Grand Valley State University
ScholarWorks@GVSU

Technical Library | School of Computing and Information Systems

2013

Analysis of Virulence Factors of Helicobacter pylori Infection and Human Migration

Yuka Kutsumi
Grand Valley State University

Follow this and additional works at: https://scholarworks.gvsu.edu/cistechlib

ScholarWorks Citation
Kutsumi, Yuka, "Analysis of Virulence Factors of Helicobacter pylori Infection and Human Migration" (2013). *Technical Library*. 159.
https://scholarworks.gvsu.edu/cistechlib/159

This Capstone is brought to you for free and open access by the School of Computing and Information Systems at ScholarWorks@GVSU. It has been accepted for inclusion in Technical Library by an authorized administrator of ScholarWorks@GVSU. For more information, please contact scholarworks@gvsu.edu.

Analysis of Virulence Factors of Helicobacter pylori Infection and Human Migration

By

Yuka Kutsumi

A paper submitted in partial fulfillment of the

Requirements for the degree of Master of Science in Medical and Bioinformatics at

Grand Valley State University.

Winter 2013

Acknowledgements

I would like to thank Dr. Guenter Tusch for always being supportive to this project. I would like to also thank GVSU Computer Information Systems department for giving me this opportunity to conduct this analysis.

Table of Contents

List of Figures

Figure Page

List of Tables

Table Page

Abstract

Helicobacter pylori is a ubiquitous bacterium found in the human stomach. Its infection is linked to various gastroduodenal diseases including peptic ulcer and gastric cancer. Despite the fact that H.pylori is present in more than two-third of the global population, only approximately 20% of infected individuals develop severe diseases. It is interesting to find out who these 20 % of the individuals are, and why they develop the diseases. One of the unique applications of H.pylori study is that it can be used as a tool for tracing human migrations since humans coevolved with their microbiomes. In this project, the diversity of bacteria and host in the H.pylori-human model was examined. Multilocus sequence typings from seven housekeeping genes in 393 H.pylori strains were analyzed using Bayesian Analysis of Population Structure software. This project gives an insight into the mechanism of co-evolution between the host and the parasites.

Introduction

Humans evolved with microbiomes, microorganisms that inhabit the human bodies. Viruses such as HPV, Hepatitis G, RNA retrovirus HTLV-1 and some bacteria have been proved to have signs of coevolution with their human host, reflecting the African-origin of the human migration history (Dominguez-Belllo *et al*, 2008). Helicobacter pylori is one of these microbiomes that grows in the mucus layer that coasts the inside of the human stomach. It is a gram-negative spiral-shaped bacterium, and its infection is linked to various gastroduodenal diseases including peptic ulcer and gastric cancer (Suerbaum and Michetti, 2002). In 1994, H.pylori infection was categorized as a group 1 carcinogen by the International Agency for Research on Cancer. Spread of H.pylori occurs through contaminated food and water or through direct mouth-to-mouth contact (Testerman T *et al*). In most populations, the bacterium is first acquired during childhood (National Cancer Institute). The Center for Disease Control and Prevention (CDC) estimates that this bacterium is present in approximately two-third of the global population with infection rates much higher in developing countries than in developed countries. Interestingly, only approximately 20% of those infected individuals develop severe diseases (National Cancer Institute). In addition, despite the high prevalence of H. pylori infection in Africa and South Asia, the incidence of gastric cancer in these areas is much lower than in other countries (Malaty, 2007). In fact, gastric cancer is second most common cancer in Asia, and more than half of the world's cancer cases arise in Eastern Asia (National Cancer Institute). It has been reported that the incidence of gastric cancer tends to decrease from north to south in East Asia (Suzuki et al, 2012). Thus, it seems the geographical location plays an important role for explaining the virulence of H.pylori.

Virulence Factor

There are a few genes considered to be H.pylori virulence factors. VacA is the second most extensively studied H.pylori virulence factor after cagA (Suzuki et al, 2012). Unlike cagA, all H.pylori strains have a functional vacA, which encodes a vacuolating cytotoxin. VacA can induce vacuolation, multiple cellular activities, including membrane-channel formation, cytochrome x release from mitochondria leading to apoptosis (Suzuki et al, 2012). The difference in vacA structure at the signal region (s1 and s2) and middle region (m1 and m2) are

discovered to have impacts on the variations in the vacuolating activity of different H.pylori strains. In vitro experiments have shown that s1m1 strains are the most toxic, followed by s1m2 strains, whereas s2m2 strains have no cytotoxic activity, and the s2m1 strains are rare (Atherton et al, 1995). Many studies in Western countries showed that individuals infected with vacA s1 or m1 H.pylori strains have an increased risk of peptic ulcer or gastric cancer compared with those with s2 or m2 strains (Atherton *et al*, 1995).

This project examines the possible reason for the varying H.pylori virulence by looking at the differences in the virulence factor (vacA) of H.pylori. By showing the geographic differences in the pathology, I hope to articulate the essential need for studying diverse ethnic groups to ensure that more people benefit from the medical study.

Materials and Method

Multilocus Sequence Typing:

Multilocus Sequence Typing (MLST) is a procedure for characterizing isolates of bacterial species using the sequences of internal fragments of seven housekeeping genes. It utilizes housekeeping genes because they are subject to purifying selection and slow evolution, and especially the variation within these genes is nearly neutral (Margo *et al*, 2008). Despite their fewer polymorphic sites in individual housekeeping genes compared with hypervariable genes, using the combined sequences of multiple housekeeping genes has been shown to provide high discriminatory power (Margo *et al*, 2008).

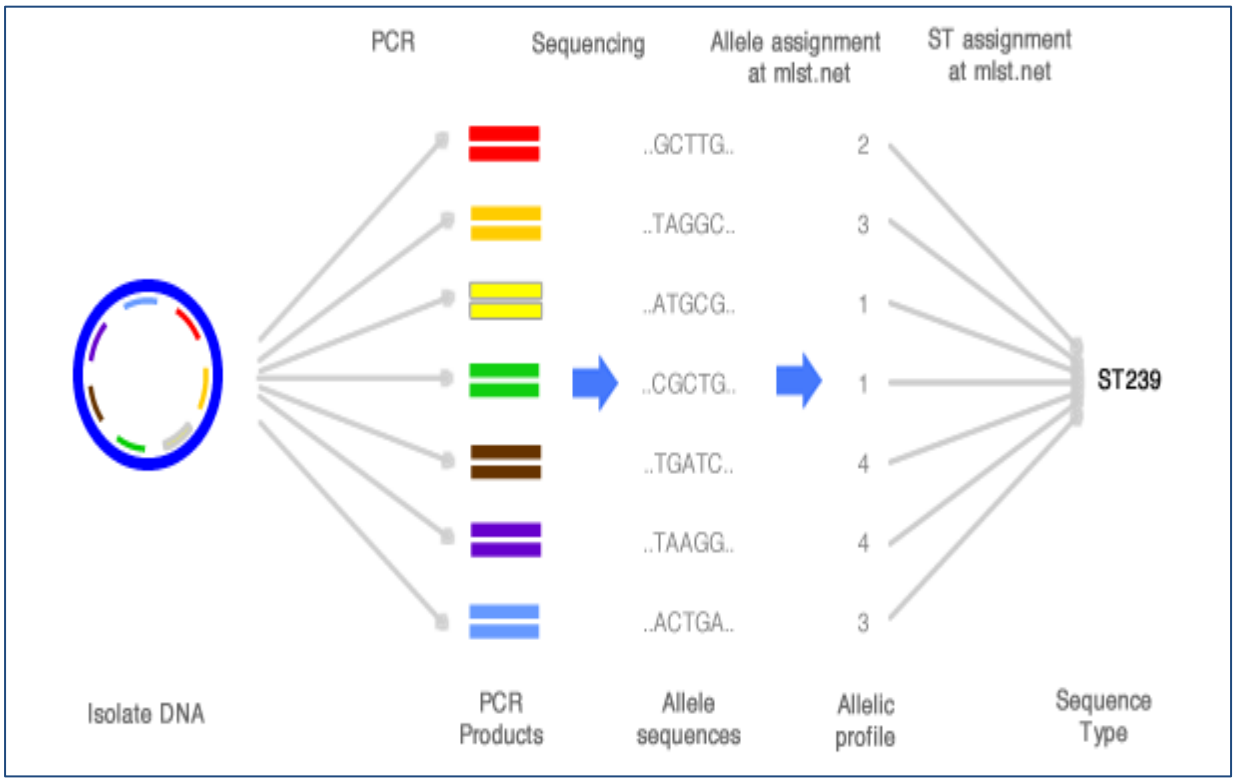


Figure 1. MLST Schematics
MLST is a genotyping method that involves the comparisons of ~450bp long nucleotide sequences from seven housekeeping genes. First, the isolated DNAs are amplified by using PCR, and then the allele sequences are compared among the seven loci. For each housekeeping gene, the different sequences present within a bacterial species are assigned as distinct allelic profile. By comparing all the allelic profiles, unique sequence typing (ST) are assigned.

