iScience, Volume *25*

Supplemental information

Helicobacter pylori genomes reveal Paleolithic human migration to the east end of Asia

Rumiko Suzuki, Naruya Saitou, Osamu Matsuari, Seiji Shiota, Takashi Matsumoto, Junko Akada, Nagisa Kinjo, Fukunori Kinjo, Kuniko Teruya, Makiko Shimoji, Akino Shiroma, Mototsugu Kato, Kazuhito Satou, Takashi Hirano, Masahiro Asaka, Kirill Kryukov, Yoshan Moodley, and Yoshio Yamaoka

Figure S1. Results of STRUCTURE analysis on MLST loci (no-admixture model), Related to STAR Methods

One vertical line corresponds to one strain and the line colors represent populations whose number is designated by K. Ten runs were executed for each K and the results were integrated by CLUMPAK. The best K was 8.

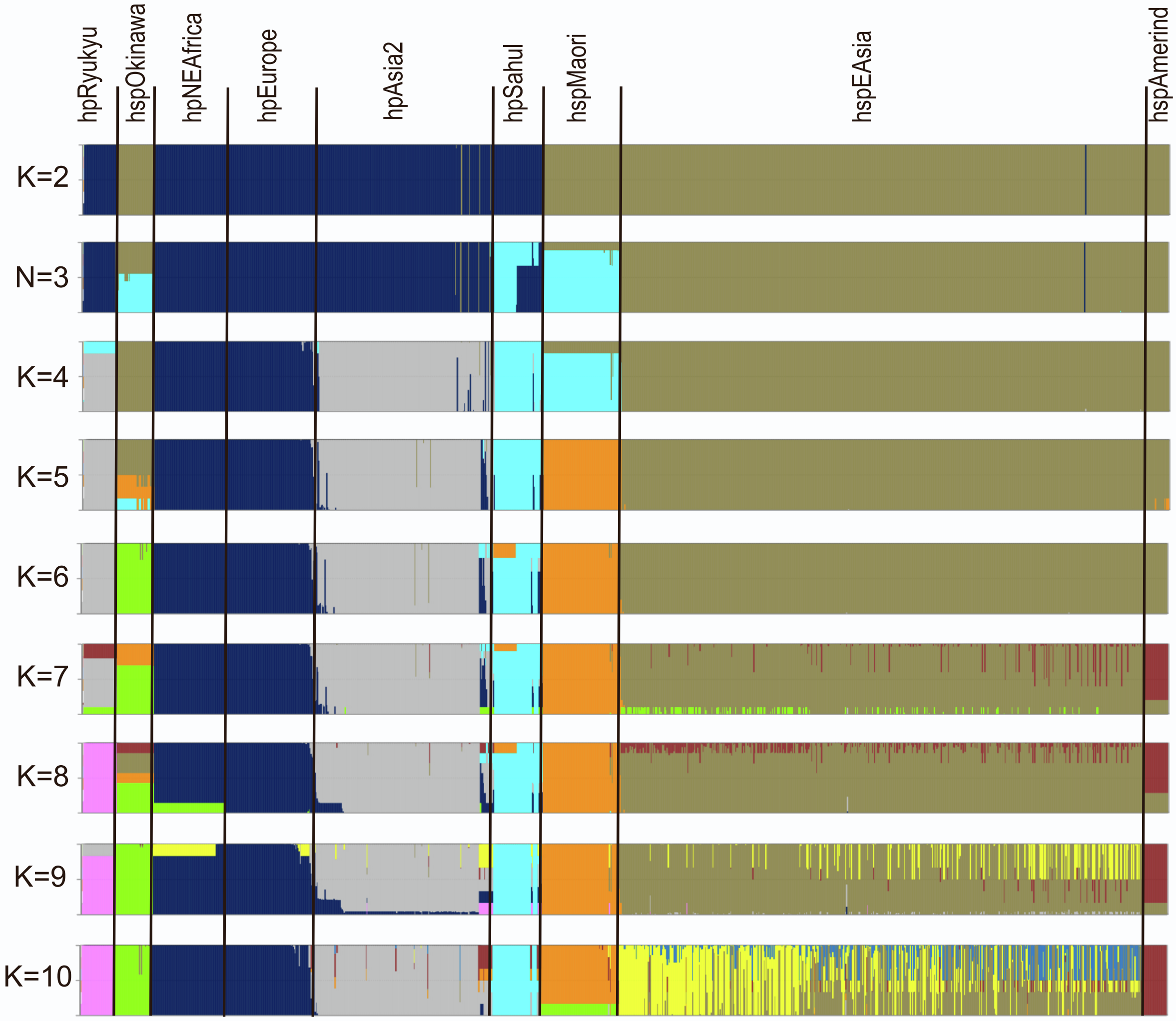


Figure S2. Results of STRUCTURE analysis on MLST loci (no-admixture model), Related to STAR Methods

