



**HAL**  
open science

# Tracing Human Ancestral Migrations Using Symbiotic Bacteria

Alexia Nguyen Trung, Bastien Boussau, Vincent Daubin

► **To cite this version:**

Alexia Nguyen Trung, Bastien Boussau, Vincent Daubin. Tracing Human Ancestral Migrations Using Symbiotic Bacteria. Groupe des Méthodes Pluridisciplinaires Contribuant à l'Archéologie (GMPCA), May 2019, Montreal, Canada. hal-03068231

**HAL Id: hal-03068231**

**<https://hal.science/hal-03068231v1>**

Submitted on 16 Dec 2020

**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.

## Tracing Human Ancestral Migrations Using Symbiotic Bacteria

Alexia Nguyen Trung, Bastien Boussau and Vincent Daubin



Alexia Nguyen Trung is a PhD student at Université Claude Bernard Lyon 1 (alexia.nguyen-trung@univ-lyon1.fr)

The genetic heritage of *Homo sapiens* is much more substantial and complex than presented by 20th century genetics. We now know that each human is the place and ultimately the product of the interaction of a multitude of symbiotic organisms, each carrying its own genome. Some of these associations are stable over the life course, and in some cases even over generations. This is the case of *Helicobacter pylori*, a bacterium generally considered to be pathogenic to humans, which is present in the stomach of 50% of the current world population (Breurec, 2011). Its transmission occurs particularly during childhood within the family (Weyermann and al., 2006), limiting the spread of infection. As a result, each strain is confined to a distinct human group and *H. pylori* typing can allow identification of different ethnic groups within the same population, even when it comes to very recent migrations (Wirth and al., 2004). For that reason, the study of strains collected from different locations allows us to spatially trace human ancestral movements and its presence in humans can be traced to the origin of *Homo sapiens* in Africa before the great/major migrations. *H. pylori* has accompanied its host during its peregrinations on the different continents, to the point that the phylogeny of the current strains traces in broad lines the history of human migrations since the *Out of Africa* migration (Falush and al., 2003; Linz and al., 2007).

Like many bacterial species, *H. pylori* has the ability to acquire genetic material from other bacteria which are not parents, by a process called horizontal gene transfers (HGT). We have previously developed a method which allows the identification of ancient events of HGT in bacterial genomes. As a result, it is possible to identify donor and recipient lineages in a phylogeny. A corollary is that we can determine which lineages are contemporary in the phylogeny of species, because HGT supposes a community of time and space. More specifically, on a phylogenetic tree, an HGT event attests that the ancestors of a donor species are older than the descendants of the recipient species. We have developed MAXTIC, a new method to use this information to date a phylogenetic tree based on HGT (Chauve et al., 2017), and showed its power to infer the timing of diversification in the tree of life (Arellano Davin, 2017).

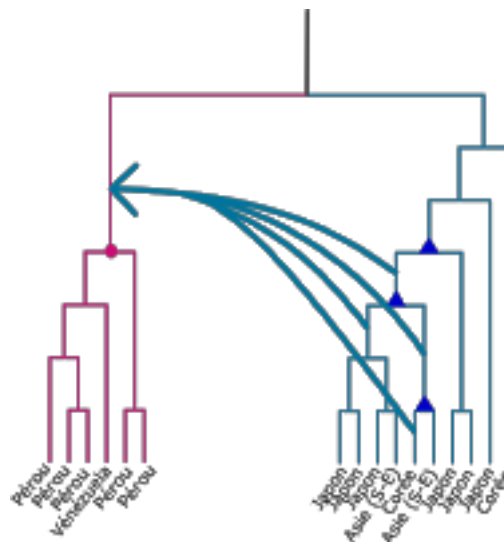
The objective of this work was to make a proof of principle that HGT in *Helicobacter pylori*, a symbiont of humans, could be used to refine dates in the history of human populations. To do this,

we retrieved a dataset of 35 strains for which a complete genome sequence was available and geographic information was confirmed (Yahara et al., 2013). We analyzed the complete genome of each of the 35 strains to optimize the chances of observing ancestral horizontal gene transfers. The goal is to compare the complete genomes of *H. pylori* strains to 1) infer as general phylogeny of the strains and 2) compare each gene history to this reference tree in order to identify all HGTs that occurred throughout this history.

More specifically:

- We clustered homologous genes together to study all gene families using MMseqs (Steinegger and Söding, 2017). We kept all families present in at least 3 genomes for further analysis. This left 1622 families for study.
- We built the tree of *H. pylori* strains based on the concatenate of those families present in all strains (502 ubiquitous genes).
- Using ALE, a probabilistic phylogenetic reconciliation method reconstructing HGT scenarios, we inferred the history of all 1622 gene families within the species tree. From this dataset, we detected 900 transfer events, which corresponded to 40 time constraints.
- Finally, based on this set of constraints, we were able to reconstruct a dated species tree for *H. pylori* using MAXTIC (Chauve and al., 2017), and assess the statistical support for the resulting chronogram. The results show that HGTs provide a strong signal which is in accordance with the current knowledge on the timing of human migrations.