Characterizing Samples

The data used to define populations and evaluate the robustness of the population assignments were downloaded from the H.pylori MLST database (http://pubmlst.org/helicobacter/). The database consisted of 15901 DNA sequences, 2456 MLST profiles, and 1551 isolates(STs). In this study, total of 393 individuals were chosen from 18 countries; Australia, New Zealand, Singapore, Korea, China, India, Estonia, Finland Germany, Spain, UK, Italy, USA, Costa Rica, Columbia, Burkina Faso, Senegal, and South Africa (Table 1). Missing allelic profiles were calculated by Python code (appendix 1).

Table I. List of sampling countries and number of individuals. Total of 393 individuals MLST data were chosen from 18 different countries. The largest sample data were collected from South Africa (N=84), whereas Senegal had the smallest number of individuals (N=5).

Country	#of individuals
Australia	21
NewZealand	25
Singapore	11
Korea	11
China	13
India	17
Estonia	11
Finland	9
Germany	22
Spain	37
UK	30
Italy	6
USA	30
Costa Rica	21
Columbia	28
Burkina Faso	12
Senegal	5
South Africa	84
Total	393

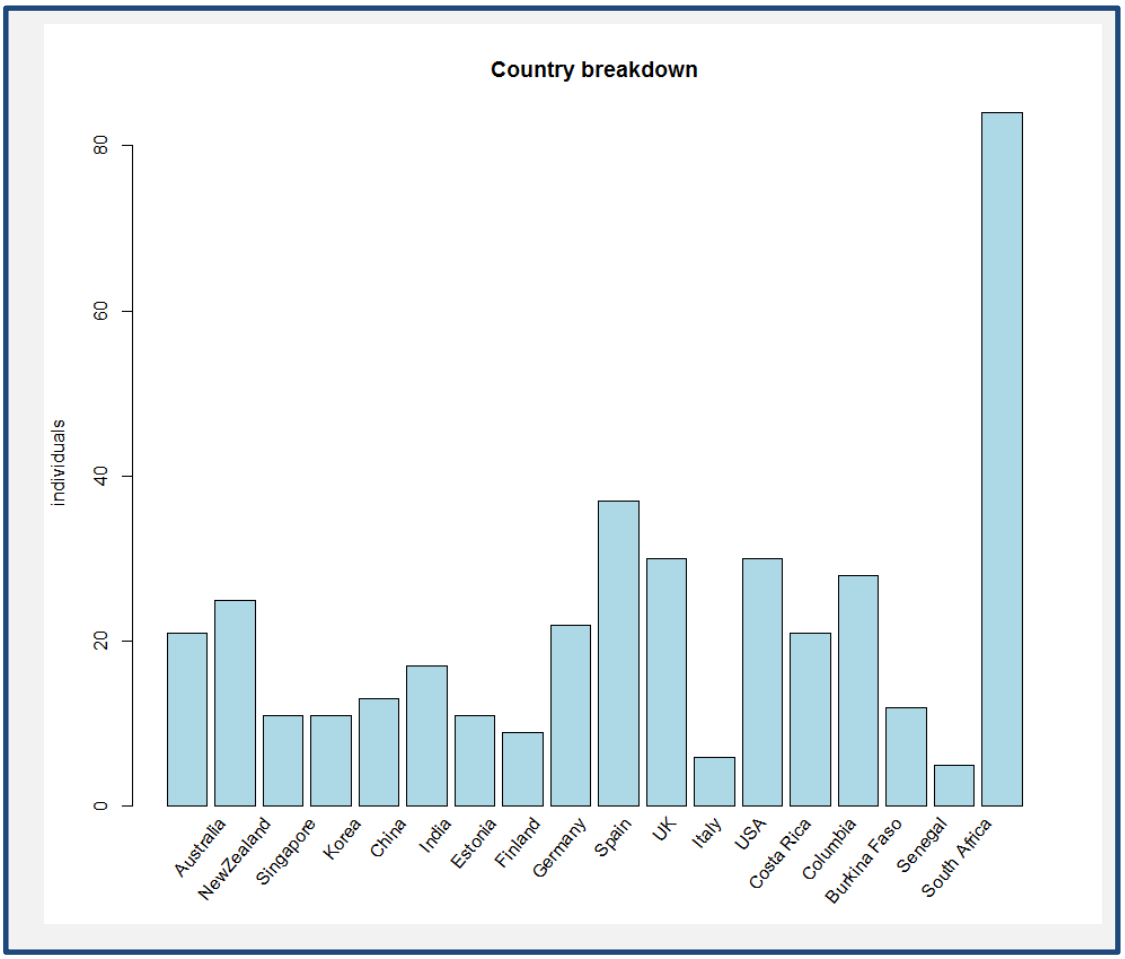


Figure 2. Sampling country distribution.
This barplot shows the number of individuals collected from each country. South Africa had largest sampling populations and Senegal had least number of individuals participated in this project. R statistical software package was used for generating this plot.

Population Analyses using BAPS

Bayesian Analysis of Population Structure (BAPS) was employed in order to assess population assignments of H.pylori using MLST data. BAPS (version 6) is a free software package for Bayesian inference of genetic structure within a given dataset (Corander J *et al*). BAPS treats both the allele frequencies of the molecular markers and the number of genetically diverged groups in population as random variables. In order to identify the genetic diversity of H.pylori, two clustering analyses were conducted. At first, the vacA gene was examined using the "clustering of individuals" module. Then, housekeeping genes (atpA, efp, mutY, ppa, trpC, ureI, phC) were examined by using the same module with linked loci option. In this module, BAPS

determines the log likelihood in 10% increments of different population divisions and subsequently calculates the most likely K values (Corander J *et al*). The likelihood of population assignment for each ST is also calculated by BAPS.

