Overview

1. Nine new anti-viral/anti-plasmid defense systems discovered with bioinformatics + cloning (Doron S, et al. Science 2018)

2. Huge family of viruses discovered through systematic removal of viral recovery biases (Kauffman K, et al. Nature 2018)

3. Methylation-based approached to affiliate reads with genomes/ mobile elements with hosts (Beaulaurier J, et al Nat Biotech 2017)

Synthesis:

(1) test new viruses on newly discovered defense systems
(2) affiliate viruses with hosts in new metagenomes with less bias
(3) discover how microbial epigenetics influence virus-host interactions

Science

RESEARCH ARTICLE

Cite as: S. Doron *et al.*, *Science* 10.1126/science.aar4120 (2018).

Systematic discovery of antiphage defense systems in the microbial pangenome

Shany Doron,* Sarah Melamed,* Gal Ofir, Azita Leavitt, Anna Lopatina, Mai Keren, Gil Amitai, Rotem Sorek†

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel.

*These authors contributed equally to this work.

†Corresponding author. Email: rotem.sorek@weizmann.ac.il

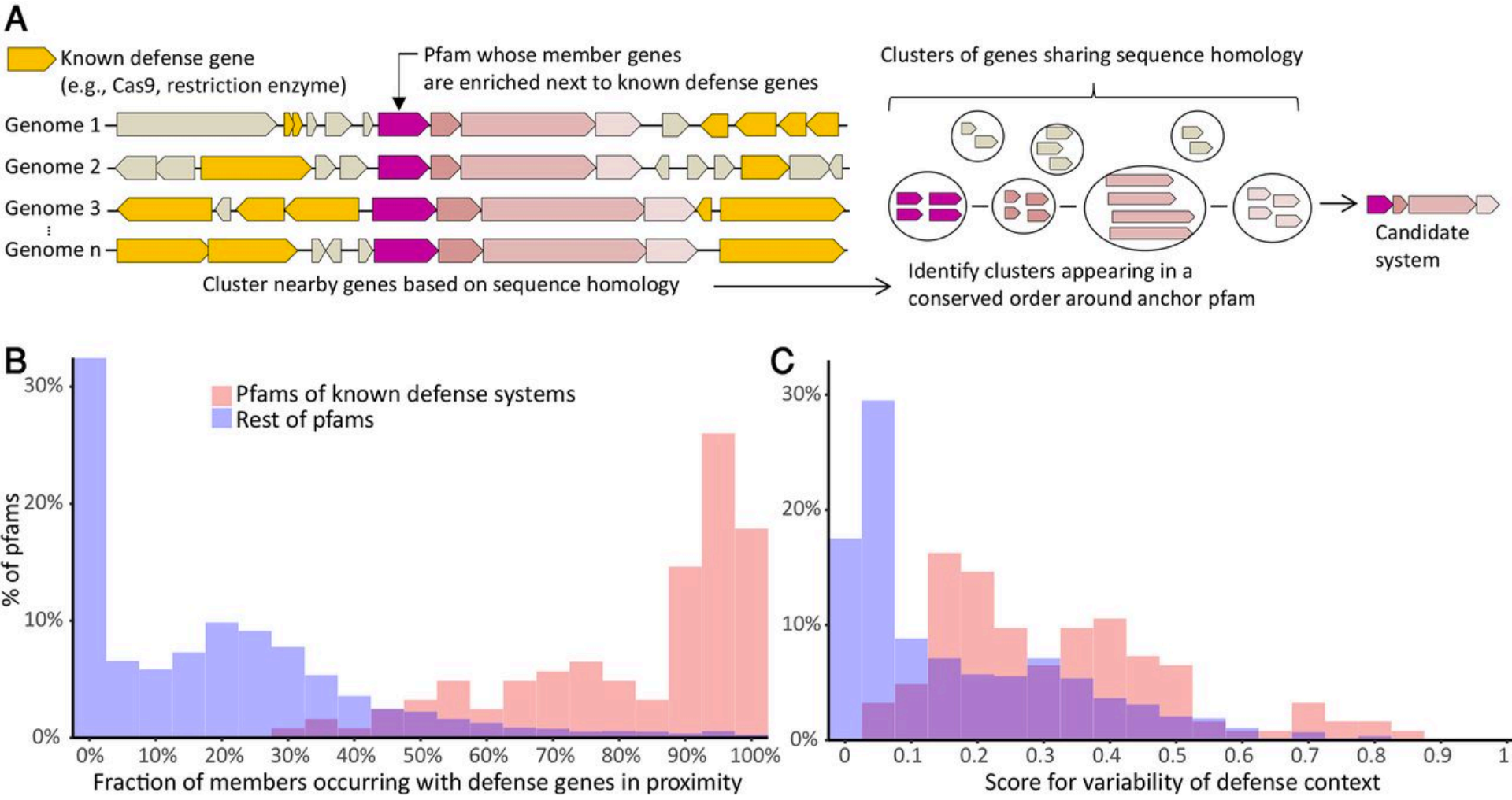


Fig. 1 Discovery of new antiphage defense systems in defense islands.
(A) Illustration of the computational analysis employed for each pfam found to be enriched in defense islands. Pfams that are enriched in the vicinity of known defense genes are identified, and their neighboring genes are clustered based on sequence homology to identify conserved cassettes that represent putative defense systems. (B) Tendency of protein families to occur next to defense genes. The genomic neighborhood for each member gene in each pfam is examined, and the fraction of member genes occurring in the vicinity (10 genes on each side) of one or more known defense genes is recorded. Pink, a set of 123 pfams known to participate in antiphage defense ("positive set"); blue, the remaining 13,960 pfams analyzed in this study. (C) Neighborhood variability score for the analyzed pfams. Score represents the fraction of pfam members occurring in different defense neighborhoods out of total occurrences of pfam members (see Methods). Pink, the 123 positive pfams; blue, a set of 576 pfams that passed the 65% threshold for fraction of members occurring with defense genes in proximity.