One vertical line corresponds to one strain and line colors represent populations whose number is designated by K. Ten runs were executed for each K and the results were integrated by CLUMPAK. The ratio of colors in a line reflect the probability of belonging to that population. The best K was 8.

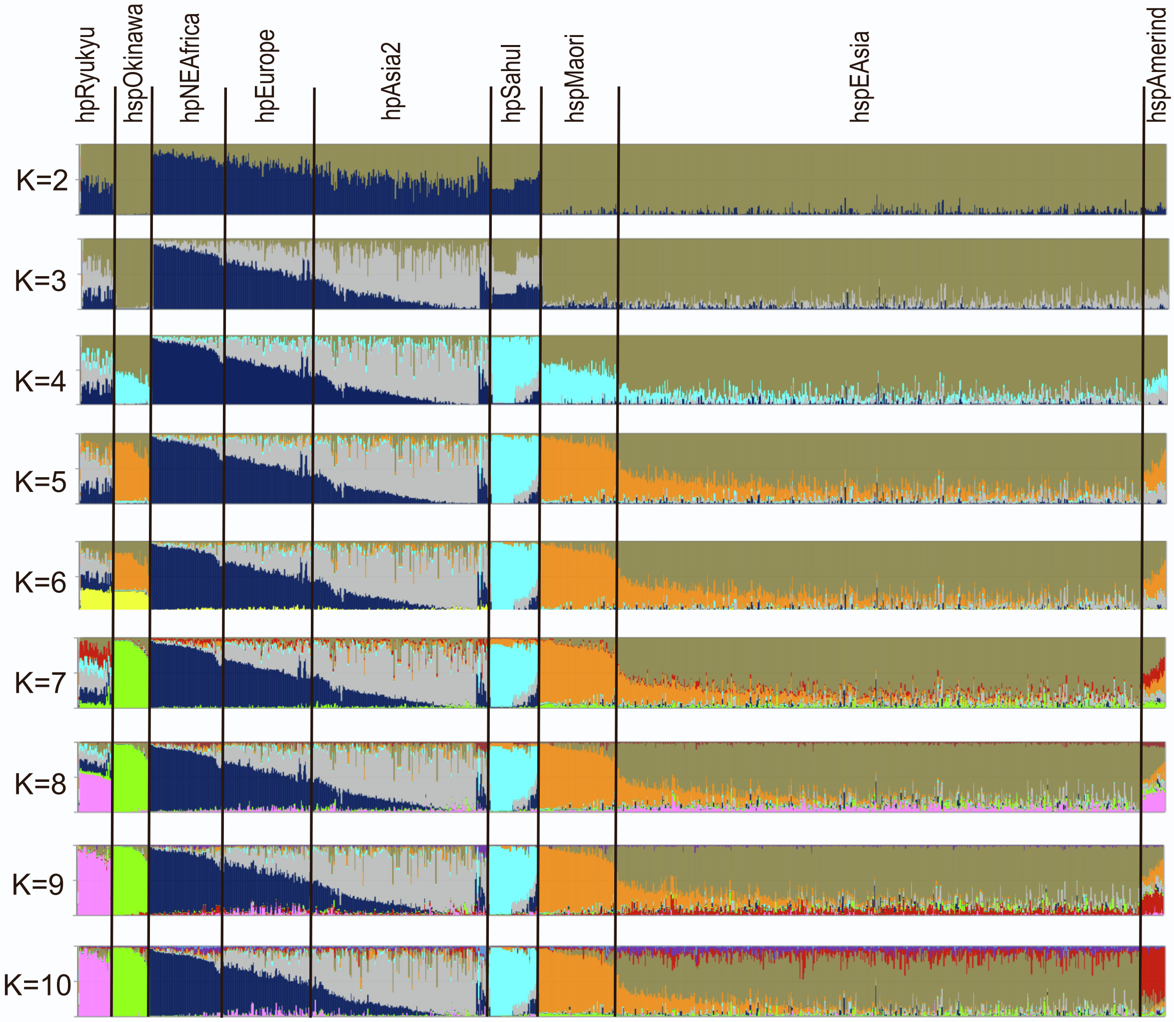
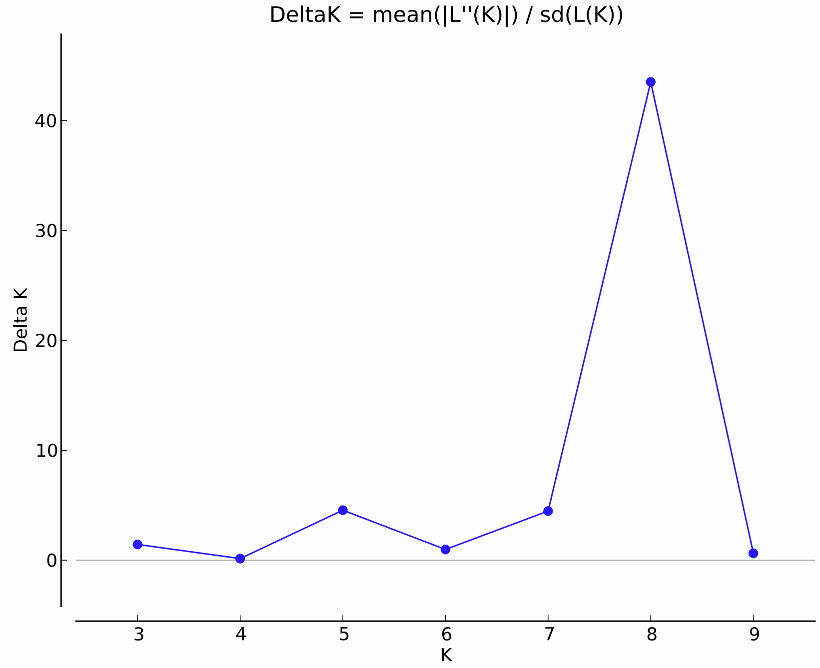