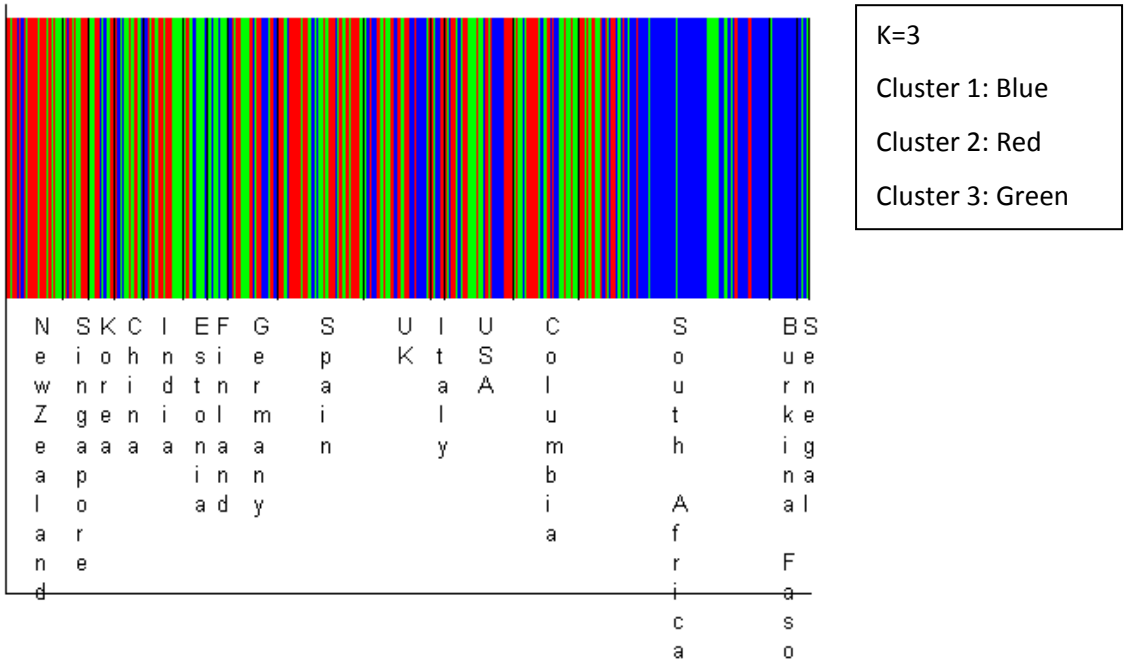
Results

Population Assignments of H.pylori based on vacA

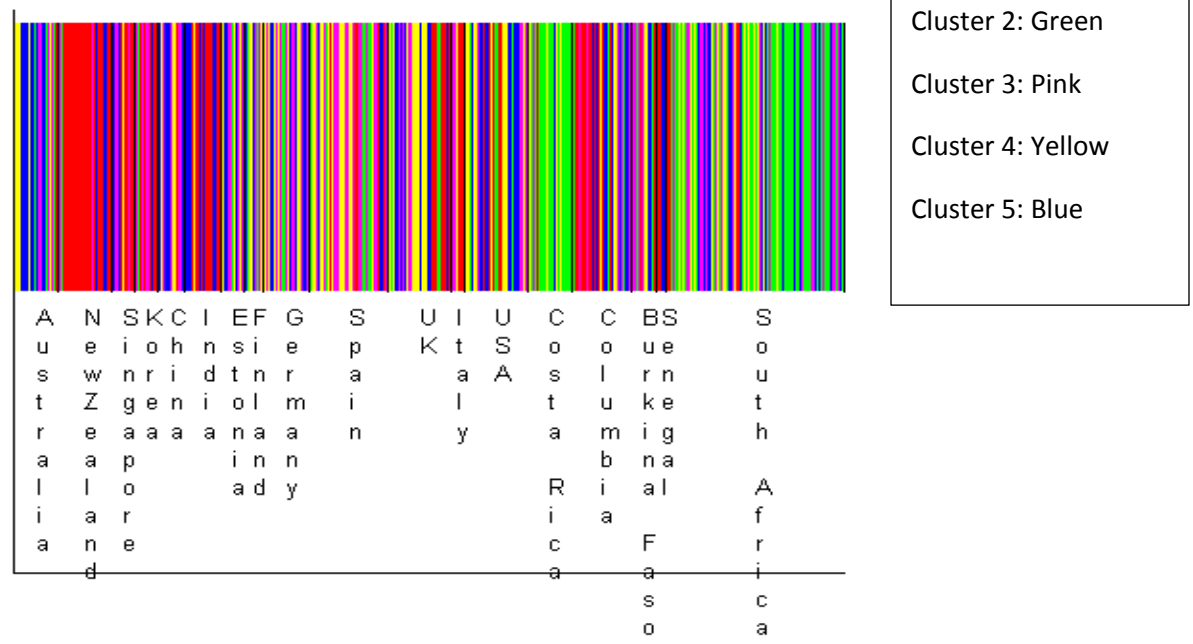
Genetic clusters of vacA were calculated by BAPS without using prior sampling information (ethnic information). Due to high missing value rate in Australia and Costa Rica, these two countries were removed from the population structure analysis. The estimated number of population was K=3. Using the calculated K value, BAPS generated a mixture clustering graphical output (Fig 3-a). Each cluster was assigned a unique color in the graphic; however the color ordering was arbitrary that we couldn't compare the colors between the analyses. Each sampling unit, in this case the individuals, is represented by a vertical bar having the color corresponding to the cluster where it was placed. It was observed that there were not many individuals who belong to cluster1 in Asian countries, whereas the majority of the individuals in African countries belonged to cluster 1.

Population Assignments of H.pylori based on seven housekeeping genes

Genetic clusters of seven housekeeping genes were also calculated to see the diversity of H.pylori among the sampling countries. The estimated number of population was K=5. Using the calculated K value, BAPS generated a mixture clustering graphical output (Fig 3-b). There are high heterogeneity observed in each country expect Eastern Asian countries (China and Korean), and New Zealand. China and Korea showed two colors, indicating the individuals sampled from there only belong to two clusters. New Zealand showed a single color for all individuals, reflecting the homogeneous genetic makeups. The individuals from South Africa showed all colors, indicating they have the same allelic frequency as the individuals from other countries sampled in this study. Also most of the individuals from African Countries showed that they belong to cluster 2 or cluster 4 (shown green and yellow), whereas the no individuals from Asian countries and Oceania countries belonged cluster 4.



a



b

Figure 3. Estimated population assignments of H.pylori genotypes based on vacA sequence typing data (a) and seven housekeeping sequence typing data (b). BAPS clustered the vacA sequence typings into three groups (a). Based on the vacA output, the majority of the individuals in African countries such as South Africa, Burkina Faso, and

Senegal were clustered as cluster1 which was colored in blue. However, not many individuals from Asia were found in cluster1 (a). In the housekeeping genes' output, five different clusters were assigned (b). The data showed high heterogeneity except New Zealand (b). African countries and Costa Rica showed many individuals belonging to cluster 2 and 4 (green and yellow), whereas no individuals from Asian countries and Oceania countries showed cluster 4, and a very few individuals belonged to cluster2.

Constructing Phylogenic Tree using housekeeping sequence typing

In order to explore the relationships among the identified clusters, a phylogenic tree was constructed by using the MLST data. A simple hierarchical clustering using UPGMA(Unweighted Pair Group Method with Arithmetic Mean) method based on Neighbor's distances (averaged over loci) showed the evolutionary relationship among the five populations (Fig 5). Five populations were then determined as South Asians, East Asians, Oceanians, Africans and Caucasians by comparing the individual ethnic information in each cluster.