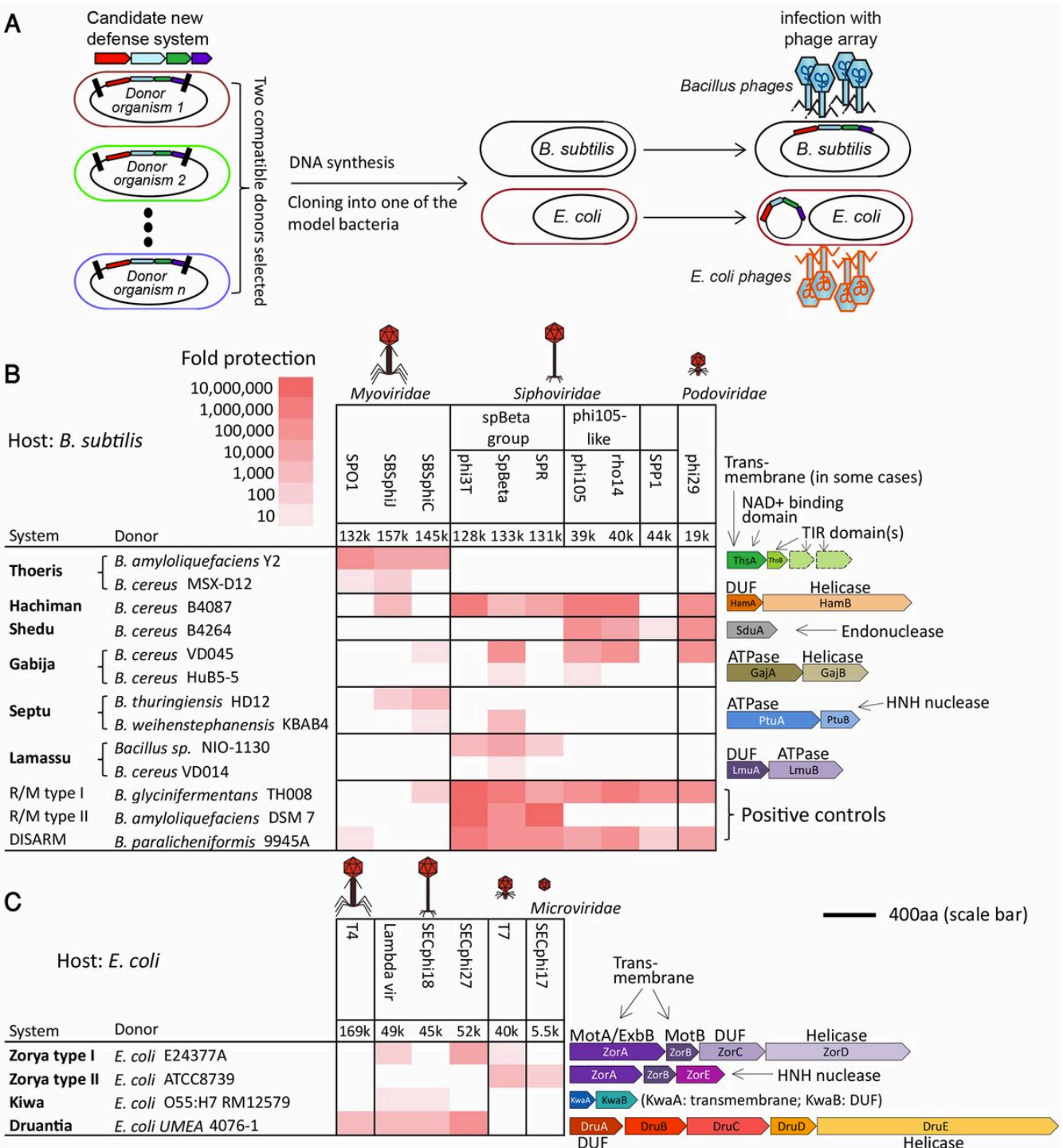


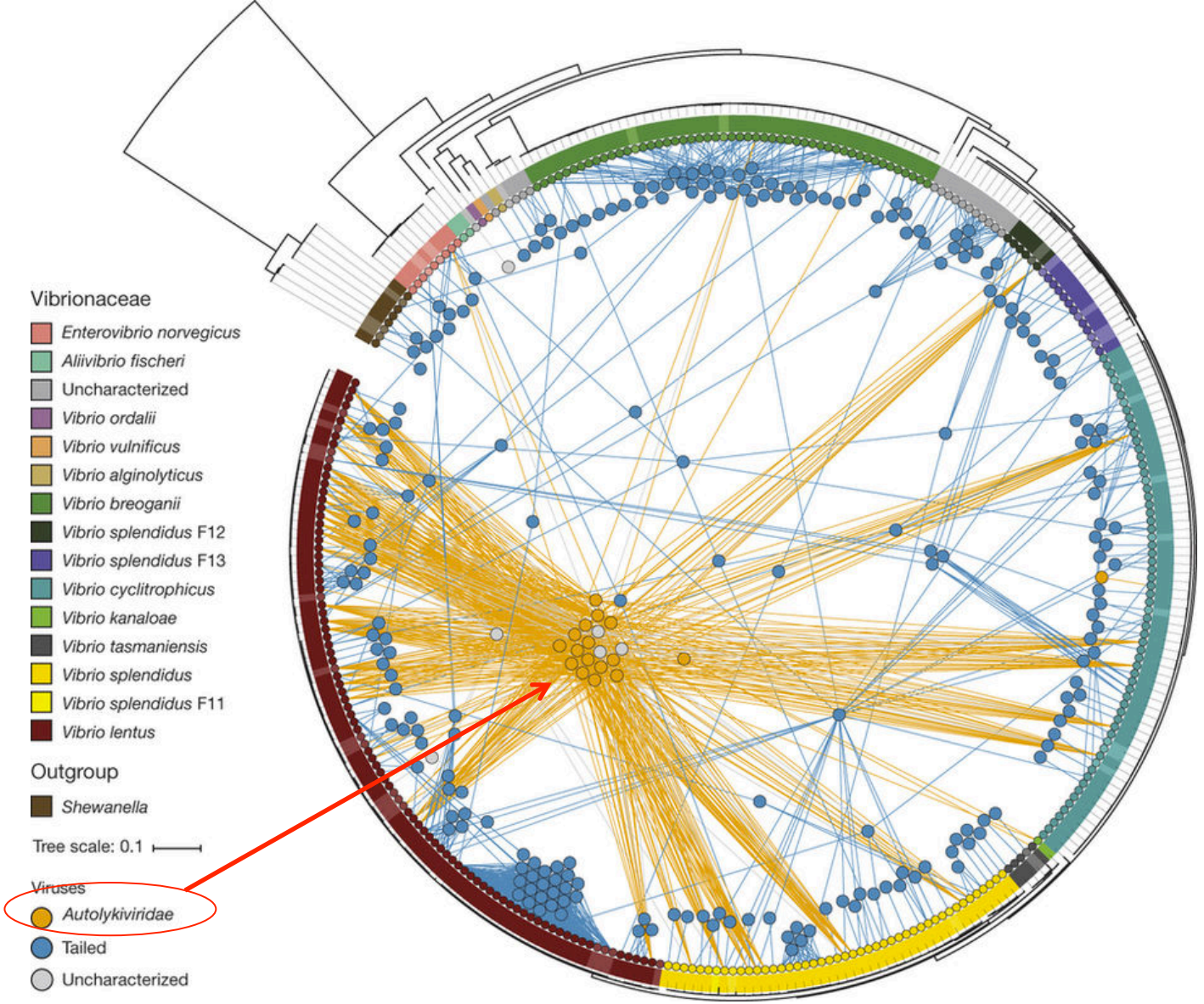
Fig. 2
Experimentally verified defense systems. (A) Flowchart of the experimental verification strategy. (B) Active defense systems cloned into B. subtilis. (C) Active defense systems cloned into E. coli. For B and C, fold protection was measured using serial dilution plaque assays, comparing the system-containing strain to a control strain that lacks the system and has an empty vector instead. Data represent average of three replicates; see figs. S2 and S3. Numbers below phage names represent phage genome size. On the right, gene organization of the defense systems, with identified domains indicated (DUF, domain of unknown function). Gene sizes are drawn to scale; scale bar represents 400 amino acids.

