black). The plot was generated with the Kimura 2-parameter, a 200-pb window, a 20-bp step without GapStrip and *homC* sequence from a Western strain as reference.

## Results

The eight allelic variants identified in segment 2 presented different frequencies among the strains tested, and geographic specificity regarding the most prevalent ones was observed (FIGURE 3): allele **AI** predominant (66,2%) and exclusive in **Western group**; allele **AIV** was predominant (94,1%) in **East Asian/Ameridian** strains and was not observed in Western strains; the **AII** allele was predominant (66,7%) in **African** strains and was the only allele present in the three geographical groups.

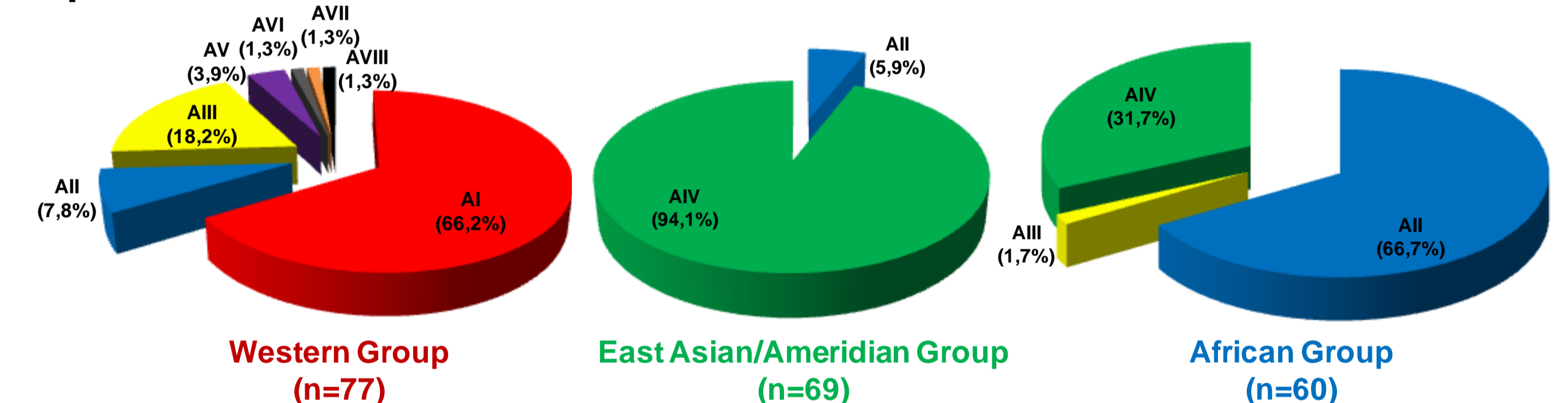


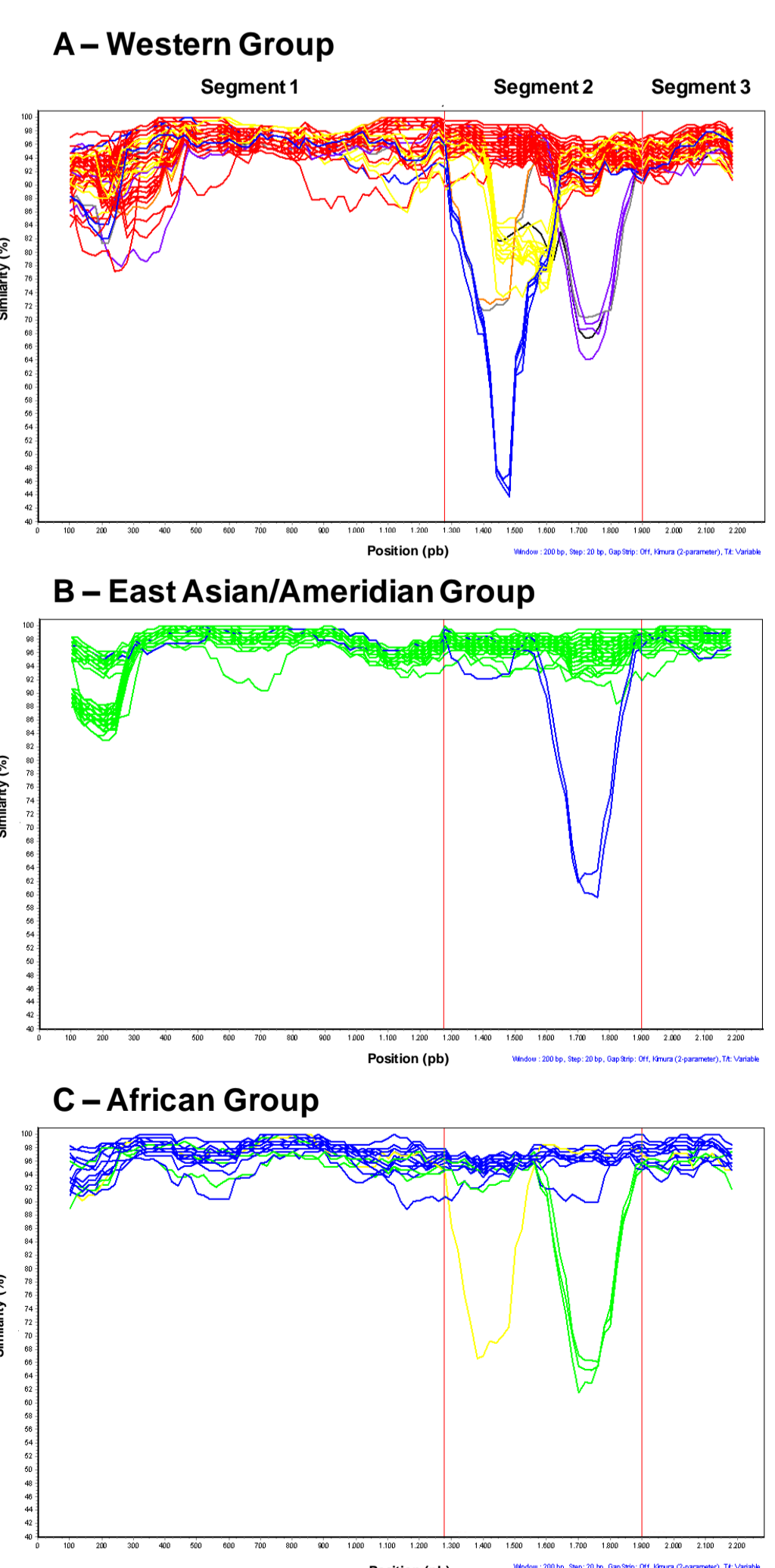
FIGURE 3: Distribution of the eight *homC* alleles in Western, East Asian/Ameridian and African strains.

The similarity plots by geographical region (FIGURE 4) show that the **African-predominant** allele (**AII**) was the most distant from the other two allelic variants predominant in **Western strains** (**AI**) and in **East Asian/Ameridian** (**AIV**).

The evidence that the segment 2 is the most polymorphic among strains from the same geographic region (FIGURE 4) but the most conserved within each allele (not shown) strongly supports its choice as the allelic region.

The **AII** allele was strongly associated with peptic ulcer disease ( $p=0.037$ ). Moreover, a more virulent genotype (*cagA+/vacAs1*) was associated with **AI** ( $p<0.01$ ) and **AIV** ( $p<0.001$ ) alleles.

FIGURE 4: Similarity plot analysis of *homC* sequences from the same geographic group. The plot was generated with the Kimura 2-parameter, a 200-pb window, a 20-bp step without GapStrip.



## Conclusions

Overall, the regular presence of *homC* and its allelic variability with geographic specificity suggest that *homC* is a host-interactive gene and a good candidate to be part of the pool of *H. pylori* OMPs involved in bacterial persistence. Moreover, the allelic variants may constitute biomarkers of the gastric disease and of the virulence of the strain.