GENERALITAT VALENCIANA
Conselleria d'Innovació, Universitats, Ciència i Societat Digital

Federation of European Microbiological Societies

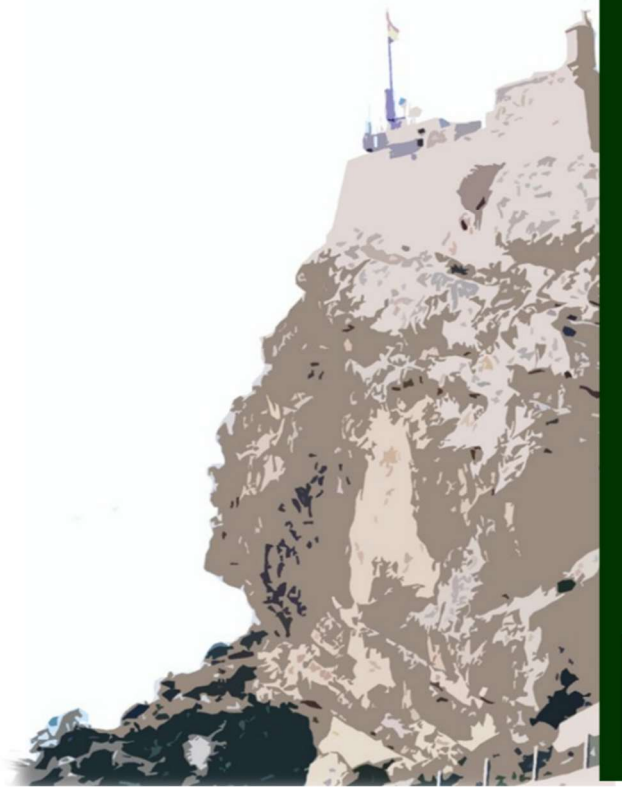
Fundació Fisabio

abstracts book

the Local Pangenome

25-28th October 2023

Alicante, Spain



Alo congress

Universitas Miguel Hernández

Universitat d'Alacant
Universidad de Alicante

INDEX

THE LOCAL PANGENOME, SAN JUAN DE ALICANTE (SPAIN), OCTOBER 25-28, 2023

UNIVERSIDAD MIGUEL HERNÁNDEZ. Edificio Severo Ochoa

Wednesday, 25th October

16.30	*Why the Local-Pangenome? Welcoming address.* ***Francisco Rodriguez-Valera***

17.15-19.15. SESSION I. **DEFINING THE PANGENOME**. *Chair:* Eduardo Rocha

17.15-17.45	**Kostas Konstantinidis** (Georgia Institute of Technology, USA). Comparative omics reveals an ANI-gap within species to define strains and the mechanism that underlies the species and subspecies gaps
17.45-18.15	**Ramunas Stepanauskas** (Bigelow Laboratory for Ocean Sciences, USA). Microbiome-wide rate and phylogenetic range of lateral gene transfer in marine prokaryoplankton
18.15-18.45	**Matthew Kellom** (JGI, USA). Metagenome-assembled genomes, isolate genomes, and pangenomics in large-scale data
18.45-19.15	**Martin Polz** (Vienna, Austria). Defining the role of the pangenome within a population genomic framework

20.00	***Informal gathering with drinks and food. "Castillo de Santa Bárbara" (by BUS)***

Thursday, 26th October

9.00-11.00. SESSION II. **GENOMICS AND THE PANGENOME**. *Chair:* Tal Dagan

9.00-9.30	**Federico Rosconi** (Boston College, USA). Bacterial pangenomes shape essentiality and gene-phenotype associations
9.30-10.00	**Aitor Blanco-Míguez** (University of Trento, Italy). A large-scale pangenome-based framework for improved meta-omics profiling
10.00-10.30	**Eduardo Rocha** (Institute Pasteur, Paris, France). The contribution of interactions between mobile elements to the bacterial pangenome
10.30-11.00	**Yusuke Okazaki** (Kyoto University, Japan). Single-cell genomics uncovers microbial genomic microdiversity and virus-host interactions that had eluded high-resolution metagenomics

11.00-11.30 Coffee break

11.30-13.30. SESSION III. **EVOLUTION OF THE PANGENOME**. *Chair:* Martin Polz

11.30-12.00	**Moritz Buck** (Swedish University of Agricultural Sciences, Sweden). Pangenome traits across Bacteria and Archaea: from genome sizes to functions
12.00-12.30	**Alex Mira** (FISABIO, Valencia, Spain). Do Darwinian evolutionary principles apply to prokaryotes?
12.30-13.00	**Tal Dagan** (University of Kiel, Germany). Pangenomes as documents of the species evolutionary history
13.00-13.30	**Carolina A. Martinez-Gutierrez** (University of North Carolina, Greensboro, USA). Prevalence and Evolutionary Implications of Genome Rearrangements in Bacteria and Archaea

13.30-14.30 Lunch Break

14.30-16.30. SESSION IV. **ECOLOGY AND THE PANGENOME**. *Chair:* Kostas Konstantinidis

14.30-15.00	**Mattias Hotzinger** (Swedish University of Agricultural Sciences, Uppsala, Sweden). The bipartite flexible genome of a freshwater bacterial species - Strain individuality and additive physiological functions
15.00-15.30	**Joshua Hamm** (NIOZ Texel, The Netherlands). Strain-level variation alters host-symbiont interaction dynamics in the *Halorubrum lacusprofundi* - *Candidatus* Nanohaloarchaeum antarcticus system
15.30-16.00	**Jose M. Haro-Moreno** (Universidad Miguel Hernández, San Juan de Alicante, Spain). O-antigen diversity in the order Pelagibacterales
16.00-16.30	**Pedro J. Cabello-Yeves** (University of Warwick, UK). Why do alpha-cyanobacteria with form 1A RuBisCO dominate aquatic habitats worldwide?

16.30-17.00 Coffee break

17.00-19.30. Flash Talks (5 min). *Chair:* Joshua Hamm.

Friday, 27th October

9.00-11.00. SESSION V. **ECOLOGY AND THE PANGENOME II**. *Chair:* Ramunas Stapanauskas

9.00-9.30	**Rachel Whitaker** (University of Illinois, Urbana-Champaign, USA). Contributions of multiscale symbiosis to the dynamics of pangenome
9.30-10.00	**Jose de la Torre** (San Francisco State University, USA). Population-level genome dynamics of archaea in hot spring microbial communities
10.00-10.30	**Felipe Coutinho** (CSIC-ICM, Barcelona, Spain). A deep dive into the genomic diversity of marine Archaea and Bacteria
10.30-11.00	**Zaki Saati-Santamaría** (Czech Academy of Sciences, Prague, Czech Republic). Analysis of the prokaryotic genome evolution within a 5.5 million years trapped chemolithotrophic cave

11.00-11.30 Coffee break

11.30-13.30. SESSION VI. **DYNAMICS OF THE PANGENOME**. *Chair:* Jose Penadés

11.30-12.00	**Miguel Rodriguez Rojas** (University of Innsbruck, Austria). Quantifying genetic variation at the edge of metagenomic resolution to characterize in situ population dynamics
12.00-12.30	**Fiona J. Whelan** (University of Nottingham, UK). Gene-gene associations and the role of selection in microbial pangenomes
12.30-13.00	**Franz Baumdicker** (University of Tübingen, Germany). Common misconceptions when disentangling the processes that shape pangenomes
13.00-13.30	**Mario Rodriguez Mestre** (INIA-CSIC, Madrid, Spain). Regions of Genome Plasticity: A Comparative Analysis of Mobility and Retention Patterns in Bacterial and Archaeal Pangenomes

13.30-14.30 Lunch Break

14.30-17.00. SESSION VII. **PANGENOMES OF ANIMAL/PLANT DENIZENS**. *Chair:* Rafal Mostowy

15.00-15.30	**Aiswarya Prasad** (University of Lausanne, Switzerland). Variation and specialization of structure and function of gut microbial community across host species
15.30-16.00	**Jaime Iranzo** (University Politécnica Madrid, Spain). High-order metabolic interdependencies dominate the human gut microbiome
16.00-16.30	**Anne Kupczok** (Wageningen University, The Netherlands). Large-scale Investigation of orphan genes in the human gut microbiome elucidates their evolutionary origins
16.30-17.00	**Sheila Roitman** (Max Planck Institute for Biology, Tuebingen, Germany). The microvirome: Expanding the plant holobiont

16.30-17.00 Coffee break

17.00-19.15. Flash Talks (5 min). *Chair:* Alex Mira.

20:30 Joint organized dinner. Finca Santa Luzia (Av. de Alicante, 38, BAJO, 03550 Sant Joan d'Alacant, Alicante)

Saturday, 28th October

9.00-11.00. **SESSION VIII. PHAGES AND PANGENOME**. *Chair:* Manuel Martinez-Garcia

9.00-9.30	**Jose Penadés** (Imperial College London, UK). The impact of bacteriophages and PICIs on bacterial evolution and virulence
9.30-10.00	**Rafal Mostowy** (Jagiellonian University, Kraków, Poland). Viewing co-evolutionary dynamics through the microbial pangenome: a multilevel analysis of phage receptor-binding proteins
10.00-10.30	**Chris Bellas** (University of Innsbruck, Austria). Bacteriophage pangenomes from metagenomes
10.30-11.00	**Maite Muniesa** (University of Barcelona). Unveiling the expansive pangenome of Crassvirales: Novel CrAss-like phage isolates reveal genetic heterogeneity and worldwide distribution

11.00-11.30. Coffee break

11.30-12.30	***Closing keynote lecture.*** **Eugene Koonin** (National Center for Biotechnology Information, Bethesda, USA). The Virosphere: Megataxonomy and Global Ecology

Why the Local-Pangenome?

Francisco Rodriguez-Valera

Evolutionary Genomics Group, Universidad Miguel Hernández, San Juan de Alicante, Spain

frvalera@umh.es

It has been more than 20 years since the group of Frederick Blattner published his Venn diagram showing the numbers of genes shared/unique that three different pathovars of *Escherichia coli* had in their genomes. A couple of years later the concept of the pan-genome was postulated by the group of Clare Fraser at TIGR. However, most prokaryotic genomicists (much less microbiologists) are far from integrating this revolution into their frame of mind. The pure culture, the workhorse of microbiology for 150 years is now biasing our interpretation of the microbial world. I too often see authors rejoicing on the description of a single isolate genome as if it represented the biology of the species as a whole. Microbiologists seem to be happy with MAGs (metagenome assembled genomes) that largely miss the flexible genome. The general perception that one single clone predominates and explains the ecology of the species still reigns. The Neodarwinian synthesis based largely on *Drosophila* is a heavy load to carry when studying osmotrophs that live in enormous populations and which diversity might be mindboggling. The mere concept of survival of the fittest is hard to apply here (who is the fittest? one cell? one clone? one species? one population?). The vastness of prokaryotes in the sequence space is astounding (*E. coli* already has a pangenome amounting to 140,000 gene families and counting, 300 new genes for each strain sequenced). The initial reaction of assigning this diversity to the interaction with host immune system has been totally discredited when several free-living bacteria and archaea have been shown to display similar gigantic gene pools. What is the size of the genetic pool of one species and where is its limit? (assuming there are species and limits). It is time that we, microbiologists, got rid of the yoke that animal and plant biology has laid upon us. It is my opinion that the time has come for a scientific revolution in the study of microbes.

