

Prokaryotic and viral genomes recovered from 787 Japanese gut metagenomes revealed microbial features linked to diets, populations, and diseases; **Cell Genomics** 2, 2022

Yoshihiko Tomofuji, Toshihiro Kishikawa, Yuichi Maeda, Kotaro Ogawa, Yuriko Otake-Kasamoto, Shuhei Kawabata, Takuro Nii, Tatsusada Okuno, Eri Oguro-Igashira, Makoto Kinoshita, Masatoshi Takagaki, Naoki Oyama, Kenichi Todo, Kenichi Yamamoto, Kyuto Sonehara, Mayu Yagita, Akiko Hosokawa, Daisuke Motooka, Yuki Matsumoto, Hidetoshi Matsuoka, Maiko Yoshimura, Shiro Ohshima, Shinichiro Shinzaki, Shota Nakamura, Hideki Iijima, Hidenori Inohara, Haruhiko Kishima, Tetsuo Takehara, Hideki Mochizuki, Kiyoshi Takeda, Atsushi Kumanogoh, Yukinori Okada

We reconstructed 19,084 prokaryotic and 31,395 viral genomes from 787 Japanese gut metagenomes as Japanese metagenome-assembled genomes (JMAG) and Japanese Virus Database (JVD), which are large microbial genome datasets for a single population. Population-specific enrichment of the *Bacillus subtilis* and *b*-porphyranase among the JMAG could derive from the Japanese traditional food natto (fermented soy- beans) and nori (laver), respectively. Dairy-related *Enterococcus*_B *lactis* and *Streptococcus thermophilus* were nominally associated with the East Asian-specific missense variant rs671:G>A in *ALDH2*, which was associated with dairy consumption. Of the species-level viral genome clusters in the JVD, 62.9% were novel. The β crAss-like phage composition was low among the Japanese but relatively high among African and Oce- anian peoples. Evaluations of the association between crAss-like phages and diseases showed significant disease-specific associations. Our large catalog of virus-host pairs identified the positive correlation be- tween the abundance of the viruses and their hosts.