Figure S3. Delta K of the STRUCTURE runs, Related to STAR Methods

Line graphs of Delta K drawn by STRUCTURE harvester. These graphs show that best K is 8 for both no-admixture and linkage models.

No-admixture model



Linkage model

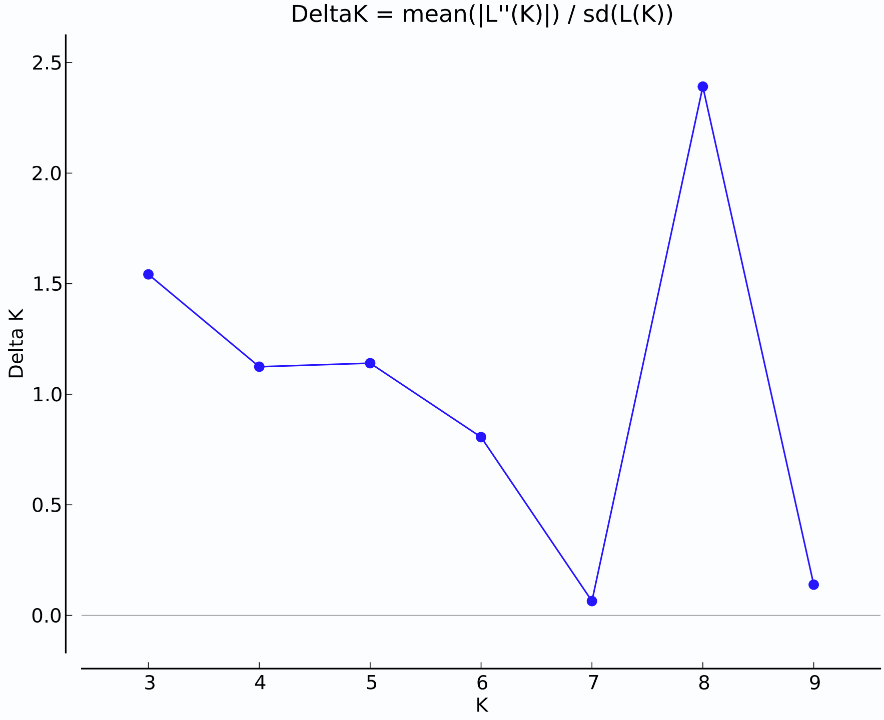


Figure S4. DAPC (discriminated analysis of principal components) plot, Related to STAR Methods

DAPC plot by the same MLST sequence data used for STRUCTURE. We set the number of clusters to 9 referring to BIC.

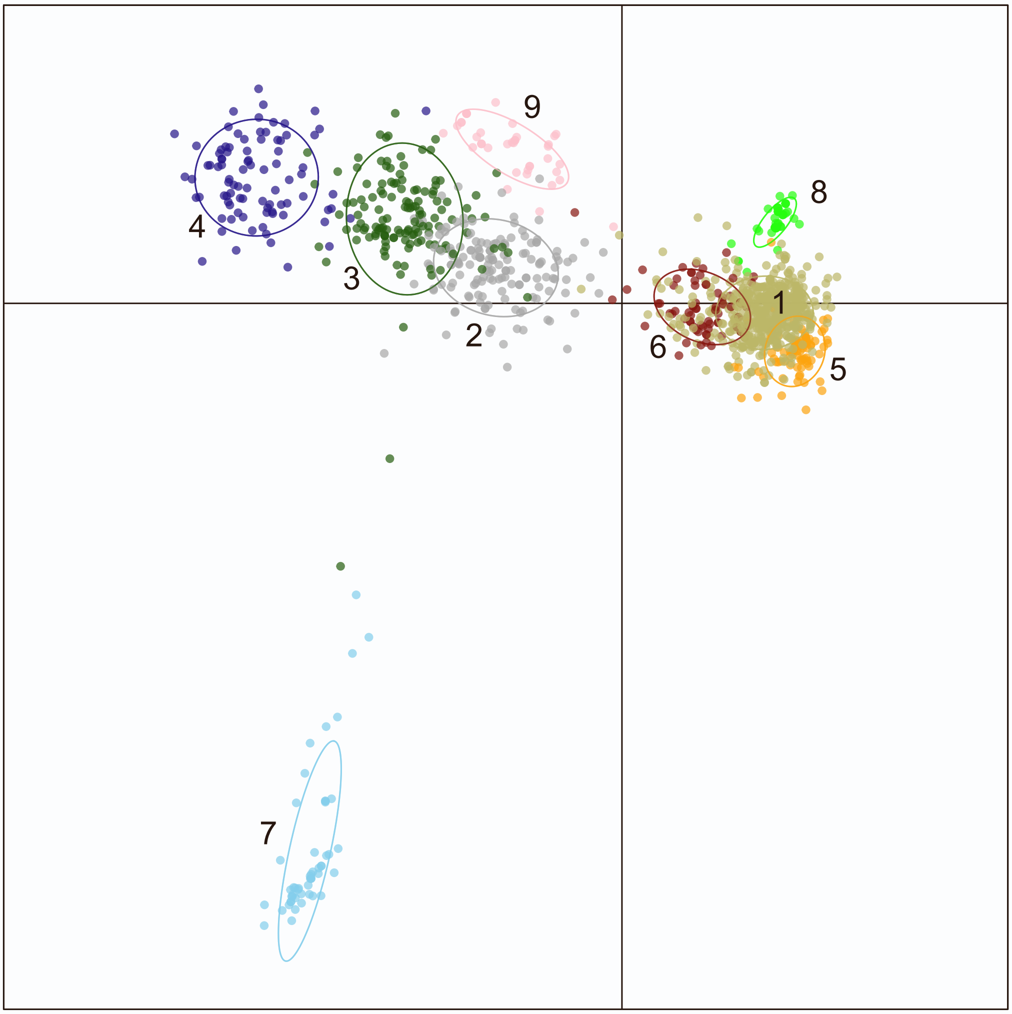


Figure S5. DAPC and STRUCTURE, Related to STAR Methods

The bar chart shows relationship between the clustering results of DAPC and STRUCTURE.