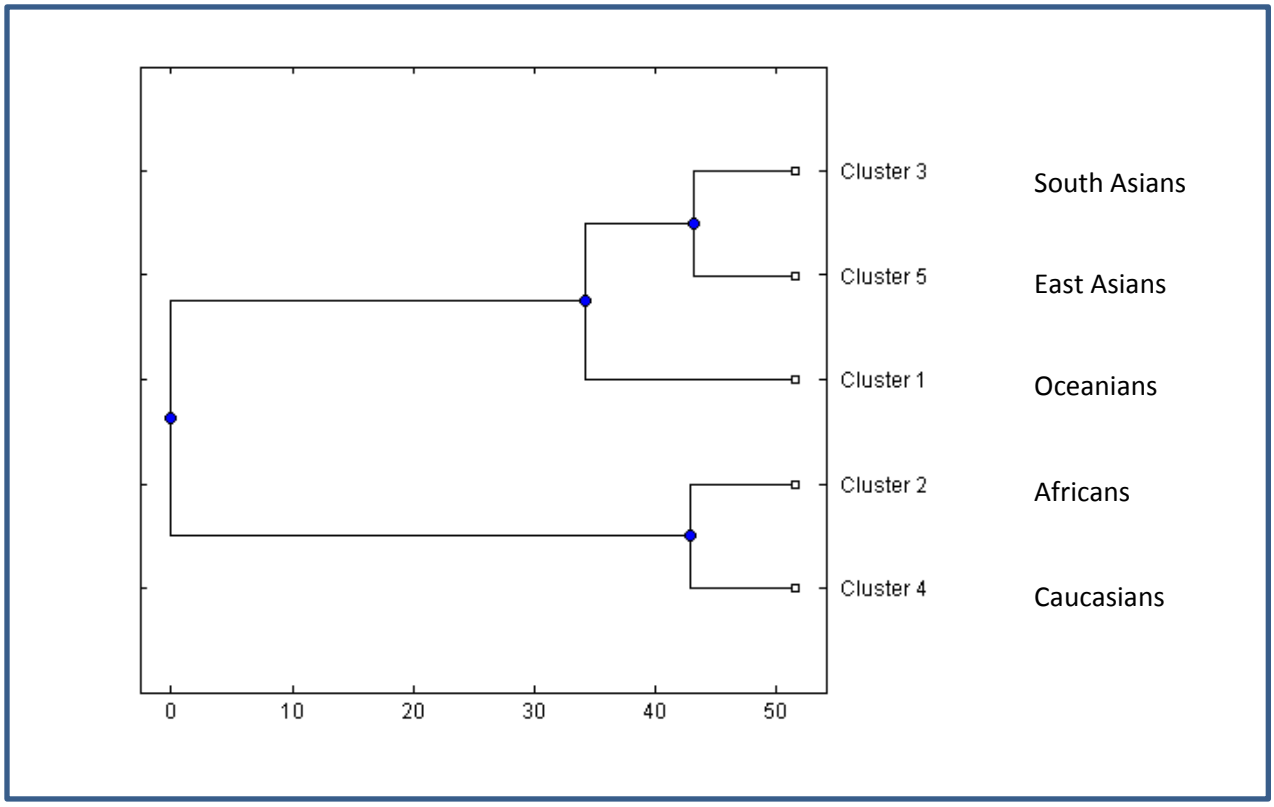


Figure 4. A phylogenic tree using MLST data showed the evolutionary relationship among the five clusters. The five clusters were then determined to be South Asians, East Asians, Oceanians, Africans, and Caucasians.

Mapping the H.pylori vacA diversity

Using the population assignment output, H.pylori vacA gene distribution was mapped in order to see the pattern of the distribution. When more than half of the individuals from each country share the same cluster, the particular color was selected, and mapped in the map(Fig 5). Australia and Costa Rica were excluded from this analysis since they did not have vacA gene data. Since USA and China showed similar ratio among more the clusters, two or three colors reflecting the cluster number were represented in the map. The further the geographical location is from Africa, the vacA gene seems to belong to cluster 3.

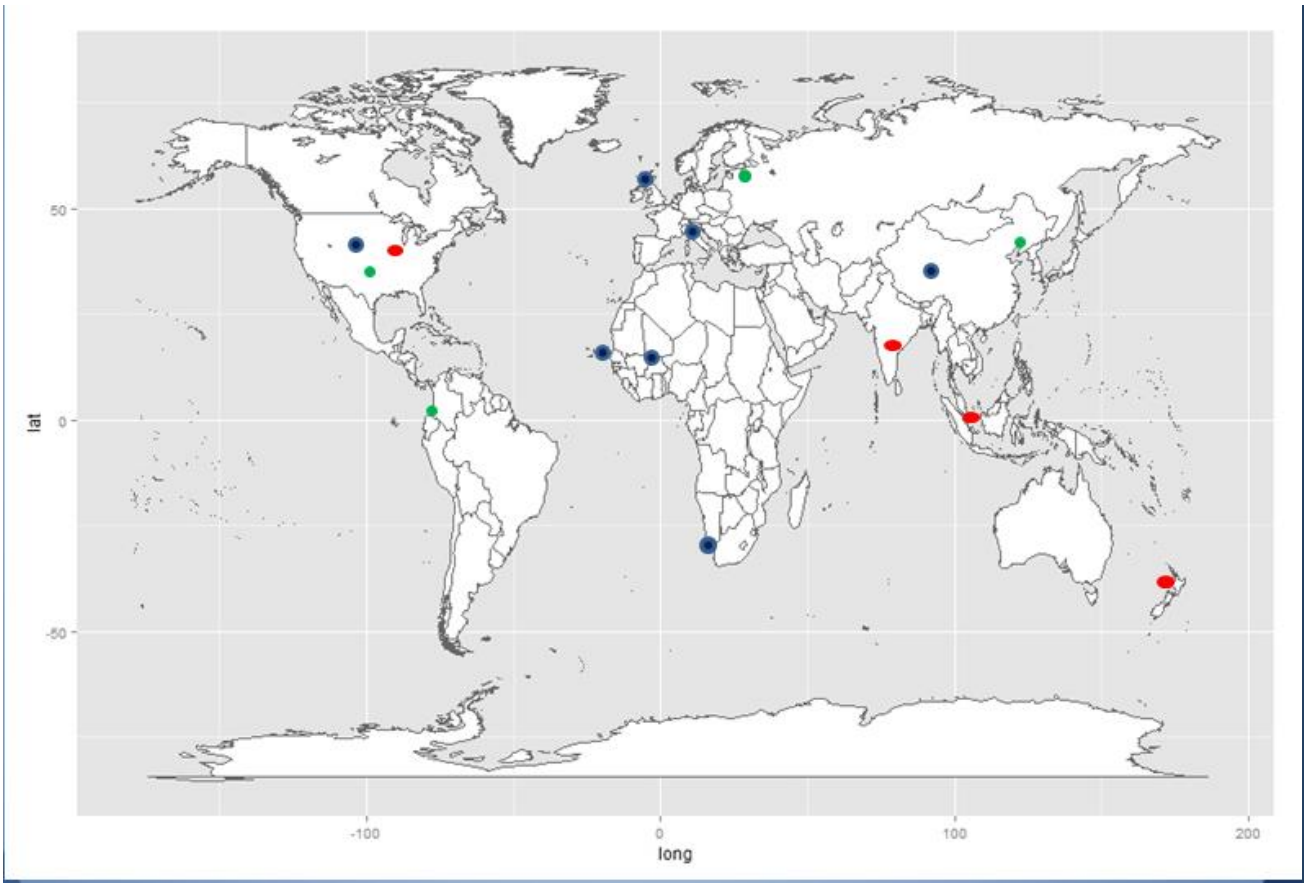


Figure 5. H.pylori Virulence Map
The population assignment of vacA was mapped. USA and China showed similar ratio among the clusters, so they have two or three clusters represented in the map. Blue, Red, Green correspond to cluster1, cluster2, cluster3 respectively. R statistical software package was used for generating the world map.

Human Migration Map: Genographic project by National Geographic Society and IBM

The Genographic Project is a multi-year genetic anthropology study by National Geographic Society and IBM, and its goal is to reconstruct the history of human migration by collecting and analyzing DNA samples from thousands of people from around the world (National Geographic Society). In this project, genetic markers on mitochondrial DNA (mtDNA), and Y-chromosomes were used to trace the participant's distant ancestry. Nearly 500,000 individuals have participated in this project with field research conducted by 11 regional centers to advance the science and understanding of migratory genealogy (IBM). This joint project confirmed that humans migrated out of Africa, and constructed a migration route map (Fig 6).

Fig 6. Human Migration Map by Genographic Project
The map shows the route of human migration from Africa known as "out of Africa" expansion of modern humans. Blue arrows show the human migration direction. This map reflects the human migration history which can be traced back to 60,000 years ago.

Discussion

H.pylori infection is a carcinogen, and it is estimated that about two-third of the world's population carries the bacteria in their stomach. Of those infected, who will develop disease is influenced by the virulence of the infecting H.pylori strain, the genetic susceptibility of the host and environmental co-factors (Atherton 1998). This project examined the virulence of the H.pylori by looking at the distribution of the H.pylori populations along with the human migration.

Virulence Factor vacA gene diversity

The *vacA* gene is well-established *H. pylori* virulence factor that interact with the host cells and disrupt downstream signaling pathways (Suzuki et al, 2012). The population assignment showed that the majority of the African individuals belonging to cluster 1, whereas the majority of the East/South Asian individuals belonging to cluster2. European individuals showed highly mixed diversity of the three clusters. This can be interpreted as the East/South Asians have the most toxic s1m1 strains, and the Africans have the non-toxic s2m2 strains, and the Europeans have both strains. This makes sense in term of the discrepancy of the virulence of H.pylori infections among different geographical locations, because East/South Asians have the most toxic structure of vacA gene, causing the H.pylori to develop the severe disease.