Long live The Pangenome!!

Session 1.

Defining the Pangenome

Comparative Omics Reveals an ANI-gap within Species to Define Strains and the Mechanism that Underlies the Species and Subspecies Gaps

Konstantinidis, Kostas

School of Civil and Environmental Engineering and School of Biological Sciences, Georgia Institute of Technology, Atlanta, USA.

kostas@ce.gatech.edu

Metagenomic analyses of cross-sectional or longitudinal samples from various habitats have recently revealed that microbial communities, including viral communities, are predominantly composed of persistent, sequence-discrete, species-like clusters or populations. Specifically, the intra-species sequence relatedness measured -for instance- by the metric of the genome-aggregate average nucleotide identity (ANI) is typically >95% and <90% ANI for inter-species comparisons. However, it remains unclear if such “discontinuities” or gaps in ANI values can be observed within species, and thus used to advance and/or standardize intra-species units such as strains and sequence types. By analyzing 330 bacterial species with at least ten genome representatives each as well as available long-read metagenomes, we show that such a discontinuity also exists between 99.2-99.8% (mean 99.5%) ANI. The 99.5% ANI threshold is largely consistent with how sequence types have been defined in previous epidemiological studies, based on identical sequences in 6-7 genes or loci, but provides clusters with ~20% higher accuracy in terms of evolutionary and gene-content relatedness of the grouped genomes. Consequently, strains should be defined at higher ANI values with >99.99% ANI proposed as the threshold based on enough high gene-content similarity at this level. More recent work has elucidated the mechanism that may underlie the existence of these species and subspecies ANI gaps, which will be highlighted in this talk. Similarities and differences with viral and microbial eukaryotic species will also be discussed. Collectively, our results advance the species concept for all microbes, and the definition and identification of subspecies units.

Recommended reading:

Rodriguez-R, L. M. et al. A natural definition for a bacterial strain and clonal complex. bioRxiv (2022).

Microbiome-wide Rate and Phylogenetic Range of Lateral Gene Transfer in Marine Prokaryoplankton

R. Stepanauskas[1*], J.M. Brown[1], U. Mai[2], M. Pachiadaki[3], O. Bezuidt[1], J.H. Munson-McGee[1], T. Chang[1], S.J. Biller[4], P.M. Berube[5], S. Mirarab[6]

[1] Bigelow Laboratory for Ocean Sciences; East Boothbay, Maine, 04544, U.S.A. [2] Department of Computer Science and Engineering, University of California; San Diego, La Jolla, California 92093, U.S.A. [3] Woods Hole Oceanographic Institution; Woods Hole, Massachusetts 02543, U.S.A. [4] Department of Biological Sciences, Wellesley College; Wellesley, Massachusetts 02481, U.S.A. [5] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology; Cambridge, Massachusetts 02139, U.S.A. [6] Department of Electrical and Computer Engineering, University of California; San Diego, La Jolla, California 92093, U.S.A.

rstepanauskas@bigelow.org

Despite the importance of lateral gene transfer (LGT) in microbial evolution, its quantification in nature has been challenging. Here, we harnessed a global, randomized collection of microbial single cell genomes and a novel computational approach to estimate rates of LGT in the epipelagic zone of tropical and subtropical ocean. We found that LGT involves both "flexible" and "core" genes, with an average cell line acquiring laterally and retaining ~6.5% of its genes every 1 million years. This translates to a net rate of LGT of several hundreds of genes L-1 day-1. Our study indicates that most genes are exchanged among closely related cells, but the phylogenetic range of LGT exceeds the contemporary definition of bacterial species, thus providing prokaryoplankton with extensive genetic resources for LGT-based adaptation.

Keywords: marine prokaryoplankton, microbial single cell genomics, lateral gene transfer

Metagenome-assembled Genomes, Isolate Genomes, and Pangenomics in Large-scale Data

Matthew Kellom

DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA

mkellom@lbl.gov

The Joint Genome Institute (JGI) is a User Facility funded by the U.S. Department of Energy. The JGI provides high-throughput DNA sequencing and analysis services to the user community in support of projects that foster genomic understanding of microbial behavior with an impact on the physical, chemical, and geochemical processes that control elemental and nutrient cycling. As part of this effort, the JGI has focused efforts on microbial sequencing and genome-resolved metagenomics that explores phylogenomic novelty at an increasingly large scale. To date, the JGI has produced tens of thousands of high-quality isolate genomes, single-amplified genomes, and metagenome-assembled genomes that are publicly available [1][2]. As the amount of genomic information continues to grow, the JGI is exploring pangenomics to consolidate and organize data in a way that facilitates visualization and analysis of core and accessory genomes as they relate to niche environmental conditions.

Keywords: metagenome-assembled genomes, single-amplified genomes, isolate genomes, data visualization, binning

References:

1) Nayfach, S., et al. (2021). A genomic catalog of Earth's microbiomes. Nature biotechnology, 39(4), 499-509.
2) Chen, I. M. A., et al. (2023). The IMG/M data management and analysis system v. 7: content updates and new features. Nucleic Acids Research, 51(D1), D723-D732.

Defining the role of the Pangenome within a Population Genomic Framework

Martin F. Polz[1,2*], David VanInsberghe[2]

[1, 2] *Centre for Microbiology and Environmental Systems Science, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria.* [2] *Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 15 Vassar St., Cambridge, MA 02138, USA.*

martin.f.polz@univie.ac.at

Despite appearing identical across their core genome, microbial isolates can differ substantially by their flexible genome, but many functions of this variable gene content remain enigmatic. In particular, our understanding of how selection shapes and structures the flexible genome is limited and hinders our ability to interpret its fine scale diversity. Here we show that a substantial proportion of the flexible genome is structured into modular gene clusters that are potentially involved in social interactions within microbial communities or carry adaptive traits, despite being at intermediate frequencies in populations. We identify population structure by considering both ecology and evolution, and systematically identify all flexible genome regions from 646 strains in nine different marine *Vibrio* populations. A substantial proportion of flexible regions are made up of combinations of modular gene clusters that are present at different frequencies in each population. The distribution and evolutionary history of these regions indicates that they are commonly gained and lost via homologous recombination across flanking sequences, potentially enabling strains to readily change complex social roles in populations. These modules exhibit stable gene content and organization over evolutionary time scales, suggesting they represent fitness peaks that strains rapidly transition between by exchanging modules. In that way, the different modules that make up individual regions are similar to alleles of the same gene, and behave like individual units of selectable diversity, so we propose they be termed heteroforms of the same region. Heteroformic regions appear to function by increasing evolvability while limiting evolutionary costs. The prevalence of heteroformic regions in marine *Vibrio* populations suggests they are an important tool for recombinogenic populations to adapt to rapidly changing environments.

Keywords: population structure, frequency dependent selection, adaptation

Session 2.

Genomics of the Pangenome

Bacterial Pangenomes Shape Essentiality and Gene-Phenotype Associations

Federico Rosconi[1], Tim van Opijnen[2]

[1]*Biology Department, Boston College, Newton, MA, USA.* [2]*Broad Institute of MIT and Harvard, Cambridge, MA, USA*
federico.rosconi@bc.edu

Whole-genome sequencing has led to the astonishing discovery that bacterial isolates from the same species can significantly differ in their genetic content. For instance, two random strains of the bacterial pathogen *Streptococcus pneumoniae* may differ by the presence and absence of hundreds of different genes. Considering that even small genetic changes can have far-reaching phenotypic consequences, we study how *S. pneumoniae's* pangenome influences two crucial aspects of its biology: gene essentiality and mechanisms required for host infection.

By using Tn-Seq, whole-genome sequencing, and RNA-Seq on a set of 36 *S. pneumoniae* clinical strains representing >68% of its pangenome, we have identified the ESSENTIALOME of the pathogen, e.g., the set of genes essential for growth in at least one strain of the collection [1]. These data highlight that some genes are essential in every strain, while for a subset, essentiality is strain-dependent. We show through different genomics approaches that this strain dependency relies on different mechanisms, including (1) toxic intermediates accumulation, (2) functional redundancy, (3) metabolite recycling, and (4) rewiring of critical pathways.

Moreover, by Tn-Seq, we have characterized, with a pan-genome perspective, the *in vivo* phenotype of the same strain collection. We find that 338 genes are required by at least two strains to invade the lungs, and within these, only 77 genes are required by every strain. An in-depth analysis of the strain-dependent required genes reveals that 1) the requirement of some genes depends on how virulent a strain is, and 2) the requirements of central cell processes correlate with the presence/absence of a poorly characterized bacteriocin. These results show that the gain or loss of a single gene could dictate the strategies required by *S. pneumoniae* to survive inside the host.

In conclusion, we demonstrate that bacterial pangenomes make gene-essentiality and specific gene-phenotype associations a fluid concept. Bacterial pangenomes are thereby not only under selection by their environment; the “ebb and flow” of genetic material between genomes generates intragenomic interactions that create additional selective pressures that we are, through the development of new genomics tools, starting to be able to map out and predict.

Keywords: Streptococcus pneumoniae, Tn-Seq, gene essentiality, *in vivo* requirements

References:
1) Rosconi et al. (2015). A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. Nature Microbiology. https://doi.org/10.1038/s41564-022-01208-7.