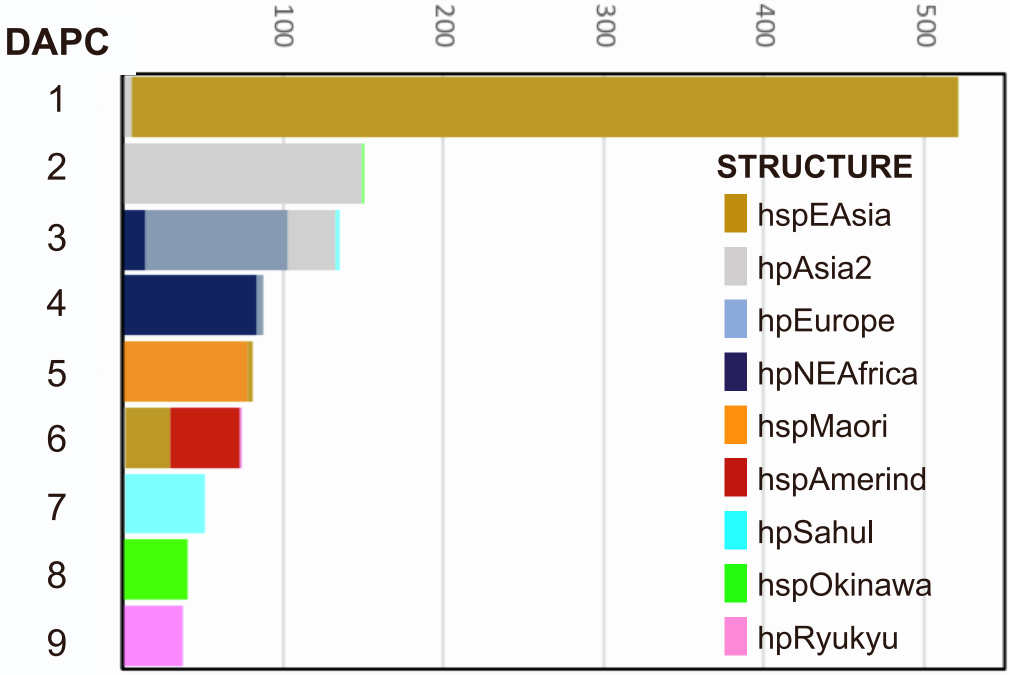


Figure S6. Phylogenetic tree by *cagA* sequences, Related to Figure 1

Phylogeny of cagA among the strains shown in Supplementary figure 6. Sixty-six strains are shown because 22 strains lacked cagA.