H.pylori genetic diversity

In order to see the genetic diversity of the H.pylori strains, MLST data with seven housekeeping loci were used for population assignment analysis. The result showed high heterogeneity among African individuals, followed by European individuals. East Asians showed the least heterogeneity, indicating there have very low diversity in H.pylori strains. People from Oceania, especially individuals from New Zealand showed homogeneous cluster reflecting that they suffered a genetic bottleneck, and the reduced genetic diversity in individuals from New Zealand is evident in the result.

Virulence Map and Human Migration Map

The H.pylori virulence map and human migration map were then compared. By looking at the virulence map (Fig 5), the further the geographical location is from Africa, the vacA tends to have the s1m1 strain which is most toxic. The population assignment from housekeeping genes also reflected that the further from Africa, the lower the genetic diversity in H.pylori. These two observations seem to be accordance with human migration map generated by Genographic project. Human migration map shows the migration routes, and according to the routes shown in the map, East Asia and South America are the furthest destinations that humans migrated. Although there were no samples available from South American countries in this project, the individuals from East Asian countries such as China and Korea showed least H.pylori diversity, and most toxic structure of vacA gene. These two findings suggest that H.pylori accompanied humans when they migrated, and the H.pylori lost its diversity along the human migrations as humans migrated further from Africa.

Ethical implication

In the past decade, researchers have dramatically improved our understanding of the genetic basis of complex chronic diseases, such as Alzheimer's disease and type 2 diabetes, through more than 1,000 genome-wide association studies (Bustamante *et al*, 2011). However, the findings from such studies are likely to have less relevance than was previously thought for the world's population as a whole. This is due to the fact that almost 96 % of subjects included in the genome-wide association studies conducted were people of European descent. Since ethnicity plays an important role when an infection like H.pylori infection happens, it is essential to study diverse ethinic geoups.

Conclusion

The geographic differences in the H.pylori pathology can be explained by the distribution of the H.pylori which was accompanied with the human migration history. The H.pylori virulence map was in accordance with the human migration map generated by Genographic Project. The genetic diversity of H. pylori was determined to play an important role in the consequences of infection in different hosts.

Literature Cited

Atherton J, Cao P, Peek J, Tummuru M, Blaser M, Cover T. 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.

Atherton J. 1998.H.pylori virulence factors. British Medical Bulletin: 54; 105-120.

Bustamante C, De La Vega F, Burchard E. 2011. Genomics for the world. Nature 475, 163-165.

Corander J, Gyllenberg M, and Koski T. 2007. Random Partition models and Exchangeability for Bayesian Identification of Population Structure. Bulletin of Mathematical Biology, 69, 797-815.

Corander J, Marttinen P, Sirén J, and Tang J. BAPS: Bayesian Analysis of Population Structure Manual v. 4.14 and 6.0 .

Correa P and Piazuelo B. 2012. Evolutionary History of the *Helicobacter pylori* Genome: Implications for Gastric Carcinogenesis. Gut Liver. 6(1): 21–28.

Domínguez-Bello G, Pérez E, Bortolini C, et al. 2008. Amerindian Helicobacter pylori strains go extinct, as European strains expand their host range. PLoS ONE 3, http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0003307

Falush D, Wirth T, Linz Bodo, and Pritchard K. 2007. Traces of human migrations in Helicobacter pylori populations. Science 299.5612: 1582-5.

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, et al. 1995. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol;12:334–345.

IBM. *The Mapping of Humanity's Family Tree*. Retrieved April 12, 2013, from http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/mappinghumanity/

Margos G, Gatewood AG, Terekhova D, et al. 2008. MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. PNAS;105:8730–8735.

National Cancer Institute. *Helicobacter pylori and Cancer.* Retrieved March 24, 2013, from http://www.cancer.gov/cancertopics/factsheet/Risk/h-pylori-cancer

National Geographic Society. *The Human Journey: Migration Routes*. Retrieved April 2, 2013, from https://genographic.nationalgeographic.com/

Suerbaum S, Michetti P. 2002. Helicobacter pylori infection. N. Engl. J. Med. 347, 1175–1186.

Suzuki R, Shiota S, and Yamaoka Y. 2012. Molecular epidemiology, population genetics, and pathogenic role of Helicobacter pylori. Elsevier Infection, Genetics and Evolution. 12:203-213

Testerman T, McGee D, Mobley H. NCBI Bookshelf. *Chapter 34: Adherence and Colonization*. Retrieved March 4, 2013, from http://www.ncbi.nlm.nih.gov/books/NBK2437/

Quattro. IP: 148.61.173.240

Appendix

A. Missing values were calculated using python code. Total percentages of missing allelic sequences were then calculated for each locus. The example output is shown in the following.

```
**********************************************************************
Python 2.7.2 (default, Jun 12 2011, 15:08:59) [MSC v.1500 32 bit (Intel)] on win32
Type "copyright", "credits" or "license()" for more information.
>>> ========== RESTART ==========
>>>
size of missing sequence of locus1
993
percentage
45.1158564289
size of missing sequence of locus2
905
percentage
44.0175097276
size of missing sequence of locus3
966
percentage
43.929058663
size of missing sequence of locus4
949
percentage
44.8912015137
size of missing sequence of locus5
1004
percentage
44.7614801605
size of missing sequence of locus6
1034
percentage
45.350877193
size of missing sequence of locus7
1031
percentage
45.0611888112
>>>
```

B. Result of the mixture clustering analysis. The individual ids were found in each cluster, which allowed the author to identify the population. The analysis was tested for ten times. The following is the sample output.

RESULTS OF INDIVIDUAL LEVEL MIXTURE ANALYSIS:
Data file: st_revised.txt
Model: independent
Number of clustered individuals: 393
Number of groups in optimal partition: 5
Log(marginal likelihood) of optimal partition: -23321.854

Best Partition:
Cluster 1: {21, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35,
36, 37, 39, 41, 42, 43, 45, 46, 52, 54, 56, 57,
61, 63, 65, 67, 68, 78, 85, 87, 88, 90, 92, 93,
94, 97, 98, 151, 163, 173, 176, 199, 200, 201, 203,
204, 205, 207, 208, 212, 213, 214, 223, 226, 228,
231, 249, 265, 267, 268, 270, 271, 273, 274, 276,
278, 280, 283, 298, 304, 305, 306, 307, 310, 311,
343, 344}
Cluster 2: {11, 16, 20, 24, 60, 69, 104, 112, 116, 128, 129,
141, 142, 149, 166, 167, 169, 170, 180, 202, 217,
229, 230, 237, 244, 246, 248, 250, 251, 252, 254,
255, 256, 259, 261, 262, 263, 264, 266, 282, 290,
296, 303, 312, 314, 315, 317, 319, 322, 324, 332,
335, 339, 346, 348, 350, 352, 358, 360, 361, 363,
364, 366, 367, 368, 369, 371, 372, 375, 376, 377,
378, 381, 383, 385, 386, 387, 390, 393}
Cluster 3: {5, 6, 7, 9, 10, 12, 15, 40, 44, 48, 51, 58, 66,
70, 73, 74, 81, 83, 84, 89, 91, 95, 96, 99, 101,
102, 106, 111, 114, 123, 132, 134, 138, 139, 145,
174, 175, 182, 184, 186, 193, 197, 198, 206, 218,
222, 224, 225, 235, 236, 238, 239, 240, 257, 275,
277, 281, 284, 288, 291, 301, 302, 308, 309, 325,
328, 330, 340, 345, 355, 357, 365, 370, 373, 380,
382, 388, 392}
Cluster 4: {1, 2, 3, 4, 8, 18, 53, 71, 76, 77, 86, 100, 105,
108, 110, 113, 118, 119, 120, 122, 124, 126, 131,
135, 136, 143, 146, 147, 150, 152, 155, 156, 157,
160, 161, 171, 178, 179, 181, 187, 190, 191, 192,
194, 195, 196, 211, 215, 216, 219, 220, 227, 232,
233, 234, 242, 245, 253, 258, 260, 279, 285, 286,
300, 316, 318, 321, 323, 331, 334, 337, 338, 342,
347, 349, 351, 353, 356, 362}
Cluster 5: {13, 14, 17, 19, 22, 23, 38, 47, 49, 50, 55, 59,