A large-scale Pangenome-based Framework for Improved Meta-omics Profiling

Aitor Blanco-Míguez[1*], Nicola Segata[1,2]

[1] Department CIBIO, University of Trento, Trento, Italy. [2] IEO, European Institute of Oncology IRCCS, Milan, Italy
aitor-blancomiguez@unitn.it

Culture-independent analyses of microbial communities have progressed dramatically in the last decade, however, most of the methods rely on reference-based information from isolate sequences. In my talk I will discuss a pangenome-based framework I am developing - ChocoPhlAn 4 - which integrates information from metagenome assemblies and microbial isolate genomes for more comprehensive meta-omic profiling. From a curated collection of 1.01 M medium-to-high quality prokaryotic reference and metagenome-assembled genomes, we define pangenomes for 30,550 species-level genome bins, 7,990 of them taxonomically unidentified at the species level. ChocoPhlAn uses massive genomic information and standardized gene calls and gene families to generate SGB-specific markers for taxonomic and strain-level profiling of metagenomes with MetaPhlAn[1], StrainPhlAn[2], and PanPhlAn[3], phylogenetic profiling of genomes and metagenomic-assembled genomes with PhyloPhlAn[4], and functional profiling of metagenomes with HUMAnN[5]. ChocoPhlAn pangenomes are in continuous expansion with the addition of novel microbial genomes and provide a uniform shared resource for subsequent meta-omics profiling.

Keywords: Meta-omics, microbiome profiling, metagenomic assembly

References:
1) Blanco-Míguez, A. et al. (2023) Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nature Biotechnology. doi:10.1038/s41587-023-01688-w.
2) Truong, D. et al. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. Genome Research. doi: 10.1101/gr.216242.116.
3) Scholz, M. et al. (2016). Strain-level microbial epidemiology and population genomics from shotgun metagenomics. Nature Methods. doi: 10.1038/nmeth.3802.
4) Asnicar, F. et al. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nature Communications. doi: 10.1038/s41467-020-16366-7.
5) Franzosa, E. A. et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. Nature Methods. doi: 10.1038/s41592-018-0176-y.

The Contribution of Interactions between Mobile Elements to the Bacterial Pan-genome

Eduardo PC Rocha

Institut Pasteur, Université Paris Cité, CNRS UMR3525, 25-28 Rue Dr Roux, 75015 Paris, France
erocha@pasteur.fr

Horizontal gene transfer driven by phages or conjugative elements allows the acquisition of complex adaptive traits and their transmission to subsequent generations. This speeds up evolutionary processes as exemplified by the acquisition of virulence traits in emerging infectious agents and by antibiotic resistance in many human pathogens. I'll describe how differences between mobile genetic elements in terms of their mechanism of vertical and horizontal transmission result in diverse patterns of gene transmission. These patterns are further modified by the interactions between mobile genetic elements within cells. As a result, the changes in pangenomes driven by horizontal transfer can have a multitude of causes and be subject to very diverse selective pressures.

Keywords: Horizontal gene transfer, mobile genetic elements, phage, conjugation, hyper-parasites

Single-cell Genomics uncovers Microbial Genomic Microdiversity and Virus-Host Interactions that had eluded High-Resolution Metagenomics

Yusuke Okazaki[1*], Yohei Nishikawa[2,3],Ryota Wagatsuma[2,4],
Haruko Takeyama[2,3,4,5], Shin-ichi Nakano[6]

1. Institute for Chemical Research, Kyoto University, Japan. 2. Computational Bio Big-Data Open Innovation Laboratory, AIST-Waseda University, Japan.3. Research Organization for Nano & Life Innovation, Waseda University, Japan. 4. Department of Life Science and Medical Bioscience, Waseda University, Japan. 5. Institute for Advanced Research of Biosystem Dynamics, Waseda Research Institute for Science and Engineering, Graduate School of Advanced Science and Engineering, Waseda University, Japan. 6. Center for Ecological Research, Kyoto University, Japan
okazaki@kuicr.kyoto-u.ac.jp

In microbial ecology, exploring the genomic diversity of bacteria and viruses through metagenomics has become a routine task. We previously used long-read sequencing and read mapping to profile intra-species nucleotide and structural microdiversity in a lake microbial ecosystem[1]. However, even such a high-resolution metagenomic approach misses an extremely diverse part of the pangenome due to the difficulty of assembling such parts. Furthermore, although we reconstructed many bacteria and viral genomes from the environment[2], identifying virus-host pairs remains challenging. To address these issues, here we performed single-cell genomics of lake bacterioplankton to uncover previously unknown genomic microdiversity and virus-host interactions. We individually sequenced over 1900 cells collected from two different water layers and two seasons in Lake Biwa and assembled over 850 genomes that satisfied our quality standards. This resulted in one of the most comprehensive single-cell genomic datasets in freshwater systems. The genomes were dominated by major bacterioplankton lineages in the lake, including members of Fonsibacter, Nanopelagicaceae, Ilumatobacteraceae, and the CL500-11 lineage. We detected phage signals in over 100 genomes that were unevenly distributed among bacterial lineages, with *Methylocystis Rhodoferax* and *Synechococcus* displaying a disproportionately high number of phage signals. We identified high-coverage circular phage genomes that shared genetic synteny across different cells of the CL500-11 lineage, which is the dominant bacterioplankton in the hypolimnion. This is the first CL500-11 phage genome reported and was enriched in samples collected during the winter mixing period, which coincided with a sudden decline in CL500-11 abundance, suggesting that this phage plays a central role in the viral shunt during the massive decline event. Our analysis also revealed hypervariable regions in bacterial genomes that had eluded previous metagenomic assemblies, demonstrating the value of single-cell genomics in studying genomic microvariants. Overall, our study provides new insights into virus-host networks and microbial genomic microdiversities in the lake that had not been identified even through intensive metagenomic efforts.

Keywords: Single cell genomics, Virus-host interaction, Microdiversity, Lake bacterioplankton

References:
1) Okazaki et al. (2022). Long-read-resolved, ecosystem-wide exploration of nucleotide and structural microdiversity of lake bacterioplankton genomes. mSystems, DOI:10.1128/msystems.00433-22
2) Okazaki et al. (2019). Genome-resolved viral and cellular metagenomes revealed potential key virus-host interactions in a deep freshwater lake. Environmental Microbiology, DOI: 10.1111/1462-2920.14816

Session 3.

Evolution of the Pangenome

Pangenome traits across Bacteria and Archaea: from genome sizes to functions

Moritz Buck, Mehrshad Maliheh, Fernando Puente Sánchez, Stefan Bertilsson

Swedish University of Agricultural Sciences

moritz.buck@slu.se

Recent advances in sequencing and bioinformatics have expanded the tree of life by providing genomes for uncultured environmentally relevant clades through metagenome-assembled and single-cell genomes (MAGs and SAGs). This expanded diversity can provide novel insights into microbial pangenome structures. We collected around 600,000 genomes, MAGs and SAGs from genbank (~400,000 genomes), the genomic catalogue of Earths microbiomes from JGI (~50,000 genomes), and an in-house database of aquatic MAGs extracted from ~20,000 public metagenomes (~170,000 genomes). With the help of the bioinformatic tool mOTUpan, this collection was used to compute pangenome traits for the roughly 3,000 species clusters with more than 20 genomes amounting to around 150.000 genomes in total. Pangenome traits (core-fraction, size, diversity, pangenome...) were correlated with genomic (genome-size, GC-content, functional traits) and phylogenetic traits (taxonomy and distances). We found that the fraction of the genome that is is encoding core gene clusters is highly variable among different species but does not correlate with genomic or phylogenetic factors. Additionally, certain classes of gene cluster are over-represented in the core genome (e.g., KEGG annotated genes) whereas genes encoding antimicrobial resistance are depleted from it.

Do Darwinian Evolutionary Principles apply to Prokaryotes?

Alex Mira[1], Francisco Rodríguez-Valera[2]

[1] Genomics & Health Department, FISABIO Foundation, Valencia, Spain. [2] Miguel Hernandez University, San Juan de Alicante, Spain

mira_ale@gva.es

Evolutionary Darwinian principles were developed in the frame of multicellular, diploid sexual organisms. In this talk, we will revise some of the main topics covered in Evolution textbooks, identifying the points where current principles apply only partially or do not apply at all to prokaryotes. Speciation, for instance, is biologically explained as the outcome of geographic and genetic isolation. In bacteria, however, it is well established that all microorganisms are everywhere, the environment selecting those that survive, and therefore the geographic isolation concept is unrealistic. Related to this, animal and plant biogeography is explained as the product of migration or by tectonic plates movement. Due to massive dispersal, however, plate tectonics is highly unlikely to influence bacterial geographic distribution. A central issue in evolution, the level at which selection takes place, is also likely to have unique features in prokaryotes. Does it happen at the individual level? If so, what is an “individual” in prokaryotes? How does the existence of clonal populations affect group selection? Are the cases of bacterial suicide under nutrient depletion examples of kin selection? We also underline that bacteria can clearly modify their phenotype and genotype during their "lifetime" by Horizontal Gene Transfer. Couldn't that be considered a case of an "acquired character" in Lamarckian terms? On the contrary, in eukaryotes, sexual reproduction implies the incorporation of new genetic variants (not even new genes) which are not expressed in the individual undergoing sex, but in their offspring. Another major contrast between eukaryotes and prokaryotes would be the source of functional innovation. In eukaryotes, especially vertebrates, differences in functions between species mainly evolve by 1) SNPs on pre-existing genes or 2) changes in gene regulation, whereas in microbes it mainly arises by the incorporation of new genes. As a consequence of that, humans and fruitflies share more genes between them than different strains of *E. coli* among themselves. Therefore, the pan-genome is mainly a prokaryotic feature that is central for understanding microbial evolution. We therefore propose that many of the basic evolutionary processes do not apply to prokaryotes and that a revised paradigm for prokaryotic evolution is required.

Keywords: Prokaryotic evolution, levels of selection, Lamarckian evolution, speciation, biogeography

Pangenomes as Documents of the Species Evolutionary History

Tal Dagan

Institute for General Microbiology, Kiel University

tdagan@ifam.uni-kiel.de

The determination of the last common ancestor (LCA) of a group of species plays a vital role in evolutionary theory. Traditionally, an LCA is inferred by the rooting of a fully resolved species tree. From a theoretical perspective, however, inference of the LCA amounts to the reconstruction of just one branch – the root branch – of the true species tree and should therefore be a much easier task than the full resolution of the species tree. Discarding the reliance on a hypothesized species tree and its rooting leads us to reevaluate what phylogenetic signal is directly relevant to LCA inference and to recast the task as that of sampling the total evidence from all gene families at the genomic scope. The LCA and root inference can thus be reformulated in the framework of statistical hypothesis testing. In my presentation I will outline an analytical procedure to formally test competing a priori LCA hypotheses and to infer confidence sets for the earliest speciation events in the history of a group of species. The liberation from a reference species tree in phylogenomic root inference further enables LCA inference for extrachromosomal elements. Demonstrative examples will be presented, including the evolution of domesticated plasmids as well as reconstruction of transitions in plasmid mobility. Approaching root inference within a statistical framework supplies a generalized phylogenomic inference methodology that is both powerful and robust.