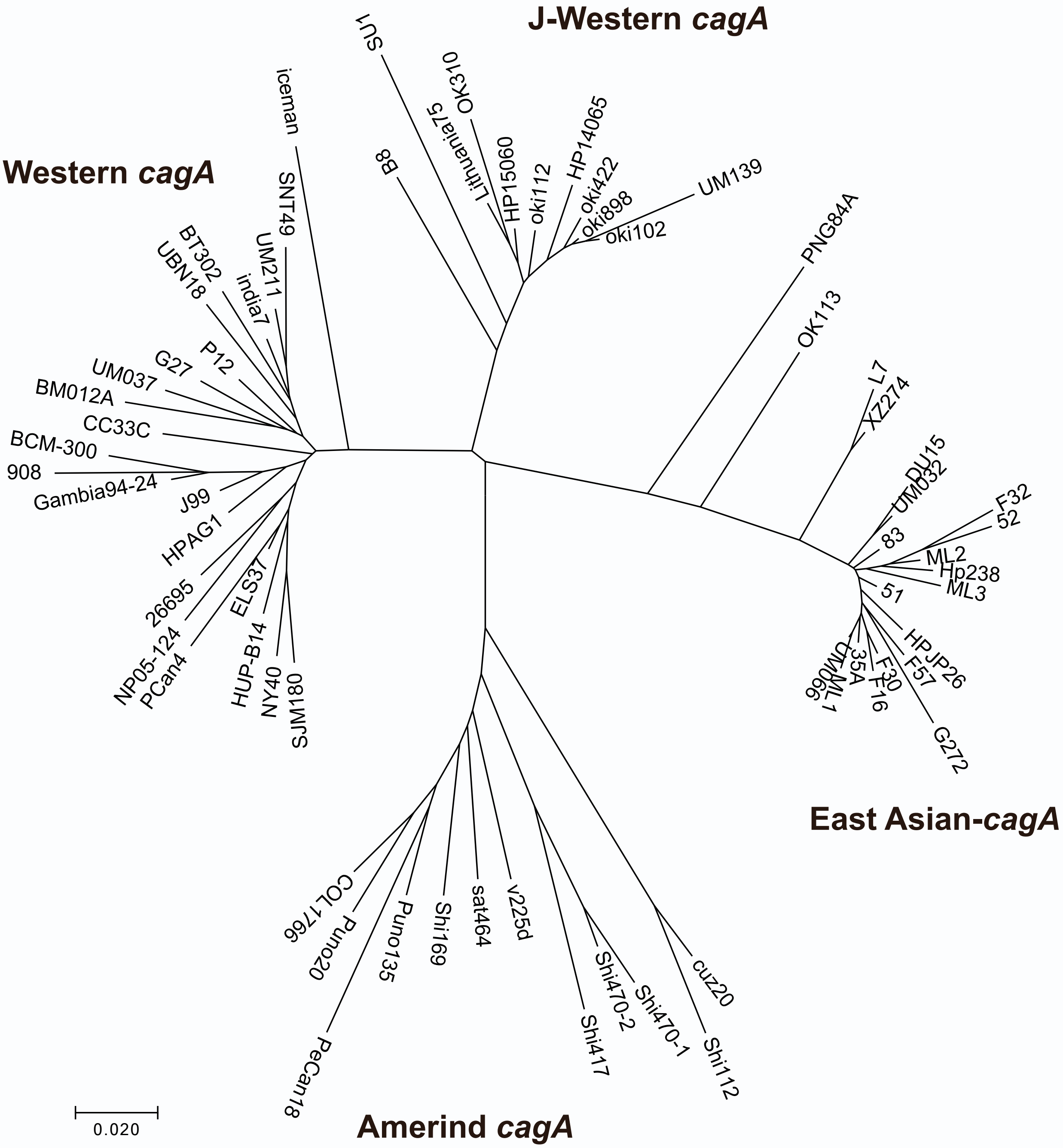


Figure S7. Chromopainting heatmap, Related to STAR Methods

The darker the color (yellow->red->blue), the stronger the connection from the donor to the recipient. The dendrogram shows population clustering. Red arrows on the left side indicate strains selected for BEAST analysis. hpRyukyu and hspOkinawa are boxed by intact and broken lines, respectively.