LETTER

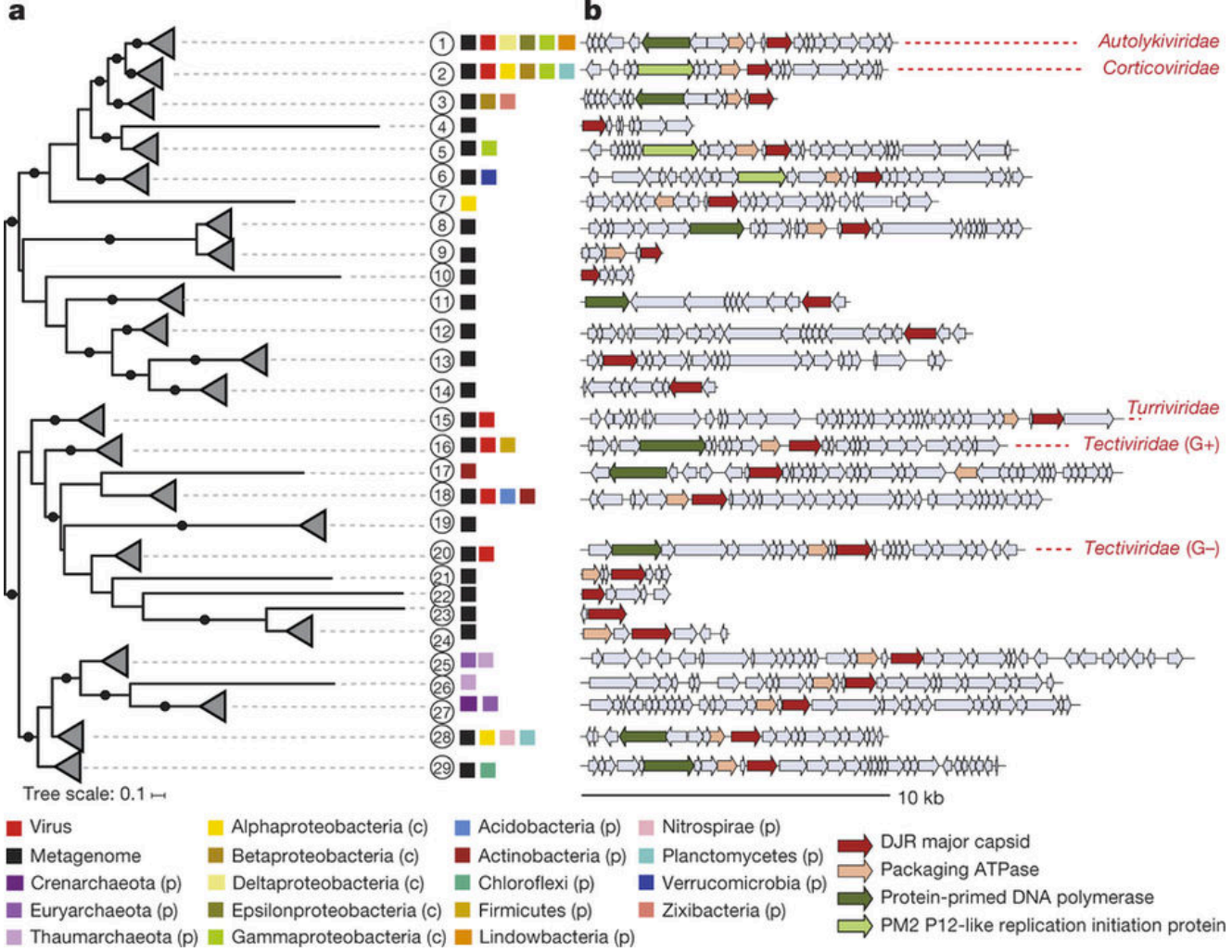
doi:10.1038/nature25474

A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria

Kathryn M. Kauffman[1], Fatima A. Hussain[1], Joy Yang[1], Philip Arevalo[1], Julia M. Brown[2]†, William K. Chang[2], David VanInsberghe[1], Joseph Elsherbini[1], Radhey S. Sharma[1]†, Michael B. Cutler[1], Libusha Kelly[2] & Martin F. Polz[1]



Inverted phylogenetic tree showing relationships among all 318 bacterial strains assayed based on concatenated alignments of hsp60 and ribosomal protein genes, and using a partitioned model in RAxML31 to allow placement of 40 strains for which only the hsp60 gene sequence was available. Isolates are predominantly non-clonal. Leaves represent bacterial isolates coloured by species. Nodes represent 247 viruses described as Autolykiviridae (n = 17), tailed (n = 224) or uncharacterized (n = 6; no genome sequence). The edges represent infections coloured by viral type.



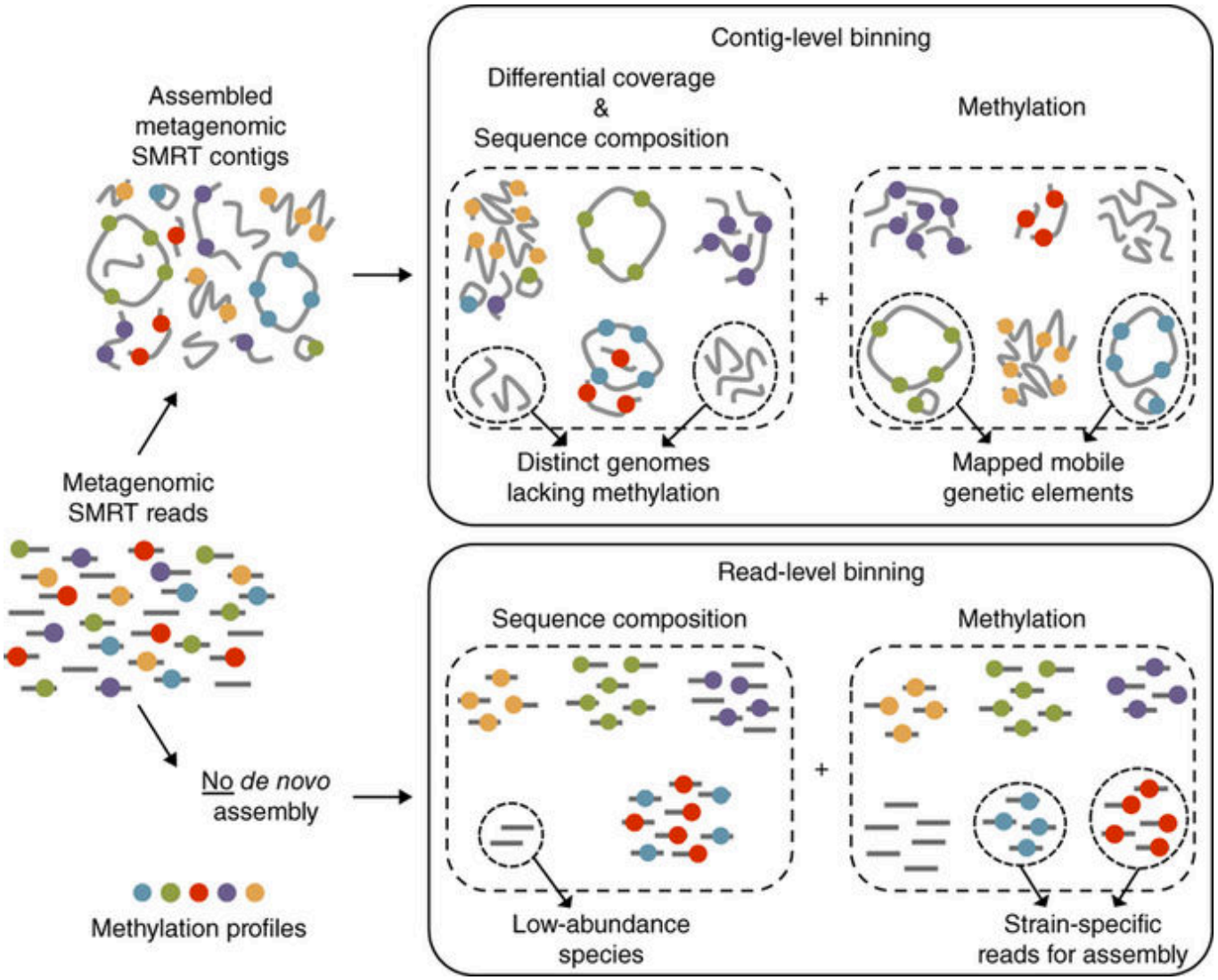
a, Phylogeny of 442 bacterial and archaeal DJR virus capsid proteins (sequences in Supplementary Data 1), including representatives of three previously described DJR virus families and sequences newly identified here; group numbers are assigned to each branch for reference, coloured blocks indicate hosts, black circles on branches indicate approximate likelihood-ratio test (aLRT) branch support ≥0.9. b, Element gene diagrams from each group show prophage host genome neighbourhoods and metagenome contigs often contain additional genes common to DJR elements (contig information in Extended Data Table 2). G+, Gram-positive; G−, Gram-negative.