Prevalence and Evolutionary Implications of Genome Rearrangements in Bacteria and Archaea

Carolina A. Martinez-Gutierrez and Louis-Marie Bobay

University of North Carolina at Greensboro
cmartinez@vt.edu

Advances in sequencing technologies during the last few decades have allowed evolutionary biologist to revisit long standing challenges in microbial genome evolution. One of these questions concerns the prevalence and evolutionary implications of genome rearrangements in microbes. Before high-throughput sequencing technologies, the study of genome rearrangements was limited to a few well-studied clades, preventing the scientific community from establishing general trends across bacteria and archaea. Here, we explored the ubiquity of genome rearrangements across the Tree of Life of prokaryotes by analyzing a broad diversity of bacterial and archaeal pangenomes. We also investigated the suitability of short- and long- read sequencing technologies for the study of genome rearrangements. Preliminary results indicate that assembly biases in short-read genomic data may lead to overestimation in the number of genome rearrangements. This finding indicates that long-read genomes must be used for the study of genome rearrangements. Moreover, genome-by-genome analysis of rearrangements suggest that these events are independent of the location across the genome relative to origin and terminus, pointing out that the movement and maintenance of genome rearrangements in prokaryotes is governed by random processes and drivers.

Session 4.

Ecology and the Pangenome

Individuality and Additive Physiological Functions

Matthias Hoetzinger[1], Mario López-Pérez[2], Fernando Puente Sánchez[1], Francisco Rodriguez-Valera[2], Martin W. Hahn[3]

[1] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden. [2] Evolutionary Genomics Group, Departamento Producción Vegetal y Microbiología, Universidad Miguel Hernández, Alicante, Spain. [3] Research Department for Limnology, University of Innsbruck, Mondsee, Austria

matthias.hotzinger@slu.se

Polynucleobacter is amongst the most prevalent bacterial genera in freshwater bacterioplankton worldwide, with *Polynucleobacter paneuropaeus* being a particularly abundant member. The species seems to be relatively young, possibly less than 100 000 years old. Among 113 strains isolated from 54 lakes and ponds across Europe, 16S rRNA gene identity is >99.9% and ANI >96.5% [1]. Genetic coherence across populations is even more conspicuous: Reads from 38 combined metagenomes of samples spanning a 2700 km range mapped to one reference genome at an ANIr of 99.1% [2]. Yet, the phylogenetic coherence is opposed by vast diversification in the flexible genome, which counted 8746 different (<95% amino acid identity) genes, surpassing the 1370 different genes of the core genome more than 6-fold. Bacterial species seem to diversify their flexible genomes at an extraordinary pace. Is this a means to better exploit their environments through additional physiological capabilities, or a necessity to survive predation by the high diversity of phages surrounding them? Functional annotation in *Polynucleobacter paneuropaeus* suggests both. And the two themes present themselves in such distinct forms, we may do well to distinguish them when seeking to understand the ecological relevance of its flexible genome. More than 86% of the flexible gene pool was present in one or a few strains only. A large part of these “strain-specific” genes seemed to be involved in cell surface glycosylation, possibly decisive for phage recognition. They were clustered in a specific region of the genome, which was different in virtually all strains. When considering the cumulative flexible gene pool, the excessive share of strain-specific genes dwarfs the group of genes that is more commonly shared among individuals and provides additional functions. Yet, when considering individual genomes, these “additive” genes dominated the flexible genome, contributing 68% of genes on average. Respective gene cassettes, encoding functions such as anoxygenic photosynthesis, may tailor adaptation to abiotic factors. If strain-specific and additive genome and their different ecological roles are similarly well distinguishable in other species, consistent terminological distinction between these two themes of the flexible genome may help to clarify its relevance for microbial ecology and evolution.

Keywords: freshwater bacteria, horizontal gene transfer, strain-specific genome, additive genome

References:

1) Hoetzinger M, Pitt A, Huemer A, and Hahn M W (2021). Continental-scale gene flow prevents allopatric divergence of pelagic freshwater bacteria. Genome Biol Evol 13(3). doi:10.1093/gbe/evab019
2) Hoetzinger M, Hahn MW, Andersson LY, Buckley N, Ramsin C, Buck M, Nuy JK, Garcia SL, Puente-Sánchez F, Bertilsson S (2023). Geographic population structure and distinct population dynamics of globally abundant freshwater bacteria. bioRxiv. doi: https://doi.org/10.1101/2023.07.13.548520.

Strain-level Variation alters Host-Symbiont Interaction Dynamics in the *Halorubrum lacusprofundi* – *Candidatus* Nanohaloarchaeum antarcticus system

Joshua Hamm, Anja Spang

Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, P.O. Box 59, NL-1790 AB Den Burg, The Netherlands.

Joshua.hamm@nioz.nl

Candidatus Nanohaloarchaeum antarcticus is an obligate symbiont of *Halorubrum lacusprofundi* and requires direct cell-cell interactions with its host for survival [1]. In the natural environment these two organisms display contrasting degrees of population heterogeneity; whilst populations of *Ca*. Nha. antarcticus display almost no genetic variation those of *Hrr*. *lacusprofundi* vary significantly, particularly with regards to secondary replicons [1,2]. Much of this variation is found within genes for cell surface structures predicted to impact interactions between *Hrr*. *lacusprofundi* and its viruses[2] which could also affect interactions with *Ca*. Nha. antarcticus. To assess the impact of *Hrr*. *lacusprofundi* genetic variation on interactions with *Ca*. Nha. antarcticus, we performed cultivation experiments investigating the capacity of a set of 24 *Hrr*. *lacusprofundi* strains isolated from hypersaline Antarctic lakes to support growth of *Ca*. Nha. antarcticus. In addition to cultivation, we carried out comparative genomics analyses to determine the level of genomic variation between these *Hrr*. *lacusprofundi* strains and identify variations that may affect the response of host strains to *Ca*. Nha. antarcticus. Our results indicate a variable susceptibility of *Hrr. lacusprofundi* strains to exploitation by *Ca.* Nha. antarcticus as a result of a suite of secondary replicons that are also likely to impact interactions with viruses and nutrient scavenging within the population.

Keywords: Nanohaloarchaeota, Symbioses, Genomics

References:

1) Hamm, J. N. et al. (2019) Unexpected host dependency of Antarctic Nanohaloarchaeota. Proc Natl Acad Sci U S A 116, 14661-14670, doi:10.1073/pnas.1905179116
2) Tschitschko, B. et al. (2018) Genomic variation and biogeography of Antarctic haloarchaea. Microbiome 6, doi:ARTN 113 10.1186/s40168-018-0495-3

O-antigen diversity in order Pelagibacterales

Jose M. Haro-Moreno, Mario López-Pérez and Francisco Rodriguez-Valera

Evolutionary Genomics Group, Dpto. Producción Vegetal y Microbiología, Universidad Miguel Hernández, Alicante, Spain

moreno.jose@umh.es

The diversity of the glycosidic envelope of prokaryotic cells in general has been a focus of microbiology for a long time. Classically it was attributed to the need to offer a wide diversity of antigenic specificity and avoid host immune systems. However, free living bacteria that never interact with immune systems display similar levels of variation. Here we have studied the order Pelagibacterales O-polysaccharide gene cluster to establish the degree of diversity and patterns of sharing among this diverse group of streamlined aquatic bacteria. We have used single-amplified genomes (SAGs) since metagenome-assembled genomes (MAGs) do not contain the flexible regions of the genomes. In Pelagibacter, there is only one major cluster of polysaccharide synthesising genes that happens to be located between the 16S rRNA-23SrRNA and the 5SrRNA genes that are very well conserved. We found 685 different (non-syntenic) gene clusters (81% of the total recovered sequences), a number comparable to the ones found within a collection of 27,334 genomes of Enterobacteriales[1]. The diversity found at a single location in BATS and within genomes belonging to the same genomospecies is remarkable. The genes also had a similar pattern of sharing indicating frequent exchange by HGT. The clusters seem to diverge with the rest of the genome indicating that exchanges are old and do not happen frequently but a few examples of “smoking guns” (nearly identical clusters) were found among SAGs, mostly those very close phylogenetically and more often than not from the same location. The results indicate that antigen diversity Is not due to interaction with the host but reflect virus predation as predicted by the CD hypothesis[2].

References:

1) Holt, K.E., Lassalle, F., Wyres, K.L. et al. Diversity and evolution of surface polysaccharide synthesis loci in Enterobacteriales. *ISME J*. 2020. 14: 1713–1730.
2) López-Pérez, M and Rodriguez-Valera, F. Pangenome evolution in the marine bacterium *Alteromonas. Genome Biology and Evolution*. 2016 May; 8(5): 1556–1570.

Why do Alpha-Cyanobacteria with Form 1A Rubisco Dominate Aquatic Habitats worldwide?

Pedro J. Cabello-Yeves, Puxty, Richard J., Scanlan, David J.

School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK

pedrito91vlc@gmail.com

RuBisCO is one of the most abundant enzymes on Earth and virtually all food webs depend on it to supply fixed carbon. Photosynthetic organisms have developed CO_2-concentrating mechanisms (CCMs) such as carboxysomes, of which two independent forms exist: alpha and beta. Thanks to these CCMs the cyanobacterial lineage has colonized habitats worldwide, being primary producers of great ecological importance. Amongst them, the genera Prochlorococcus/Synechococcus, the two most abundant photosynthetic taxa on Earth, dominate oceanic ecosystems. These marine picocyanobacteria possess a form IA RuBisCO and α-carboxysomes (α-cyanobacteria). The remainder of the cyanobacterial radiation possess β-carboxysomes and form IB RuBisCOs (β-cyanobacteria), including freshwater unicellular and filamentous bloom-forming taxa comprising model organisms used in laboratories worldwide e.g. Microcystis, S. elongatus and Synechocystis. However, we isolated and sequenced the genomes of new abundant and widespread unicellular freshwater picocyanobacteria that are phylogenetically closer to their marine counterparts which fall inside the well-known cluster 5 picocyanobacterial clade. What distinguishes these marine and freshwater microbes? Overall larger genomes (≈2.9 Mb), %GC content (≈64%) and flexible genomes (ca. 50% of total genomic repertoire) have been observed in freshwaters compared to marine (2.5 Mb, 58.5 GC% and 33% of genes accounting to the flexible genome) isolates. Genomic novelties/differences across the salinity divide highlighted acidic proteomes and specific salt adaptation pathways in marine isolates, while freshwater strains possessed distinct ion/potassium channels, permeases, fatty acid desaturases, and more neutral/basic proteomes. However, regardless of their origin, they all possess form IA RuBisCOs and α-carboxysomes among their shared genomic features, which eliminates salinity as the main driver of the acquisition of the alpha carboxysome. Why then two forms of carboxysomes exist and why do α-cyanobacteria dominate aquatic habitats? With a combination of genomics/bioinformatics, genetic engineering, photophysiology of α/β representatives and biochemistry of RuBisCO forms IA/IB we hypothesize that marine/freshwater α-cyanobacteria dominate large, stable water masses, characterized by well-buffered pHs and relatively slow changes in carbonate chemistry. This will be crucial for accurately predicting the biosphere's response to changing CO_2, Ocean's acidification, pH and carbonate chemistry and developing biotechnological applications aimed at improving plant growth.