Figure 3 shows an example of the type of information provided by HGT on the relative timing of diversification of *H. pylori* strains. In this example, ALE detects HGTs from the East-Asian clade (in blue) to the American clade (in fuchsia). According to these events, the ancestor of Asian donors is older than the descendants of the recipient lineage. This is in accordance with the history of the populations, because *Homo sapiens* first migrated to Asia before reaching the Americas.



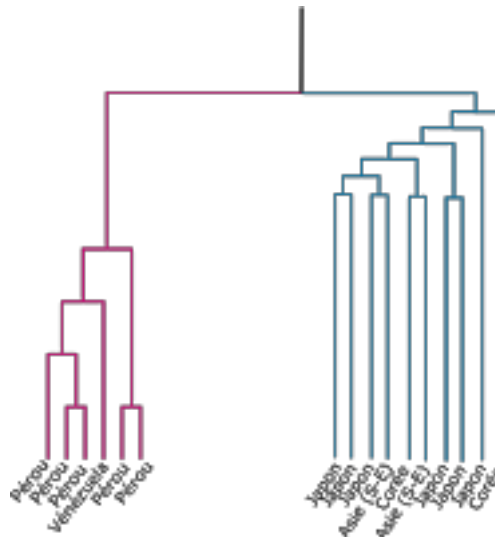


Figure 3. (left) ALE has detected gene transfers from East-Asian clade (in blue) to the American clade (in fuchsia). These transfers attest that the donor's ancestors (represented by a triangle) are older than the recipient's sons (represented by a circle); (right) From the constraints of fig a, our phylogenetic tree should then be represented in this way: the East-Asian ancestors are older than the American.

This study demonstrates that there is a strong potential in using *Homo sapiens* symbionts for dating human migrations. We show that HGT is abundant in *H. pylori* genomes, and that it provides strong statistical support for a phylogeographic sequence which corresponds to current knowledge on human migrations. There is little to be learned on such a small instance (only 35 strains of *H. pylori*), but it opens the door to a larger study using the >1000 genomes of *H. pylori* available in public databases. Interestingly, ancient *H. pylori* DNA has been recently extracted and sequenced from mummies. The inclusion of this data into our dataset will be of great interest as they provide direct absolute dates to calibrate our phylogenies. With this in mind, we have recently developed and implemented in RevBayes (Höhna et al., 2016) a new method to combine dating information from HGT with molecular clock methods to further improve dating.

## References:

- Breurec S., 2011. *Helicobacter pylori*: migrations humaines et cancer gastrique. Sciences agricoles. Université Paris Sud - Paris XI.
- Weyermann, M., Adler, G., Brenner, H., Rothenbacher, D., 2006. The mother as source of *Helicobacter pylori* infection. *Epidemiol. Camb. Mass* 17, 332–334.
- Wirth, T., Wang, X., Linz, B., Novick, R.P., Lum, J.K., Blaser, M., Morelli, G., Falush, D., Achtman, M., 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.
- Falush, D., Wirth, T., Linz, B., K Pritchard, J., Stephens, M., Kidd, M., J Blaser, M., Y Graham, D., Vacher, S., Perez Perez, G., Yamaoka, Y., Megraud, F., Otto, K., Reichard, U., Katzowitzsch, E., Wang, X., Achtman, M., Suerbaum, S., 2003. Traces of Human Migrations in *Helicobacter pylori* Populations. *Science* 299.
- 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.
- Arellano Davin, A., 2017. Utilisation des transferts horizontaux de gènes pour dater des phylogénies. Lyon.
- Cédric Chauve, Akbar Rafiey, Adrian Davin, Celine Scornavacca, Philippe Veber, et al. MaxTiC:Fast ranking of a phylogenetic tree by Maximum Time Consistency with lateral gene Transfers. 2017.hal-01532738v1
- Yahara, K., Furuta, Y., Oshima, K., Yoshida, M., Azuma, T., Hattori, M., Uchiyama, I., Kobayashi, I., 2013. Chromosome Painting. In: *Silico in a Bacterial Species Reveals Fine Population Structure*. MEB, Molecular Biology and Evolution 30, 1454–1464.
- Höhna, S., Landis, M.J., Heath, T.A., Boussau, B., Lartillot, N., Moore, B.R., Huelsenbeck, J.P., Ronquist, F., 2016. RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. *Syst. Biol.* 65, 726–736.
- Szöllösi Gergely J., Davin Adrián Arellano, Tannier Eric, Daubin Vincent and Boussau Bastien, 2015. Genome-scale phylogenetic analysis finds extensive gene transfer among fungi 370 *Phil. Trans. R. Soc. B*
- Steinberger, M., Söding, J., 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.*