62, 64, 72, 75, 79, 80, 82, 103, 107, 109, 115,
117, 121, 125, 127, 130, 133, 137, 140, 144, 148,
153, 154, 158, 159, 162, 164, 165, 168, 172, 177,
183, 185, 188, 189, 209, 210, 221, 241, 243, 247,
269, 272, 287, 289, 292, 293, 294, 295, 297, 299,
313, 320, 326, 327, 329, 333, 336, 341, 354, 359,
374, 379, 384, 389, 391}

Changes in log(marginal likelihood) if indvidual i is moved to group j:

ind	1	2	3	4	5
1:	-20.0	-20.0	-19.9	.0	-19.9
2:	-13.6	-13.6	-13.5	.0	-13.5
3:	-30.6	-30.6	-30.5	.0	-30.5
4:	-11.7	-11.7	-11.6	.0	-11.6
5:	-.2	-.2	.0	-.2	-.1
6:	-.2	-.2	.0	-.2	-.1
7:	-.2	-.2	.0	-.2	-.1
8:	-24.6	-24.6	-24.5	.0	-24.5
9:	-.2	-.2	.0	-.2	-.1
10:	-.2	-.2	.0	-.2	-.1
11:	-.1	.0	.0	-.1	.0
12:	-.2	-.2	.0	-.2	-.1
13:	-.2	-.2	-.1	-.2	.0
14:	-.2	-.2	-.1	-.2	.0
15:	-.2	-.2	.0	-.2	-.1
16:	-.1	.0	.0	-.1	.0
17:	-.2	-.2	-.1	-.2	.0
18:	-7.0	-7.0	-6.9	.0	-6.9
19:	-.2	-.2	-.1	-.2	.0
20:	-.1	.0	.0	-.1	.0
21:	.0	-.1	.0	-.1	.0
22:	-.1	-11.8	-11.7	-11.8	.0
23:	-.2	-5.9	-5.9	-5.9	.0
24:	-.1	.0	.0	-.1	.0
25:	.0	-42.3	-42.2	-42.3	-42.2
26:	.0	-42.3	-42.2	-42.3	-42.2
27:	.0	-19.6	-19.5	-19.6	-19.5
28:	.0	-35.1	-35.0	-35.1	-35.0
29:	.0	-18.3	-18.2	-18.3	-18.2
30:	.0	-37.3	-37.2	-37.3	-30.7
31:	.0	-43.7	-43.6	-43.7	-43.6
32:	.0	-43.7	-43.6	-43.7	-43.6
33:	.0	-29.3	-29.2	-29.3	-29.2
34:	.0	-19.6	-19.5	-19.6	-13.7
35:	.0	-23.5	-23.4	-23.5	-23.4

36:	.0	-.1	.0	-.1	.0
37:	.0	-23.5	-23.4	-23.5	-23.4
38:	-.2	-.2	-.1	-.2	.0
39:	.0	-36.5	-36.4	-36.5	-36.4
40:	-6.0	-6.0	.0	-6.0	-5.9
41:	.0	-43.0	-42.9	-43.0	-42.9
42:	.0	-32.0	-31.9	-32.0	-31.9
43:	.0	-36.6	-36.5	-36.6	-36.5
44:	-6.0	-6.0	.0	-6.0	-5.9
45:	.0	-5.9	-5.8	-5.9	-5.8
46:	.0	-31.3	-31.2	-31.3	-31.2
47:	-.2	-.2	-.1	-.2	.0
48:	-.2	-.2	.0	-.2	-.1
49:	-.2	-.2	-.1	-.2	.0
50:	-.2	-.2	-.1	-.2	.0
51:	-.2	-.2	.0	-.2	-.1
52:	.0	-35.1	-35.0	-35.1	-35.0
53:	-.1	-.1	.0	.0	.0
54:	.0	-.1	.0	-.1	.0
55:	-.2	-.2	-.1	-.2	.0
56:	.0	-5.9	-5.9	-5.9	-5.9
57:	.0	-29.2	-29.1	-29.2	-29.1
58:	-.2	-.2	.0	-.2	-.1
59:	-.2	-.2	-.1	-.2	.0
60:	-.1	.0	.0	-.1	.0
61:	.0	-19.0	-18.9	-19.0	-18.9
62:	-.2	-.2	-.1	-.2	.0
63:	.0	-6.6	-6.5	-6.6	-6.5
64:	-.2	-.2	-.1	-.2	.0
65:	.0	-6.7	-6.6	-6.7	-6.6
66:	-.2	-.2	.0	-.2	-.1
67:	.0	-.1	.0	-.1	.0
68:	.0	-19.0	-18.9	-19.0	-18.9
69:	-.1	.0	.0	-.1	.0
70:	-.2	-.2	.0	-.2	-.1
71:	-.1	-.1	.0	.0	.0
72:	-.2	-.2	-.1	-.2	.0
73:	-.2	-.2	.0	-.2	-.1
74:	-.2	-.2	.0	-.2	-.1
75:	-.2	-.2	-.1	-.2	.0
76:	-.1	-.1	.0	.0	.0
77:	-.1	-.1	.0	.0	.0
78:	.0	-.1	.0	-.1	.0
79:	-.2	-.2	-.1	-.2	.0
80:	-.2	-.2	-.1	-.2	.0
81:	-.2	-.2	.0	-.2	-.1

82:	-.2	-.2	-.1	-.2	.0
83:	-11.8	-11.8	.0	-11.8	-11.7
84:	-23.5	-23.5	.0	-23.5	-23.4
85:	.0	-33.2	-33.1	-33.2	-33.1
86:	-.1	-.1	.0	.0	.0
87:	.0	-7.0	-6.9	-7.0	-6.9
88:	.0	-13.6	-13.5	-13.6	-13.5
89:	-35.2	-35.2	.0	-35.2	-35.1
90:	.0	-29.3	-29.2	-29.3	-29.2
91:	-.2	-.2	.0	-.2	-.1
92:	.0	-7.1	-7.0	-7.1	-7.0
93:	.0	-39.8	-39.7	-39.8	-39.7
94:	.0	-39.8	-39.7	-39.8	-39.7
95:	-.2	-.2	.0	-.2	-.1
96:	-.2	-.2	.0	-.2	-.1
97:	.0	-29.3	-29.2	-29.3	-29.2
98:	.0	-5.9	-5.8	-5.9	-5.8
99:	-.2	-.2	.0	-.2	-.1
100:	-.1	-.1	.0	.0	.0
101:	-.2	-.2	.0	-.2	-.1
102:	-.2	-.2	.0	-.2	-.1
103:	-.2	-.2	-.1	-.2	.0
104:	-.1	.0	.0	-.1	.0
105:	-.1	-.1	.0	.0	.0
106:	-.2	-.2	.0	-.2	-.1
107:	-.2	-.2	-.1	-.2	.0
108:	-.1	-.1	.0	.0	.0
109:	-.2	-.2	-.1	-.2	.0
110:	-7.0	-7.0	-6.9	.0	-6.9
111:	-.2	-.2	.0	-.2	-.1
112:	-5.9	.0	-5.8	-5.9	-5.8
113:	-.1	-.1	.0	.0	.0
114:	-.2	-.2	.0	-.2	-.1
115:	-.2	-.2	-.1	-.2	.0
116:	-5.9	.0	-5.8	-5.9	-5.8
117:	-.2	-.2	-.1	-.2	.0
118:	-.1	-.1	.0	.0	.0
119:	-.1	-.1	.0	.0	.0
120:	-5.9	-5.9	-5.8	.0	-5.8
121:	-.2	-.2	-.1	-.2	.0
122:	-5.9	-5.9	-5.8	.0	-5.8
123:	-.2	-.2	.0	-.2	-.1
124:	-.1	-.1	.0	.0	.0
125:	-.2	-.2	-.1	-.2	.0
126:	-.1	-.1	.0	.0	.0
127:	-.2	-.2	-.1	-.2	.0