Session 5.

Ecology of the Pangenome II

A role for local adaptation in the microbial pangenome

Rachel J. Whitaker[1], Armun Liaghat[2], Mercedes Pascual[2], Jiayue Yang[1], Ruben Sanchez-Nieves[1] and Changyi Zhang[1]

[1] Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana, IL 61801. [2] New York University, New York, NY

rwhitakr@illinois.edu

Much of the microbial pangenome content and dynamics are defined by active symbiotic interactions between mobile genetic elements and their microbial hosts. The co-evolutionary trajectories of these symbiotic interactions critically depend upon the biotic and abiotic context in which they evolve. Therefore, we argue that the local microbial pangenome will be composed of traits that are locally adaptive to both microbial hosts and their genetic symbionts. We present evidence for local adaptation between *Sulfolobus islandicus* and its viruses from two isolated populations in Kamchatka, Russia and Yellowstone National Park and discuss implications for the local pangenome dynamics.

Keywords: Local adaptation, Archaea, Sulfolobus, Virus, Plasmid, CRISPR, Genomics

Population-level Genome Dynamics of Archaea in Hot Spring Microbial Communities

José R. de la Torre

Department of Biology, San Francisco State University, San Francisco, CA, USA
jdelator@sfsu.edu

Understanding how microbial populations adapt and evolve is critical to understanding the complex role that microbes play in our world. However, many microbial population genetics studies include cultivation biases or have focused on fast-growing, heterotrophic organisms. Here we have taken a cultivation-independent, metagenome-based approach to investigate the population-level genetic diversity of naturally occurring populations of chemolithoautotrophic ammonia-oxidizing archaea (AOA). AOA are among the most widely-distributed organisms on the planet and play a key role in driving the global nitrogen cycle. Although first discovered in marine environments, they are also found across a wide variety of aquatic (marine and freshwater) and terrestrial habitats, as well as associated with a number of vertebrates and invertebrates. Phylogenomic analyses indicate that the earliest diverging lineages are found in terrestrial hydrothermal systems. The presence of these thermophilic AOA (ThAOA) in physically isolated hot springs make them a good model system to investigate intrapopulation dynamics because of the strong barriers to gene flow and migration in these environments. To investigate patterns of microdiversity of ThAOA in naturally-occurring populations, we recruited reads from hot spring metagenomes from Europe, North America and Asia to a variety of ThAOA reference genomes. Our analyses have found large variations in the amount of genetic diversity found across sites. Some of ThAOA populations showed large amounts of nucleotide diversity, while others were practically clonal. Across all spring, the vast majority of genes, were under purifying selection. This pattern held true for genes in both the core and variable genome, and was also true across functional categories. In addition, the observed recombination frequency was considered low-moderate when compared to previously characterized microbial populations. This suggests that ThAOA live under genome-wide purifying selective pressures with few, but select, horizontal gene transfer events. This work has begun to reveal the population dynamics of the ecologically essential ThAOA, displaying a genome-wide selective pressure that appears stable across temporal and spatial scales.

Keywords: archaea, hot spring, thermophilic, ammonia-oxidizing, microdiversity

A Deep Dive into the Genomic Diversity of Marine Archaea and Bacteria

Felipe H Coutinho[1], Pablo Sánchez[1], Pierre Ramond[1], Marta Sebastián[1], Pablo Sánchez[1], Carlos M. Duarte[2], Dolors M Vaqué[1], Josep M Gasol[1], Silvia G Acinas[1]

1 - Department of Marine Biology and Oceanography, Institut de Ciències del Mar (ICM), CSIC, Barcelona, Spain. 2 - Red Sea Research Centre (RSRC) and Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia.

fhernandes@icm.csic.es, sacinas@icm.csic.es

Environmental gradients structure the structure of marine communities of Archaea and Bacteria, as well as the composition of their genomes. We sought to determine how the depth gradient influences the genome size, metabolic richness, coding density, and prevalence of prophages and molecular defense systems in dominant marine Archaea and Bacteria, throughout the epipelagic, mesopelagic, and bathypelagic zones. Thus, we generated 76 metagenomes across eleven sites at tropical and subtropical regions, sampled from the surface down to 4,000 m deep, during the Malapsina expedition. A total of 1,228 medium and high-quality, non-redundant, metagenome assembled genomes were obtained. These MAGs belonged to 22 bacterial and 5 archaeal phyla, and represented both ubiquitous and abundant, as well as rare low abundance taxa of the marine microbiome. A total of 509 MAGs genomes were derived from new species, one of which represented a new order within the class Planctomycetes. Genomes of Alphaproteobacteria, Gammaproteobacteria, Bacteroidota, and Actinobacteriota, tended to increase in size from the epipelagic towards the bathypelagic zone. Meanwhile, MAGs derived from Marinisomatota and Chloroflexota displayed a bimodal distribution with peaks of genome length at the mesopelagic and bathypelagic. MAGs derived from SAR324 and Thermoproteota were notable exceptions to this pattern, as the first tended to display the largest genomes at the epipelagic, while the latter displayed no association between genome size and depth. Overall, the larger genomes from deeper layers displayed lower coding density, and higher metabolic richness, prevalence of signal transduction genes, prophages, and defense systems, although the strength of the associations between these variables and depth varied significantly among taxa and zones. These trends were associated with a shift from a predominantly free-living to particle-attached lifestyle throughout the depth gradient, accompanied by a decrease in the availability of organic matter. We demonstrated that genome streamlining favors oligotrophic, free-living, epipelagic photoautotrophs and photoheterotrophs, as well as mesopelagic chemolitoautotrophs, while genome enlargement benefits copiotrophic, particle-attached, chemoheterotrophs at all zones. These results demonstrated how the ecological niches occupied by marine Archaea and Bacteria drive the composition of their genomes, which has important implications for the understanding of their ecology and evolution.

Keywords: deep ocean; genome evolution; metabolic richness; prophages; molecular defense systems

Analysis of the Prokaryotic Genome Evolution within a 5.5 million years trapped Chemolithotrophic Cave

Zaki Saati-Santamaría[1,2,3], Paula García-Fraile[2,3,4], Alena Nováková[3], Lorena Carro[1,2], Daniel Morais[5,6], Tomáš Větrovský[3], Miroslav Kolarik[1]

[1] Institute of Microbiology of the Czech Academy of Sciences, Vídeňská, Prague, Czech Republic. [2] Departamento de Microbiología y Genética, Universidad de Salamanca, Salamanca, Spain. [3] Institute for Agribiotechnology Research (CIALE), Villamayor, Salamanca, Spain. [4] Associated Unit USAL-CSIC (IRNASA), Salamanca, Spain. [5] Biological Institute of São Paulo – Vila Mariana, Av. Conselheiro Rodrigues Alves 1252, São Paulo, SP 04014-002, Brazil. [6] Norwegian College of Fishery Science, UiT the Arctic University of Norway, Muninbakken 21, Tromsø 9019, Norway

zakisaati@usal.es

The study of the prokaryotic pangenome evolution is of utmost importance to understand key processes in the ecology of microbes. There are hundreds of pangenome studies among taxa found in different ecosystems, which helps to deduce genomic divergences linked with some niches' occupancies or with phenotypic traits. However, despite its importance, there are not many opportunities to unveil the pangenome evolution with innovative approaches.
The Movile Cave (Romania) represent a unique chemolithotrophic environment to study genome evolution events, since it has been sealed for ~5.5 million years from the outer life, outer microbiomes and climatic shifts[1]. Therefore, it is arguably that life in this cave should have been evolved differentially than that out of the cave. Indeed, since this cave was discovered in 1986, tens of endemic novel species (animal invertebrates, bacteria, fungi, etc) have been found in there.

Here we used the prokaryotic DNA from several locations of the Movile Cave as an innovative model to study the evolution of microbes, based on (I) that the cave's microbiota have been shaped by distinct evolutionary forces, and (II) that the long exclusion of their microbial taxa from the outer genetic material serves as a resource to understand the time effect on prokaryotic evolution. We were able to assemble several high quality metagenome-assembled-genomes (MAGs) and we analyze their genomic innovations in comparison with genomes and MAGs from closely related taxa found out of this cave. We found that our MAGs have suffered a high divergence in gene content. We also found that there has not been many horizontal gene transfer events within the microbiomes. Interestingly, we found that the virome of this habitat carry auxiliary metabolic genes that may impact in the outcomes of the microbiome functionalities. Finally, we described some *Candidatus* novel taxa found within this cave based on the MAG data.

In sum, our study provides a novel insight into the (pan)genome evolution of prokaryotes. Overall, we found that 5.5MY is not sufficient time to have a very deep impact in the evolution of prokaryotes, likely due to the absence of selective pressures.

Keywords: Genome evolution; Horizontal Gene Transfer; Comparative genomics; Chemolithotrophs' pangenomes
References:
1) Kumaresan et al. (2014). Microbiology of Movile Cave—a chemolithoautotrophic ecosystem. Geomicrobiology Journal, https://doi.org/10.1080/01490451.2013.839764

Session 6.

