

MICROBIAL GENOMICS

Plenty of single cells in the sea

Environmental microorganisms exhibit marked and poorly understood genetic variation, even within the same nominal species, that cannot currently be resolved using metagenomics. A large-scale single-cell genomics study of planktonic bacteria and archaea that inhabit the surface ocean now broadens insights into the heterogeneity and genomic composition of this marine microbiome, which has essential roles in global biogeochemical processes but remains understudied owing to difficulties in cultivation.

Using a randomized cell selection strategy, the team generated and sequenced an untargeted library of single amplified genomes (SAGs) of prokaryoplankton retrieved from 28 tropical and subtropical surface ocean water samples. This approach yielded 12,715 partial genomes (an estimated average of 38% of each cell's genome) with no detectable contamination, a data set the authors termed Global Ocean Reference Genomes Tropics (GORG-Tropics).

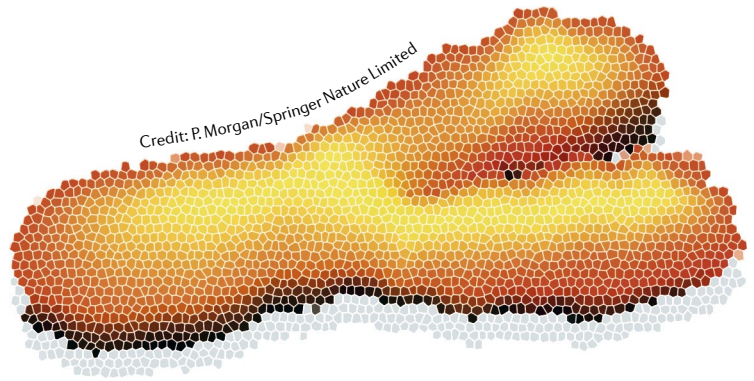
Pairwise average nucleotide identity (ANI) measurements indicated that very few bacterial and archaeal cells in the ocean are sufficiently similar to each other to be called the same species using current species definitions. This large degree of genomic individuality provides a plausible explanation for the challenges in representing complex microbiomes in metagenomic assemblies. Taxonomic assignment based on complete or near-complete 16S rRNA gene sequences was possible for 5,536 SAGs. Identification of genes involved in energy and secondary metabolisms in this large, randomized set of genomes provided a bioprospecting roadmap for bioenergy, pharmaceuticals and other applications.

Fragment recruitment, an alignment-based binning method whereby individual reads are aligned to homologous reference genomes, was used to determine the extent of global diversity represented in GORG-Tropics. This analysis showed that GORG-Tropics outperforms existing reference genome databases, as did a reference data set consisting of a subset of 6,236 genomes, which was obtained from a single, 0.4 ml seawater sample aliquot from the Sargasso Sea. The authors suggest that “a substantial fraction of the global coding potential of marine prokaryoplankton resides in each tiny parcel of ocean water, due to effective mixing on a global scale”.

This study supports a growing body of work using single-cell genomic approaches for cultivation-independent assessment of complex, natural microbial ecosystems, broadens our perspective regarding the extent of genomic heterogeneity in natural microbiomes and offers new tools for improved interpretation of the numerous meta-omics studies of the marine microbiome. As microbial genetic variation can provide a selective advantage when conditions change, a detailed understanding of it might help predict how these communities respond to climate change.

Linda Koch

ORIGINAL ARTICLE Pachiadaki, M. G. et al. Charting the complexity of the marine microbiome through single-cell genomics. *Cell* **179**, 1623–1635 (2019)



GENOMICS

Getting wise about lost Ys

Mosaic loss of chromosome Y (LOY) in leukocytes is the most common form of clonal mosaicism, but its causes and effects are not well understood. Now, a study in *Nature* identifies genetic determinants of LOY, providing insight into how it might occur and its relationship to cancer and other ageing-related diseases.

Using a custom computational approach, Thompson et al. detected LOY in 41,791 of 205,011 (20%) men aged 40–70 years in the UK Biobank. Heritability of LOY was estimated to be 31.7%, and a genome-wide association study identified 156 independent associations with a range of effect sizes (OR 1.03–2.02) and risk allele frequencies (0.25–99.8%). The robustness of these associations for sample type, measurement technique and ancestry was supported by replication in three independent data sets (23andMe, deCODE and Biobank Japan).

Potentially causal variants were identified for each association by fine-mapping. Subsequent analyses indicated that LOY-associated variants are enriched in haematopoietic stem cells and involved in biological pathways associated with cell cycle regulation and the DNA damage response. Moreover, the variants are preferentially found close to cancer-related genes, in keeping with the known association between clonal mosaicism and increased risk of developing haematological cancers. Using a polygenic risk score comprising the 156 LOY variants, the authors demonstrated that the association

between genetic risk of LOY and increased cancer risk also extends to male-specific cancers (prostate cancer and testicular germ cell tumour) and non-sex-specific non-haematological cancers (glioma, renal cell carcinoma, lung cancer and colorectal cancer).

Importantly, in women LOY variants are associated with increased risk of female-specific cancers (breast, ovarian and endometrial cancers) and with age at natural menopause, indicating that the physiological effects associated with LOY cannot be caused solely by loss of Y-encoded functions. Instead, LOY in leukocytes is likely to be a biomarker of genome instability in other cell or tissue types.

The authors propose that the genetic mechanisms predisposing to LOY are shared with genome instability and cancer across multiple cell types. Furthermore, the nature of the genes and pathways associated with LOY suggest it is initiated by chromosome mis-segregation during mitosis and that defects in detecting and eliminating aneuploid cells facilitate subsequent clonal expansion.

Future experiments to test these theories will shed further light on the mechanisms involved in LOY and whether loss of the 45 proteins encoded on the Y-chromosome has any direct physiological consequences.

Dorothy Clyde

ORIGINAL ARTICLE Thompson, D. J. et al. Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**, 652–657 (2019)