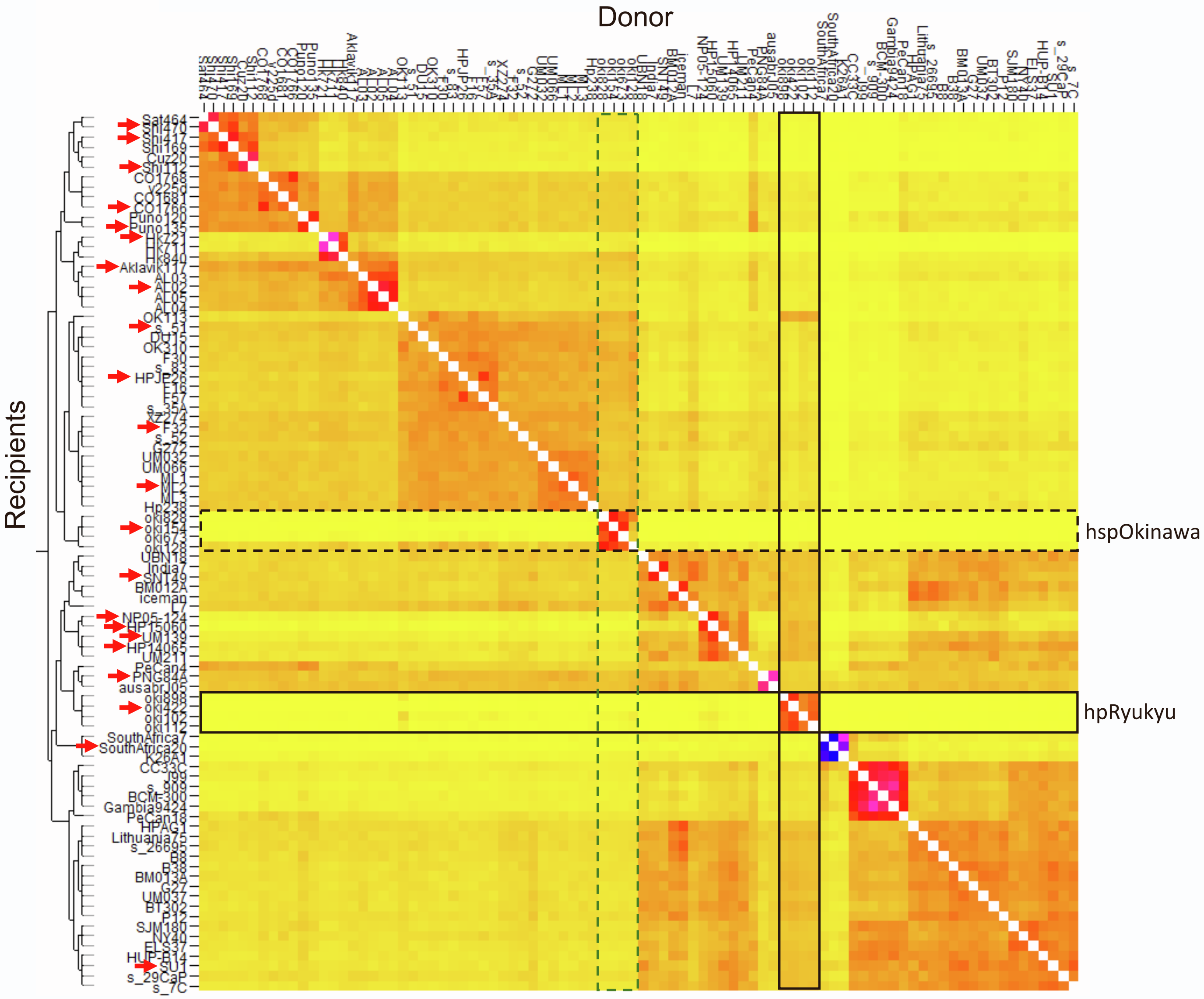


Table S1. Sampling places and numbers of strains for the 8 clusters defined by STRUCTURE, Related to Figure 1

STRUCTURE	Sampling place	Number
hpRyukyu	Okinawa, Japan	35
hspOkinawa	Okinawa, Japan	37
hpNEAfrica	Ethiopia	48
	Nigeria	8
	Bangladesh	6
	Thailand	3
	Algeria	2
	Somalia	2
	Myanmar	1
	NewCaledonia	1
	Taiwan	1
hpEurope	Iran	34
	Bangladesh	19
	Thailand	11
	Nepal	10
	China	9
	Laos	4
	Bhutan	2
	Myanmar	2
	Australia	1
	Japan	1
hpAsia2	Thailand	49
	Nepal	40
	Bangladesh	26
	India	25
	Myanmar	13
	Bhutan	5
	Laos	5
	China	4
	Philippines	3
	Malaysia	2
	Australia	1
	Kazakhstan	1
	SouthAfrica	1
	The Netherlands	1
hpSahul	Australia	25
	PapuaNewGuinea	24
hspMaori	Taiwan	37
	NewZealand	23
	NewCaledonia	13
	Australia	2
	Philippines	2
	Japan	1
hspEAsia	China	161
	Japan	97
	Thailand	86
	Bhutan	57
	Taiwan	30
	SouthKorea	24
	Vietnam	21
	Laos	12
	Singapore	9
	Malaysia	8
	India	6
	Cambodia	5
	Myanmar	4
	NewCaledonia	2
	France	1
	Nepal	1
	Peru	1
	UK	1
hspAmerind	Alaska, USA	8
	Japan	5
	Canada	4
	Colombia	4
	Peru	1
	Venezuela	1