Dynamics of the Pangenome

Quantifying Genetic Variation at the edge of Metagenomic Resolution to characterize *in situ* Population Dynamics

Luis M Rodriguez-R[1], Esteban Bustos Caparros[2], Alexandra Meziti[3,5], Despina Tsementzi[4,5], Ramón Rossello-Mora[2], Kostas T Konstantinidis[5]

[1] Department of Microbiology and Digital Science Center, University of Innsbruck, Innsbruck, Austria. [2] Marine Microbiology Group, Department of Animal and Microbial Biodiversity, Mediterranean Institutes for Advanced Studies (IMEDEA, CSIC-UIB), Esporles, Spain. [3] Department of Marine Sciences, University of the Aegean, Myrtilini, Greece [4] Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta GA, USA. [5] School of Civil and Environmental Engineering and School of Biological Sciences, Georgia Institute of Technology, Atlanta GA, USA

miguel.rodriguez@uibk.ac.at

The genetic variation that accumulates in natural populations (*i.e.*, members of the same species in a given community) is typically termed microdiversity. While it is well documented that natural prokaryotic populations exhibit genomic variation, a systematic framework for the quantification of microdiversity from metagenomic methods is lacking. In this lecture I will describe a conceptual framework for the quantification of microdiversity that explicitly distinguishes between gene-content and allelic diversity, and present computational methods that push the boundaries of what is quantifiable through metagenomics for individual populations. I will showcase the approaches in two aquatic environments: a natural riverine system in Kalamas River (Greece) and an experimental setup exploring adaptations of halophilic prokaryotes to solar salterns. Importantly, I will present evidence on the statistical limitations of current approaches for the quantification of microdiversity from metagenomes as well as potential computational solutions.

Keywords: Microdiversity, Metagenomics, Environmental Genomics, Natural Populations

Gene-Gene Associations and the Role of Selection in Microbial Pangenomes

Fiona J Whelan

School of Life Sciences, University of Nottingham, Nottingham, United Kingdom

fiona.whelan@nottingham.ac.uk

The variability between strains of evolutionary-related microbes can have important implications on a bacteria's mechanisms of self-defence, virulence, and – more broadly – on the ability to co-exist as part of mixed microbial communities. Bacteria are unique in the rate with which they undergo horizontal gene transfer (HGT), leading to some genes being acquired independent of lineage. We hypothesize that the variability in gene content observed in related bacterial strains is maintained by selection and thus that gene gain and loss is not random but instead reflects functional patterns within the genome. If this is the case, we might expect sets of genes to be gained and/or lost together across a set of strains which would manifest as patterns of gene-gene co-occurrence (i.e., association) and/or avoidance (i.e., dissociation). To test this, we developed software called Coinfinder which can detect sets of genes present more than we expect by chance across a set of genomes. Using *Pseudomonas sp.* as a case study, we find that the vast majority of abundant accessory genes with signatures of HGT are involved in gene-gene co-occurrence relationships, indicating that the strain-to-strain variability within *Pseudomonas sp.* (and other microbes) is most likely a result of natural selection. Similarly, when we examine the longitudinal variability of *P. aeruginosa* in the airways of individuals with cystic fibrosis, we find ample genetic diversity in the population which fluctuates over the course of pulmonary exacerbation, further indicating the importance of genetic diversity in mixed microbial communities.

Keywords: pangenomes, gene-gene associations, coinfinder, *Pseudomonas aeruginosa*, cystic fibrosis

Common Misconceptions when Disentangling the Processes that Shape Pangenomes

Anne Kupczok[1], Christian Resl[2], Hannah Götsch[3,4], **Franz Baumdicker**[2,4,5]

[1] Bioinformatics Group, Wageningen University & Research, Netherlands. [2] Cluster of Excellence "Controlling Microbes to Fight Infections", Mathematical and Computational Population Genetics, University of Tübingen, Germany. [3] Faculty of Mathematics, University of Vienna, 1090 Vienna, Austria. [4] Vienna Graduate School of Population Genetics, Austria. [5] Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, Germany

franz.baumdicker@uni-tuebingen.de

Most pangenomes contain many accessory genes of low and intermediate frequencies. Different population genetics processes contribute to the shape and composition of these pangenomes, namely selection and fitness-independent processes such as gene transfer, gene loss, and migration. However, the relative importance of these processes is highly debated.
Here, we first argue that the debate around prokaryotic pangenomes arose due to the imprecise application of population genetics models [1]. Most importantly, two different processes of horizontal gene transfer act on prokaryotic populations, which are frequently confused, despite their fundamentally different behavior. Gene transfer that acquires genes from distantly related organisms, termed here acquiring gene transfer (AGT), introduces new variants into the populations and has thus an effect comparable to the process of mutation in nucleotide sequences. In contrast, gene gain by transfer within the population, termed here spreading gene transfer (SGT), has an effect on gene frequencies that is identical to the effect of positive selection on single genes. In this talk, we show that selection and fitness-independent population genetic processes affecting pangenomes can be indistinguishable at the level of single gene dynamics.

Nevertheless, when considering the joint distribution of all accessory genes across individuals of a population, population genetics processes are fundamentally different. To understand to which degree the different processes shape pangenome diversity, we propose that the development of comprehensive models and simulation tools is mandatory. Furthermore, we need to identify summary statistics and measurable features that can distinguish between the processes to analyze the joint distribution of accessory genes, which, in turn, introduces other potential pitfalls that require careful attention.

In the second part of the talk, we present a phylogenetic method to distinguish between SGT and AGT alongside some of these pitfalls. Particularly, we highlight the effect of sampling bias and genome-wide linkage, two commonly overlooked aspects of pangenome analysis that can lead to misinterpretations of gene co-occurrence patterns, gene frequency spectra, and pangenome accumulation curves. Finally, by using simulations based on the parameters in the real data, we can reliably estimate the relative importance of SGT and AGT in a pangenome.

Keywords: pangenome evolution, population genetics, horizontal gene transfer, gene co-occurrence, sampling bias

References:

1) Franz Baumdicker and Anne Kupczok. (2023). Tackling the Pangenome Dilemma Requires the Concerted Analysis of Multiple Population Genetic Processes. Genome Biology and Evolution. DOI: https://doi.org/10.1093/gbe/evad067

Regions of Genome Plasticity: a Comparative Analysis of Mobility and Retention Patterns in Bacterial and Archaeal Pangenomes

João Botelho[1], **Mario Rodríguez-Mestre**[2], Jaime Iranzo[1]

1 Centro de Biotecnología y Genómica de Plantas (CBGP), Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC), Madrid, Spain. 2 Section of Microbiology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark.

joao.tbotelho@upm.es

Genome plasticity is a hallmark of bacterial and archaeal populations. Understanding the underlying mechanisms is crucial for deciphering their adaptation to diverse environments and response to selective pressures. Pangenomes contain regions of genome plasticity (RGPs) that can differ even among closely related strains. These regions harbor genes responsible for various phenotypic traits like virulence, complex molecule biosynthesis, antibiotic resistance, phage resistance, and other stress responses. Therefore, studying RGPs is essential for understanding the evolution and adaptation of bacterial and archaeal populations, with potential applications in clinical microbiology and biotechnology.

In this ongoing project, we analyzed RGPs in bacterial and archaeal genomes and investigated correlations between different types of co-transferred genes. Our analysis followed a two-stage approach. Firstly, we used the PPanGGOLiN software to identify RGPs based on the presence of contiguous accessory genes and clustered them into families. Subsequently, we performed a comparative analysis of the identified RGPs, with a focus on compatibility between different RGPs and co-transfer of genes.

Our study encompassed 90,000 high-quality genomes from 500 prokaryotic species, resulting in the identification of approximately 4 million RGPs. We examined the conservation of these RGPs across taxonomic levels and explored co-occurrence patterns, which shed light on inter-genomic conflicts and cooperation. Additionally, we assessed the mobility of functional modules within RGPs, including defense systems, antibiotic resistance cassettes, and biosynthetic clusters. To achieve this, we developed a scoring system based on gene distances and taxonomic markers, and analyzed the presence of regulator genes that could account for the co-location of these functional modules.

Furthermore, we employed tools capable of detecting remote homology relationships and investigated structural and functional connections between proteins typically found in RGPs. Our approach utilized a deep profile clustering strategy with relaxed sequence identity and coverage thresholds, along with protein language-model embeddings comparisons, to achieve high sensitivity and identify remote homologs. The results of this study provide insights into the evolutionary dynamics and functional significance of RGPs in bacterial and archaeal pangenomes. Additionally, the resulting database of remote homologs in RGPs serves as a valuable resource for the scientific community.

Keywords: Regions of genome plasticity, mobility, co-transferred genes, functional modules, remote homologs

Session 7.

Pangenomes of animal / plant denizens

Variation and Specialization of Structure and Function of Gut Microbial Community Across Host Species

Aiswarya Prasad[1*], Asha Devi Pallujam[2], Rajath Siddaganga[3], Ashwin Suryanarayanan[3], Axel Brockmann[3], Sze Huei Yek[2,4], Philipp Engel[1]

[1]University of Lausanne, Switzerland, [2]Monash University, Malaysia, [3]National Centre for Biological Sciences, Bangalore, India, [4]Universiti Malaysia Sabah

aiswarya.prasad@unil.ch

Honeybees harbor specialized microbial communities in their gut, which makes them exciting models for understanding the evolution of gut microbiota-host interactions across animals. Metagenomic studies of the well-known *Apis mellifera* and *Apis cerana* honeybees have revealed that different host species harbor discreet microbial populations within genera and a vast difference in genetic and functional diversity [1]. While the western honeybee *A. mellifera* gut microbiome is well-studied, little is known about the communities in the gut of wild honeybee species mainly found in Asia. In particular, the dearth of genomic data has limited our understanding of the functional potential and evolution of gut microbes across honeybees. With the help of collaborators in Malaysia and South India, we sampled and shotgun sequenced the gut microbiota of 200 individual honeybees sampled from six to nine colonies of five host species (A. mellifera, A. cerana, A. dorsata, A. florea, A. andreniformis). Thus, we recovered >500 high-quality metagenome-assembled genomes (MAGs), which we clustered into magOTUs (with >95% average nucleotide identity) to obtain high-resolution insight into the taxonomic, phylogenetic and functional diversity of the honeybee gut microbiota across honeybee species. We find that the diversity of the microbial community in terms of magOTUs varies across host species. Distinct magOTUs from the same genus were found in different host species. A few magOTUs are found prevalently across host species, while others are found uniquely in particular host species independent of geography. Gene flow analyses suggested host-specific pangenomes even within magOTUs. While MAGs of the same genus share a core set of orthologous genes, distinct orthologous groups are often associated with particular magOTUs. We are further exploring the functional gene content across groups using a community-level approach informed by the metagenomic binning process. So far, we find that some predominant magOTUs not formerly known to be important core microbiota members of social bees contain diverse arsenals of pectinase, hemicellulase and cellulase genes important for the breakdown and digestion of the recalcitrant pollen diet of honeybees. Further analysis will reveal how the structure and functions in the pangenomes of gut microbes have evolved across host species.