128:	-.1	.0	.0	-.1	.0
129:	-.1	.0	.0	-.1	.0
130:	-.2	-.2	-.1	-.2	.0
131:	-.1	-.1	.0	.0	.0
132:	-.2	-.2	.0	-.2	-.1
133:	-.2	-.2	-.1	-.2	.0
134:	-.2	-.2	.0	-.2	-.1
135:	-6.0	-6.0	-5.9	.0	-5.9
136:	-.1	-.1	.0	.0	.0
137:	-.2	-.2	-.1	-.2	.0
138:	-6.0	-6.0	.0	-6.0	-5.9
139:	-6.0	-6.0	.0	-6.0	-5.9
140:	-.2	-.2	-.1	-.2	.0
141:	-.1	.0	.0	-.1	.0
142:	-.1	.0	.0	-.1	.0
143:	-.1	-.1	.0	.0	.0
144:	-.2	-.2	-.1	-.2	.0
145:	-.2	-.2	.0	-.2	-.1
146:	-12.5	-12.5	-12.4	.0	-12.4
147:	-.1	-.1	.0	.0	.0
148:	-.2	-.2	-.1	-.2	.0
149:	-.1	.0	.0	-.1	.0
150:	-35.9	-35.9	-35.8	.0	-35.8
151:	.0	-.1	.0	-.1	.0
152:	-.1	-.1	.0	.0	.0
153:	-.2	-.2	-.1	-.2	.0
154:	-.2	-.2	-.1	-.2	.0
155:	-.1	-.1	.0	.0	.0
156:	-.1	-.1	.0	.0	.0
157:	-5.9	-5.9	-5.8	.0	-5.8
158:	-.2	-.2	-.1	-.2	.0
159:	-.2	-.2	-.1	-.2	.0
160:	-5.9	-5.9	-5.8	.0	-5.8
161:	-35.9	-35.9	-35.8	.0	-35.8
162:	-.2	-.2	-.1	-.2	.0
163:	.0	-6.0	-5.9	-6.0	-5.9
164:	-6.0	-6.0	-5.9	-6.0	.0
165:	-17.7	-17.7	-17.6	-17.7	.0
166:	-.1	.0	.0	-.1	.0
167:	-7.6	.0	-7.6	-7.6	-7.6
168:	-17.7	-17.7	-17.6	-17.7	.0
169:	-5.9	.0	-5.8	-5.9	-5.8
170:	-.1	.0	.0	-.1	.0
171:	-6.6	-6.6	-6.5	.0	-6.5
172:	-.2	-.2	-.1	-.2	.0
173:	.0	-35.1	-35.0	-35.1	-35.0

174: -.2 -.2 .0 -.2 -.1
175: -.2 -.2 .0 -.2 -.1
176: .0 -41.0 -40.9 -41.0 -40.9
177: -.2 -.2 -.1 -.2 .0
178: -6.0 -6.0 -5.9 .0 -5.9
179: -7.0 -7.0 -6.9 .0 -6.9
180: -.1 .0 .0 -.1 .0
181: -.1 -.1 .0 .0 .0
182: -.2 -.2 .0 -.2 -.1
183: -6.0 -6.0 -5.9 -6.0 .0
184: -.2 -.2 .0 -.2 -.1
185: -.2 -.2 -.1 -.2 .0
186: -.2 -.2 .0 -.2 -.1
187: -7.0 -7.0 -7.0 .0 -7.0
188: -.2 -.2 -.1 -.2 .0
189: -.2 -.2 -.1 -.2 .0
190: -35.1 -35.1 -35.0 .0 -35.0
191: -.1 -.1 .0 .0 .0
192: -7.0 -7.0 -6.9 .0 -6.9
193: -.2 -.2 .0 -.2 -.1
194: -35.1 -35.1 -35.0 .0 -35.0
195: -.1 -.1 .0 .0 .0
196: -6.6 -6.6 -6.5 .0 -6.5
197: -.2 -.2 .0 -.2 -.1
198: -6.0 -6.0 .0 -6.0 -5.9
199: .0 -40.0 -39.9 -40.0 -39.9
200: .0 -46.8 -46.7 -46.8 -46.7
201: .0 -.1 .0 -.1 .0
202: -.1 .0 .0 -.1 .0
203: .0 -28.5 -28.4 -28.5 -28.4
204: .0 -27.5 -27.4 -27.5 -27.4
205: .0 -20.9 -20.9 -20.9 -20.9
206: -.2 -.2 .0 -.2 -.1
207: .0 -27.5 -27.4 -27.5 -27.4
208: .0 -12.4 -12.4 -12.4 -12.4
209: -.2 -.2 -.1 -.2 .0
210: -.2 -.2 -.1 -.2 .0
211: -.1 -.1 .0 .0 .0
212: .0 -29.9 -29.9 -29.9 -29.9
213: .0 -24.1 -24.0 -24.1 -24.0
214: .0 -6.0 -5.9 -6.0 -5.9
215: -12.4 -12.4 -12.3 .0 -12.3
216: -6.6 -6.6 -6.5 .0 -6.5
217: -.1 .0 .0 -.1 .0
218: -11.8 -11.8 .0 -11.8 -11.7
219: -5.9 -5.9 -5.8 .0 -5.8

220: -13.1 -13.1 -13.0 .0 -13.0
221: -.2 -.2 -.1 -.2 .0
222: -.2 -.2 .0 -.2 -.1
223: .0 -24.0 -23.9 -24.0 -23.9
224: -11.8 -11.8 .0 -11.8 -11.7
225: -.2 -.2 .0 -.2 -.1
226: .0 -7.1 -7.0 -7.1 -7.0
227: -6.0 -6.0 -5.9 .0 -5.9
228: .0 -12.5 -12.4 -12.5 -12.4
229: -5.9 .0 -5.8 -5.9 -5.8
230: -5.9 .0 -5.8 -5.9 -5.8
231: .0 -24.0 -23.9 -24.0 -23.9
232: -17.5 -17.5 -17.4 .0 -17.4
233: -5.9 -5.9 -5.8 .0 -5.8
234: -17.6 -17.6 -17.5 .0 -17.5
235: -5.9 -5.9 .0 -5.9 -5.9
236: -5.9 -5.9 .0 -5.9 -5.9
237: -.1 .0 .0 -.1 .0
238: -.2 -.2 .0 -.2 -.1
239: -.2 -.2 .0 -.2 -.1
240: -.2 -.2 .0 -.2 -.1
241: -.2 -.2 -.1 -.2 .0
242: -17.6 -17.6 -17.5 .0 -17.5
243: -.2 -.2 -.1 -.2 .0
244: -.1 .0 .0 -.1 .0
245: -5.9 -5.9 -5.8 .0 -5.8
246: -13.5 .0 -13.4 -13.5 -13.4
247: -.2 -.2 -.1 -.2 .0
248: -5.9 .0 -5.8 -5.9 -5.8
249: .0 -5.9 -5.8 -5.9 -5.8
250: -6.0 .0 -5.9 -6.0 -5.9
251: -7.0 .0 -6.9 -7.0 -6.9
252: -20.4 .0 -20.3 -20.4 -20.3
253: -.1 -.1 .0 .0 .0
254: -7.6 .0 -7.6 -7.6 -7.6
255: -18.7 .0 -18.6 -18.7 -18.6
256: -7.6 .0 -7.6 -1.8 -7.6
257: -.2 -.2 .0 -.2 -.1
258: -5.9 .0 -5.8 .0 -5.8
259: -6.6 .0 -6.5 -6.6 -6.5
260: -17.6 -17.6 -17.5 .0 -17.5
261: -12.5 .0 -12.4 -12.5 -12.4
262: -6.0 .0 -5.9 -6.0 -5.9
263: -27.0 .0 -26.9 -27.0 -26.9
264: -14.2 .0 -14.1 -14.2 -14.1
265: .0 -5.9 -5.8 -5.9 -5.8