Table S2. Sampling places and numbers of strains for the 9 clusters defined by DAPC, Related to STAR Methods.

DAPC	Country	Number
1	China	155
	Japan	96
	Thailand	83
	Bhutan	44
	Taiwan	30
	SouthKorea	22
	Vietnam	20
	Laos	12
	Singapore	9
	Malaysia	8
	India	7
	Cambodia	5
	Myanmar	5
	France	1
	Nepal	1
	NewCaledonia	1
	Peru	1
	UK	1
2	Thailand	46
	Nepal	25
	Bangladesh	24
	India	24
	Myanmar	12
	Bhutan	4
	Philippines	3
	Malaysia	2
	China	1
	Kazakhstan	1
	Laos	1
	SouthAfrica	1
	The Netherlands	1
3	Iran	34
	Nepal	23
	Bangladesh	19
	Thailand	13
	China	10
	Laos	7
	Bhutan	3
	Myanmar	3
	Australia	2
	Japan	1
4	Ethiopia	48
	Bangladesh	8
	Nigeria	8
	Thailand	3
	Algeria	2
	Nepal	2
	Somalia	2
	NewCaledonia	1
	Taiwan	1
5	Taiwan	37
	NewZealand	23
	NewCaledonia	13
	Australia	2
	Philippines	2
	Thailand	2
	Japan	1
	Laos	1
6	Bhutan	13
	China	8
	Alaska, USA	8
	Japan	6
	Canada	4
	Colombia	4
6	SouthKorea	2
	Thailand	2
	NewCaledonia	1
	Okinawa, Japan	1
	Peru	1
	Venezuela	1
	Vietnam	1
7	Australia	25
	PapuaNewGuinea	24
8	Okinawa, Japan	37
9	Okinawa, Japan	34

Table S3. Helicobacter pylori genomes we added in this study, Related to Figure 1.

Name	Location	Accession number	Genome size (bp)	GC%
AL02	Alaska, USA	CP058250	1,575,789	39.11
AL03	Alaska, USA	CP058288	1,652,551	38.71
AL04	Alaska, USA	CP058287	1,639,336	38.75
AL05	Alaska, USA	CP058286	1,621,023	38.94
BT302	Bhutan	CP058285	1,619,269	38.91
CO1681	Colombia	CP058284	1,531,394	39.12
CO1766	Colombia	CP058283	1,633,485	38.84
CO1768	Colombia	CP058282	1,622,089	38.98
Hk711	Hokkaido, Japan	CP058281	1,687,334	38.64
Hk721	Hokkaido, Japan	CP058280	1,645,786	38.9
Hk840	Hokkaido, Japan	CP058279	1,622,038	38.79
NP05-124	Nepal	CP058252	1,639,844	38.84
oki102	Okinawa, Japan	CP006820	1,633,212	38.81
oki112	Okinawa, Japan	CP006821	1,637,925	38.81
oki128	Okinawa, Japan	CP006822	1,553,826	38.97
oki154	Okinawa, Japan	CP006823	1,599,700	38.8
oki422	Okinawa, Japan	CP006824	1,634,852	38.83
oki673	Okinawa, Japan	CP006825	1,595,058	38.92
oki828	Okinawa, Japan	CP006826	1,600,345	38.8
oki898	Okinawa, Japan	CP006827	1,634,875	38.83
UBN18	Thailand	CP058251	1,656,897	38.78