Keywords: Metagenomics, honeybee gut microbiome, host-associated, Gene functional potential, pangenome

References:

1) Ellegaard, K. M., Suenami, S., Miyazaki, R., & Engel, P. (2020). Vast differences in strain-level diversity in the gut microbiota of two closely related honey bee species. *Current Biology*, *30*(13), 2520-2531. https://doi.org/10.1016/j.cub.2020.04.070

High-order Metabolic Interdependencies Dominate the Human Gut Microbiome

Jaime Iranzo[1,2*], Adrián Aguirre-Tamaral[1], Jorge Calle-Espinosa[1]

[1] Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM) – Instituto Nacional de Técnica y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain. [2] Institute for Biocomputation and Physics of Complex Systems (BIFI), Universidad de Zaragoza, Zaragoza, Spain

jaime.iranzo@upm.es

Over the last decade, advances in metagenomics have revolutionized the study of microbial communities. Large-scale initiatives, like the Earth Microbiome Project, have shed light on the composition and functional diversity of complex microbiomes associated with different environments and hosts. However, identifying cross-species interactions remains a major challenge in microbial ecology. Because interactions determine key properties of microbial communities, such as stability and resilience to perturbations, the lack of robust methods to infer ecological interactions limits our ability to develop targeted interventions for microbiome management.

In this work, we tested two approaches to predict metabolic interactions in the human gut microbiome. Starting from public data from the Integrative Human Microbiome Project, we first mapped the taxonomic profiles of 132 microbiomes to curated genome-scale metabolic models (GSMM) from the AGORA2 database [1]. Alternatively, we generated de novo GSMM for genomic bins directly assembled from the metagenomic reads of each microbiome. Despite the uncertainty associated with assembly and binning, de novo GSMM arguably produce more accurate representations of the metabolic capabilities present in the local pangenome, which may not coincide with those of reference genomes. Finally, we assessed the potential for cross-species metabolic cooperation with the software SMETANA [2], considering a panel of standard culture media and representative nutrient profiles for the human gut.

We found a high potential for metabolic complementation in the gut microbiome, that involves not only pairwise cross-feeding, but also higher order interdependencies. Metabolic complementation often requires the joint participation of 3-5 species. This observation is qualitatively robust with respect to the media composition, although the exact interaction profiles depend on nutrient availability. The median number of interdependent species is smaller if using de novo GSMM, which stresses the importance of considering locally adapted pangenomes when inferring metabolic interdependencies.

Our results have major implications for the description of microbial communities, suggesting that high-order hypergraphs provide better representations of ecological interdependencies than classical mutualistic networks. Remarkably, the introduction of such structures in models of population dynamics can result in the stable maintenance of species-rich communities. Thus, high-order metabolic interdependencies may contribute to explaining the high microbial diversity observed in the human gut.

Keywords: microbiome, cross-feeding, ecological network, flux balance analysis, metabolic modeling
References:

1) Heinken et al. (2023). Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. Nat. Biotechnol. DOI: 10.1038/s41587-022-01628-0.
2) Zelezniak et a. (2015). Metabolic dependencies drive species co-occurrence in diverse microbial communities. Proc. Natl. Acad. Sci. USA. DOI: 10.1073/pnas.1421834112.

Large-scale Investigation of Orphan Genes in the Human Gut Microbiome Elucidates their Evolutionary Origins

Nikolaos Vakirlis[1,2], **Anne Kupczok**[3*]

[1] Institute For Fundamental Biomedical Research, B.S.R.C. "Alexander Fleming", Vari, Greece. [2] Institute for General Microbiology, Kiel University, Kiel, Germany. [3] Bioinformatics Group, Wageningen University, Wageningen, The Netherlands

anne.kupczok@wur.nl

Orphan genes, i.e., genes that lack homologs outside a given species, are ubiquitous in all domains of life. In prokaryotes, orphans can originate (i) in the native genome via de novo evolution from non-genic regions or alternative frames of existing genes, or by rapid divergence and remodeling, or (ii) in a foreign genome, including viruses, followed by horizontal transfer. However, strong quantitative evidence supporting either scenario is lacking.

Here we performed a systematic, large-scale analysis of orphan genes from human gut prokaryotes[1]. After exhaustive filtering, we identified more than 3 million orphans in 4,644 species, which lack similarity to other prokaryotes and have no known functional domains. We find that a given species pangenome contains on average 2.6% orphan genes, which are mostly rare within a species. Overall, orphan genes use optimal codons less frequently, and their proteins are more disordered than those of conserved (i.e., non-orphan) genes. Importantly, the GC content of orphan genes in a given genome closely matches that of conserved ones. In contrast, the 5% of orphans that share similarity to known viral sequences have distinct characteristics that set them apart from the rest of the orphans, including lower GC content. By identifying the genomic region from which they evolved in closely related species, we provide evidence for native origination for a small subset of orphan genes and find that these orphans also differ in their properties from the remaining orphans. Finally, predicting orphan function by examining functional annotations in operon-like arrangements suggests that some orphan genes are membrane-related and involved in spore germination.

Our results support that orphans emerge due to multiple routes, challenging the notion that external elements such as phages and plasmids are the primary source of prokaryotic genetic novelty. Importantly, origination in the native genome might provide a constant inflow of mostly transient genes into the cloud genome of prokaryotic pangenomes, where some orphans may prove adaptive, facilitating evolutionary innovation.

Keywords: Orphan genes, Lateral gene transfer, de novo gene evolution, human gut microbiome, cloud pangenome

References:

1) Almeida et al. (2015). A unified catalog of 204,938 reference genomes from the human gut microbiome. Nature Biotechnology. doi:10.1038/s41587-020-0603-3.

The Microvirome: Expanding the Plant Holobiont

Sheila Roitman and Detlef Weigel

Max Planck Institute for Biology, Tuebingen, Germany 72076
sheilaroitman@tuebingen.mpg.de

Eukaryotic organisms harbor large communities of microorganisms forming an holobiont, considered to be a single ecological and evolutionary unit. In recent years, bacterial community dynamics and their effect on the plant holobiont have been the subject of many studies. In spite of this, little is known regarding the role that bacteriophages play in shaping those bacterial communities. In my work I intend to set the basis for understanding the role of the microvirome in plant colonization and development, by studying Arabidopsis thaliana associated bacteria and phages, in laboratory and natural settings. Following a multilevel approach, I expect to gain a mechanistic understanding of the way phages affect plant-associated bacterial communities, deepening our basic understanding of the plant holobiont, and phage-host interactions in an oligotrophic environment. These findings can be projected to other significant plant-microbes systems, and be the foundation to design phage-based solutions to pest management in agriculture.

Keywords: Bacteriophages, Plant microbiome, Arabidopsis thaliana.

Session 8.

Phages and Pangenome

The Impact of Bacteriophages and PICIs on Bacterial Evolution and Virulence

Jose Penadés

Imperial College London, UK

j.penades@imperial.ac.uk

Bacteriophage-mediated transduction plays a pivotal role in driving bacterial evolution. Bacteriophages possess the ability to efficiently mobilise large segments of the bacterial chromosome through a process known as lateral transduction (LT). More recently, our research has unveiled an additional mechanism of chromosomal gene transfer facilitated by a significant group of pathogenicity islands called phage-inducible chromosomal islands (PICIs). This mechanism, termed PICI lateral co-transduction (LcT), also contributes to the transfer of chromosomal genes. Our findings suggest that LcT may exhibit unique mechanistic characteristics, potentially making it more frequent than LT in promoting high-frequency gene transfer. This, in turn, could have a profound impact on bacterial evolution. During this presentation, we will delve into the influence of both LT and LcT on the structural aspects of bacterial genomes. Furthermore, we will explore their implications in the evolution of virulence and antimicrobial resistance. Our research results anticipate unprecedented roles for phages and PICIs in the emergence of novel bacterial clones, shedding light on previously uncharted territories in bacterial evolution.

Viewing Co-Evolutionary Dynamics through the Microbial Pangenome: a Multi-level Analysis of Phage Receptor-Binding Proteins

Rafal J. Mostowy

Malopolska Centre of Biotechnology, Jagiellonian University in Krakow, Poland
rafal.mostowy@uj.edu.pl

Bacteria and their viruses have been co-evolving for billions of years. Our group aims to deepen the understanding of this process by focusing on the highly specific interactions between bacterial surface polysaccharides and phage receptor-binding proteins (RBPs). While our work contributes to the broader understanding of the prokaryotic pangenome by elucidating the phage component, we also posit that the pangenome serves as a valuable framework for studying these co-evolutionary processes. This approach allows us to probe highly diverse genetic regions in both bacteria and phages, offering a novel lens through which to examine their co-evolution.

Using a dataset of 133,574 representative phage proteins, we employed highly sensitive homology detection techniques to quantify instances of domain mosaicism. Our findings reveal that RBPs are hotspots of evolutionary innovation, often undergoing diversification via domain shuffling. This modularity transcends taxonomic and ecological boundaries, suggesting an ongoing co-evolutionary arms race with bacterial hosts.

To further delve into the structural diversity of RBPs, we employed AlphaFold2-based modelling. Our results indicate that structural similarity is a strong predictor of sugar specificity, offering a novel layer of understanding to phage-bacteria interactions. Additionally, we leveraged Genome-Wide Association Studies (GWAS) to identify novel RBPs specific to prevalent bacterial K-types, thereby expanding our understanding of phage host range.

Our multi-level approach to studying phage proteins not only provides fresh insights into the co-evolutionary dynamics between bacteria and phages but also has practical implications. By offering a deeper understanding of regions under strong diversifying selection, our work enables more accurate predictions of specific phage phenotypes, such as RBPs that recognise particular bacterial sugars. This advances us one step closer to predicting the host range of phages that infect bacteria considered major public health threats, like *Klebsiella pneumoniae*. Thus, our research serves as a cornerstone for both theoretical understanding and practical applications in microbial ecology, phage therapy, and public health.

Keywords: Polysaccharide capsule, horizontal gene transfer, Red Queen, structural modelling, genome-wide association studies

Bacteriophage Pangenomes from Metagenomes

Christopher Bellas

Department of Ecology, Universität Innsbruck, 6020 Innsbruck, Austria
christopher.bellas@uibk.ac.at

Virus Metagenomic Assembled Genomes (MAGs) offer insights into the diversity of viral populations. However, each MAG represents a composite assembly of variants from a virus's pangenome, meaning information about highly flexible genomic regions can be lost. By interrogating multiple bacteriophage communities across simple, analogous environments, phages with nearly identical core genomes can be repeatedly detected across large geographical distances or many years apart, revealing the individual variants and the underlying community structure. This shows that in a natural phage community, identical core genomes encode a number of flexible genes, including host recognition receptors and genes to counter bacterial defences in every location, as predicted by the constant diversity model. Notably, the individual phage variants are not fixed, as flexible genes from the pangenome are shuffled in each location by recombination, increasing diversity.