266:	-.1	.0	.0	-.1	.0
267:	.0	-39.3	-39.2	-39.3	-39.2
268:	.0	-20.5	-20.4	-20.5	-20.4
269:	-.2	-.2	-.1	-.2	.0
270:	.0	-31.7	-31.6	-31.7	-31.6
271:	.0	-6.6	-6.5	-6.6	-6.5
272:	-6.0	-6.0	-5.9	-6.0	.0
273:	.0	-5.9	-5.8	-5.9	-5.8
274:	.0	-19.4	-19.3	-19.4	-19.3
275:	-.2	-.2	.0	-.2	-.1
276:	.0	-32.8	-32.7	-32.8	-32.7
277:	-12.4	-12.4	.0	-12.4	-12.4
278:	.0	-12.4	-12.3	-12.4	-12.3
279:	-6.0	-6.0	-5.9	.0	-5.9
280:	.0	-.1	.0	-.1	.0
281:	-12.4	-12.4	.0	-12.4	-12.4
282:	-.1	.0	.0	-.1	.0
283:	.0	-5.9	-5.8	-5.9	-5.8
284:	-6.6	-6.6	.0	-6.6	-6.6
285:	-6.0	-6.0	-5.9	.0	-5.9
286:	-12.5	-12.5	-12.4	.0	-12.4
287:	-.2	-.2	-.1	-.2	.0
288:	-.2	-.2	.0	-.2	-.1
289:	-.2	-.2	-.1	-.2	.0
290:	-.1	.0	.0	-.1	.0
291:	-.2	-.2	.0	-.2	-.1
292:	-.2	-.2	-.1	-.2	.0
293:	-.2	-.2	-.1	-.2	.0
294:	-.2	-.2	-.1	-.2	.0
295:	-.2	-.2	-.1	-.2	.0
296:	-.1	.0	.0	-.1	.0
297:	-.2	-.2	-.1	-.2	.0
298:	.0	-12.9	-12.8	-12.9	-12.8
299:	-.2	-.2	-.1	-.2	.0
300:	-.1	-.1	.0	.0	.0
301:	-.2	-.2	.0	-.2	-.1
302:	-.2	-.2	.0	-.2	-.1
303:	-.1	.0	.0	-.1	.0
304:	.0	-5.9	-5.8	-5.9	-5.8
305:	.0	-29.2	-29.1	-29.2	-29.1
306:	.0	-36.2	-36.1	-36.2	-36.1
307:	.0	-7.1	-7.0	-7.1	-7.0
308:	-.2	-.2	.0	-.2	-.1
309:	-.2	-.2	.0	-.2	-.1
310:	.0	-.1	.0	-.1	.0
311:	.0	-6.7	-6.6	-6.7	-6.6

312:	-.1	.0	.0	-.1	.0
313:	-.2	-.2	-.1	-.2	.0
314:	-32.1	.0	-32.0	-32.1	-32.0
315:	-37.9	.0	-37.9	-37.9	-37.9
316:	-17.7	-17.7	-17.6	.0	-17.6
317:	-13.9	.0	-13.8	-13.9	-13.8
318:	-.1	-.1	.0	.0	.0
319:	-.1	.0	.0	-.1	.0
320:	-41.0	-41.0	-41.0	-41.0	.0
321:	-17.6	-17.6	-17.5	.0	-17.5
322:	-33.1	.0	-33.0	-33.1	-33.0
323:	-11.8	-11.8	-11.7	.0	-11.7
324:	-17.7	.0	-17.6	-17.7	-17.6
325:	-12.6	-12.6	.0	-12.6	-12.5
326:	-19.4	-19.4	-19.3	-19.4	.0
327:	-6.7	-.8	-6.6	-6.7	.0
328:	-13.2	-13.2	.0	-13.2	-13.1
329:	-.2	-.2	-.1	-.2	.0
330:	-.2	-.2	.0	-.2	-.1
331:	-.1	-.1	.0	.0	.0
332:	-20.9	.0	-20.8	-20.9	-20.8
333:	-17.6	-17.6	-11.7	-17.6	.0
334:	-17.7	-17.7	-17.6	.0	-17.6
335:	-17.7	.0	-17.6	-17.7	-17.6
336:	-35.2	-35.2	-35.1	-35.2	.0
337:	-24.1	-24.1	-24.0	.0	-24.0
338:	-25.1	-19.4	-25.0	.0	-25.0
339:	-39.3	.0	-39.2	-39.3	-39.2
340:	-.2	-.2	.0	-.2	-.1
341:	-.2	-6.0	-5.9	-6.0	.0
342:	-13.5	-13.5	-13.4	.0	-13.4
343:	.0	-29.3	-29.2	-29.3	-22.7
344:	.0	-29.3	-29.2	-29.3	-29.2
345:	-18.3	-18.3	.0	-18.3	-11.8
346:	-19.8	.0	-19.7	-19.8	-19.7
347:	-25.1	-19.4	-25.0	.0	-25.0
348:	-22.1	.0	-16.2	-22.1	-22.0
349:	-25.2	-25.2	-25.2	.0	-25.2
350:	-7.0	.0	-1.1	-7.0	-6.9
351:	-6.6	-6.6	-6.5	.0	-6.5
352:	-5.9	.0	-5.8	-5.9	-5.8
353:	-.1	-.1	.0	.0	.0
354:	-.2	-.2	-.1	-.2	.0
355:	-.2	-.2	.0	-.2	-.1
356:	-.1	-.1	.0	.0	.0
357:	-.2	-.2	.0	-.2	-.1

358:	-5.9	.0	-5.8	-5.9	-5.8
359:	-17.6	-17.6	-11.7	-17.6	.0
360:	-27.4	.0	-27.3	-27.4	-27.3
361:	-5.9	.0	-5.8	-5.9	-5.8
362:	-.1	-.1	.0	.0	.0
363:	-13.2	.0	-13.1	-13.2	-13.1
364:	-21.8	.0	-15.9	-21.8	-21.7
365:	-6.0	-6.0	.0	-6.0	-5.9
366:	-21.8	.0	-15.9	-21.8	-21.7
367:	-13.1	.0	-13.0	-13.1	-7.2
368:	-21.8	.0	-15.9	-21.8	-21.7
369:	-7.0	.0	-6.9	-7.0	-6.9
370:	-18.3	-12.5	.0	-18.3	-18.3
371:	-30.6	.0	-30.5	-30.6	-30.5
372:	-5.9	.0	.0	-5.9	-5.8
373:	-17.7	-11.2	.0	-17.7	-17.6
374:	-7.0	-7.0	-7.0	-7.0	.0
375:	-13.1	.0	-1.4	-13.1	-13.0
376:	-21.5	.0	-21.5	-21.5	-21.5
377:	-13.2	.0	-13.1	-6.7	-13.1
378:	-17.6	.0	-17.5	-17.6	-17.5
379:	-12.9	-12.9	-12.8	-12.9	.0
380:	-.2	-.2	.0	-.2	-.1
381:	-12.8	.0	-7.0	-12.8	-12.8
382:	-11.8	-4.4	.0	-11.8	-11.7
383:	-5.9	.0	-5.8	-5.9	-5.8
384:	-13.6	-13.6	-13.5	-13.6	.0
385:	-11.7	.0	-11.6	-11.7	-11.6
386:	-13.9	.0	-13.8	-13.9	-13.8
387:	-31.7	.0	-31.6	-31.7	-31.6
388:	-6.7	-6.7	.0	-6.7	-6.6
389:	-6.1	-6.1	-6.0	-6.1	.0
390:	-6.0	.0	-5.9	-6.0	-5.9
391:	-11.8	-17.7	-17.6	-17.7	.0
392:	-11.9	-6.1	.0	-11.9	-11.9
393:	-6.6	.0	-6.5	-6.6	-6.5

KL-divergence matrix in PHYLIP format:

5					
Cluster_1	0.000	8.806	8.677	8.751	8.614
Cluster_2	8.806	0.000	8.438	8.574	8.522
Cluster_3	8.677	8.438	0.000	8.484	8.384
Cluster_4	8.751	8.574	8.484	0.000	8.482
Cluster_5	8.614	8.522	8.384	8.482	0.000