nature biotechnology

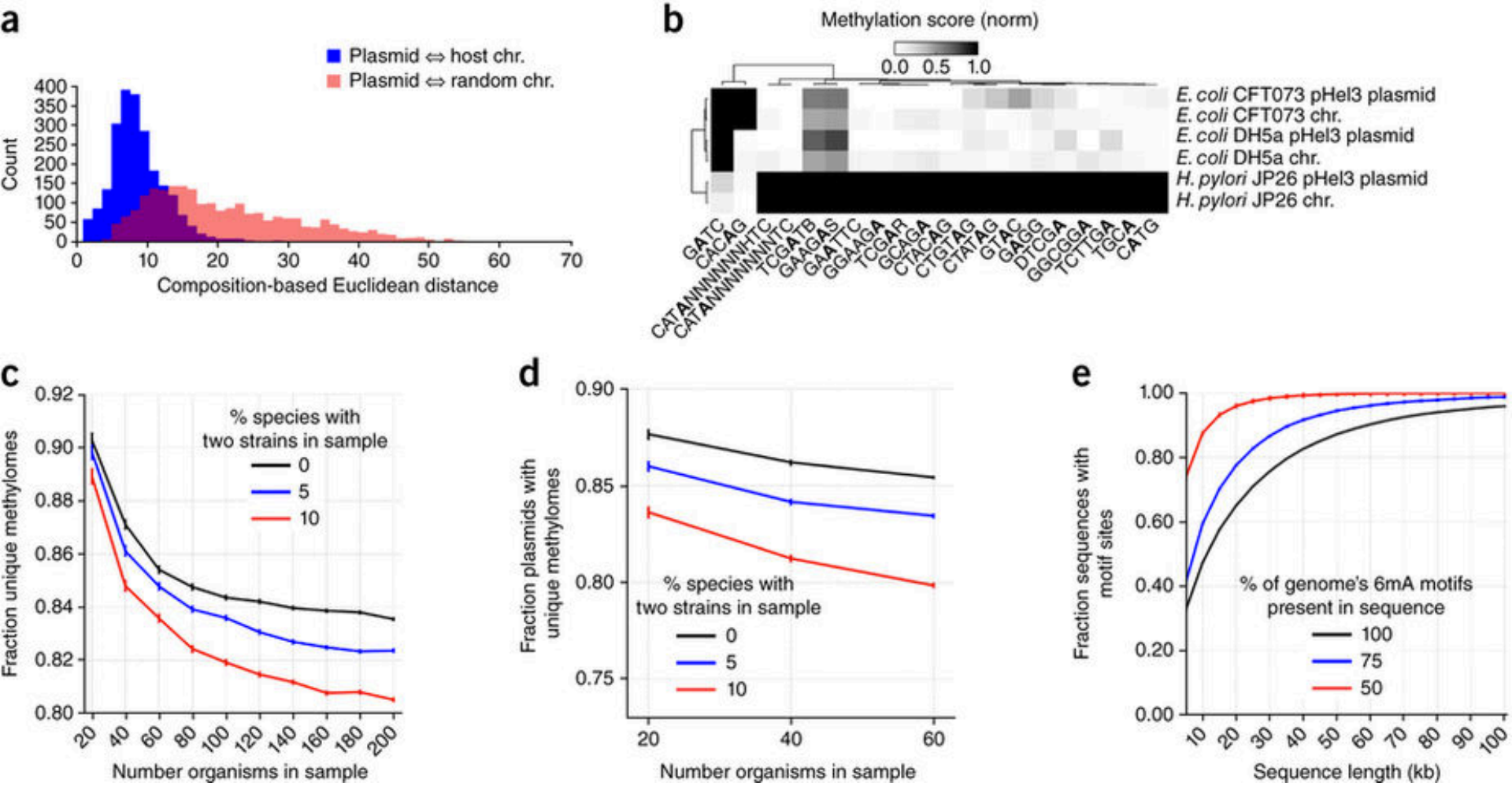
ARTICLES

Metagenomic binning and association of plasmids with bacterial host genomes using DNA methylation

John Beaulaurier[1,2], Shijia Zhu[1,2], Gintaras Deikus[1,2], Ilaria Mogno[1–3], Xue-Song Zhang[4], Austin Davis-Richardson[5], Ronald Canepa[5], Eric W Triplett[5], Jeremiah J Faith[1–3], Robert Sebra[1,2,6], Eric E Schadt[1,2,6] & Gang Fang[1,2]



Given a set of metagenomic shotgun SMRT sequencing reads, one can either assemble them into contigs for contig-level binning or can directly perform read-level binning without de novo assembly. A widely used approach for unsupervised binning of metagenomic contigs uses coverage (and its covariance across multiple samples) and sequence composition profiles, but these can be complemented by methylation profiles to better segregate contigs with similar sequence composition and coverage covariance, as well as to map mobile genetic elements to contigs from their host bacterium in the microbiome sample. Read-level binning by sequence composition can isolate reads from low-abundance species that do not assemble into contigs, while read-level binning by methylation profiles can segregate reads from multiple strains for the purpose of separate, strain-specific de novo genome assemblies. These methylation and composition features can be combined with abundance features to maximize binning resolution.



(a) Histogram of sequence-based Euclidean distance between 5-mer frequency vectors of plasmid and chromosome sequences, showing the distance between plasmids and their host chromosome (blue; based on 2,278 bacterial plasmids and their known hosts), as well as the distance between plasmid and randomly sampled chromosomes from other species (red). (b) Heatmap showing methylation profiles for the pHel3 plasmid and its three hosts: E. coli CFT073, E. coli DH5α, and H. pylori JP26. The methylation profile of pHel3 across 20 motifs matches the host from which it was isolated. (c) Simulation analysis (1,000 iterations) using 878 SMRT sequenced bacterial genomes in the REBASE database showing expected number of genomes with a unique 6mA methylome as a function of community size and presence of multi-strain species in the community. (d) Simulation analysis (1,000 iterations) using 155 SMRT sequenced genomes with known plasmids in the REBASE database showing expected number of genomes with a unique 6mA methylome as a function of community size and presence of multi-strain species in the community. (e) Simulation analysis (500 iterations) using 878 SMRT sequenced genomes in the REBASE database showing the expected sequence lengths required to capture at least one instance of the methylation motifs in a genome. As expected, capturing at least one instance of some, but not all, of the methylation motifs reduces the required sequence length.