Keywords: metagenomics, pangenome, bacteriophage

Unveiling the Expansive Pangenome of Crassvirales: Novel CrAss-like Phage Isolates Reveal Genetic Heterogeneity and Worldwide Distribution

María Dolores Ramos-Barbero, Clara Gómez-Gómez, Laura Sala-Comorera, Lorena Rodríguez-Rubio, Sara Morales-Cortes, Elena Mendoza-Barberá, Gloria Vique, Daniel Toribio-Avedillo, Anicet R. Blanch, Elisenda Ballesté, Cristina Garcia-Aljaro, **Maite Muniesa**

Departament de Genètica, Microbiologia i Estadística. Universitat de Barcelona. Diagonal 643.
mmuniesa@ub.edu

Crassvirales, also known as crAss-like phages, represent a plentiful group of bacteriophages specific to the human gut, initially discovered through computational analysis. While there has been a proposal to employ crAss-like phages as indicators of human fecal presence, the limited success in cultivating only seven strains of these phages has impeded comprehensive research. In this study, we present a breakthrough in the isolation and genetic characterization of 25 novel crAss-like phages, designated as crAssBcn. These newly discovered phages infect Bacteroides intestinalis, falling under the taxonomical classification of Crassvirales, genus Kehishuvirus. Based on their genomic diversity, we have classified them into six distinct species. The genomes of crAssBcn phages exhibit similarities to crAss001, the first isolated virion, but notable genomic and amino acid distinctions set them apart from other crAss-like phages within the same family. Intriguingly, crAssBcn phages are more prevalent in fecal metagenomes worldwide compared to crAss001. This significant study expands the roster of known crAss-like phage isolates. The remarkable abundance and heterogeneity observed in these phages prompt a vital question: which member of the Crassvirales group should be chosen as the primary human fecal marker?

Keywords: Bacteriophage, CrAssphage, human virome, microbial source tracking, phage evolution

Closing keynote lecture

The Virosphere: Megataxonomy and Global Ecology

Eugene V. Koonin

National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, USA.

koonin@ncbi.nlm.nih.gov

All or nearly all organisms are hosts to multiple viruses that collectively appear to be the most abundant biological entities in the biosphere. With the recent advances in metagenomics and metatranscriptomics, the diversity of the known viruses and their gene repertoires dramatically expanded. Comparative analysis of these diverse viruses using advanced computational methods culminated in the reconstruction of the evolution of major groups of viruses and provided for the construction of virus 'megataxonomy' that has been formally adopted by the International Committee on Taxonomy of Viruses. This comprehensive taxonomy consists of 6 virus realms that are considered to be monophyletic based on the conservation of hallmark proteins involved in either capsid structure formation or genome replication and many kingdoms, phyla and classes comprising these realms. The major taxa of viruses substantially differ in their host ranges and accordingly in their ecological associations. I will outline the virus megataxonomy and the latest discoveries that will likely lead to reassessment of some of the major taxa. I will then review the correspondence between virus taxonomy and the distributions of viruses across hosts and ecological niches, as well as the abundance of viruses in different habitats.

17.00-19.30 FLASH TALKS I. 5 min. Chair: Joshua Hamm

Order	Author	Tittle
1	Puente Sánchez, Fernando	Turning metagenomics into metapangenomics: the next frontier in microbial ecology
2	Brouns, Roos Chiara Antoinette	Predicting the environment of microbial communities from community and pangenome data
3	Valiente Mullor, Carlos	Closing the pangenome: openness analysis of >40.000 bacterial genomes reveals the limits of Heaps law
4	von Meijenfeldt, Bastiaan	A large pangenome allows for a broad niche range
5	García-González, Neris	Improving core genome gene alignments for large-scale evolutionary and transmission studies
6	Rubio Portillo, Esther	What we can expect from pangenome analysis using MAGs
7	Walter Costa, Maria Beatriz	Predicting Salinity Preferences from Bacterial Genomes with Machine Learning
		TIME FOR QUESTIONS (10 min)
8	Enav, Hagay	SynTracker differentiates evolution via mutation or recombination on a per-species basis in complex microbiomes
9	Jean, Mainguy	PPanGGOLiN V2: technical enhancement and new features to analyze thousands of prokaryotic genomes
10	Arnoux, Jérôme	PANORAMA: comparative pangenomics tools to explore interspecies diversity of microbial genomes
11	Maistrenko, Oleksandr	Population genomics of Archaea
12	Feriel, Bouderka	Evolution of the Patescibacteria symbiotic lifestyle
13	Palacín Lizarbe, Carlos	Nitrate-reducing microbes in winter in sediments of large boreal lakes affected by browning and mining
14	Richy, Etienne	Exploring Microbial Processes and Genomic Islands in Deadwood Decomposition: Integrating Metagenomics and Metatranscriptomics
		TIME FOR QUESTIONS (10 min)
15	Pardeshi, Lakhansing Arun	Pangenomics to understand the emergence of a new *Pectobacterium brasiliense* pathovar
16	García-Fraile, Paula	Comparative enomics of the Pseudomonas pangenome reveals host- and environment-specific evolution
17	Suárez-Moo, Pablo	Analysis of the pangenome in Thaumarchaeota from the western Mediterranean sea in seasons with mixed and stratified water column
18	Albers, Justine	A global pangenome of marine nitrite-oxidizing bacteria highlights differences in nutrient acquisition strategies
19	Knudsen, Anne Kalinka Sand	Uncovering Methane Sink Potential and Habitat Preferences of Methanotrophs in Denmark
20	Vestergaard, Sofie Zacho	Genomic diversity of Zoogloea: Exploring secrets of biopolymer production in the world of activated sludge
21	Liébana Gracía, Raquel	Exopolysaccharide biosynthesis genes clusters in *Alteromonas*: a role in carbon remineralization and environmental adaptation
22	Zaragoza, Asier	A novel machine learning prediction tool for extracellular proteins doubles current size of the marine secretome
		TIME FOR QUESTIONS (10 min)

1.

Turning metagenomics into metapangenomics: the next frontier in microbial ecology

Fernando Puente, Matthias Hoetzinger[2], Moritz Buck[1], Stefan Bertilsson[1]

[1]Swedish University of Agricultural Sciences. [2]Norwegian University of Life Sciences

fernando.puente.sanchez@slu.se

Traditionally, studies on microbial ecology have been conducted by using species as the basic eco-evolutionary units. However, prokaryotic species have in fact access to a wide pool of genes (the pangenome) far greater that what can be fit into a single genome. Because of this, individuals from the same species can display strong variability in terms of both allelic composition and gene content. This intra-species diversity can substantially affect the dynamics of species and ecosystems, and hence elucidating its origin, extent and consequences is central to the future development of microbial ecology. In spite of this, its analysis is still far from being a common practice in microbial ecology studies. Here we introduce our ongoing efforts to bridge this gap, and bring the analysis of environmental pangenome dynamics (metapangenomics) to the general public.

First, we have developed SuperPang, an algorithm producing pangenome assemblies from a set of input genomes of varying quality [1]. This provides a modular view of the pangenome, producing a complete and non-redundant assembly that preserves synteny and contains both coding and non-coding regions. The assembly is partitioned into core and accessory regions, and can be used as a reference for mapping back metagenomic reads in order to track the dynamics of species, alleles and accessory genes across samples. Crucially, our method works with Metagenome-Assembled Genomes (MAGs) as the sole input, allowing us to build and study pangenomes directly from metagenomic datasets without the need for high quality reference genomes. Due to this, it can be integrated into existing metagenomic pipelines, allowing researchers to analyze their metagenomes with a pangenome-centric intra-species perspective.

We are currently expanding our SqueezeMeta pipeline [2] in order to make it suitable for the automated analysis of metapangenomes. With this, researchers without extensive bioinformatics experience can produce annotated pangenome assemblies from raw metagenomic reads in a single step process. We also provide an analysis library that allows easy exploration of the resulting datasets.

In this communication we aim to introduce our approach, illustrate its potential with real world examples, and gauge the needs of the community to guide future development.

2.

Predicting the environment of microbial communities from community and pangenome data

Rose Brouns, Maria Beatriz Walter Costa, Swapnil Doijad, Bas E. Dutilh

Viral Ecology and Omics Group (VEO), Cluster of Excellence Balance of the Microverse, Friedrich Schiller University Jena, 07745 Jena, Germany

rosebrouns.veo@gmail.com

Abiotic factors, including salinity, temperature, and pH, play a crucial role in shaping the microbial community structure. Salinity, for example, has been observed to stimulate the growth of specific species while inhibiting others. As microbial species exist within complex communities that engage in continuous interactions with the environment, we aim to understand the impact of abiotic factors on these communities. Considering that these abiotic factors are crucial in shaping microbial communities, we expect that conversely, they can be predicted based on community features. Being able to predict how microbial communities are shaped in relation to abiotic factors can impact microbiome research.

To explore this, we adopted "big data" methods. We collected data from the MGnify database, which included taxonomic composition, biome descriptions, and salinity levels from 1705 samples across 24 global study sites.

Using dimension reduction techniques, we visualized the data based on salinity and biome. Although some separation was observed, clear clusters did not form, indicating the complexity of microbial communities. Therefore, we continued with machine learning analysis which can capture more complex patterns in the data. Our Random Forest models could classify samples into high or low salinity groups based on community composition at different taxonomic ranks with a mean accuracy of 96-97%.

To cross-validate our results, we conducted differential abundance analysis (DAA), identifying taxa significantly differing between high and low saline environments. Comparing these results with the key taxa from the Random Forest models reinforced our findings.

In conclusion, this research uses machine learning, dimension reduction, and differential abundance analysis to investigate the relationship between taxonomic composition, salinity levels. We were able to predict salinity levels with high accuracy based on taxonomic composition alone, which is consistent with the results of DAA.

For next steps, we will expand our scope and try to predict other abiotic factors such as pH and temperature and use more input data such as the local pangenome. Within these results we will investigate unexpected patterns in search of new biological adaptations to the abiotic environment. Moreover, our research opens up the possibility to incorporate abiotic factors into predictive microbiome models.

3.

Closing the pangenome: openness analysis of >40.000 bacterial genomes reveals the limits of Heap's law

Carlos Valiente-Mullor[1*], Fernando González-Candelas[2]

[1] *Joint Research Unit Infection and Public Health FISABIO-Univ. Valencia, Institute for Integrative Systems Biology (I2SysBio, CSIC-UV). Valencia, Spain.* [2] *CIBER in Epidemiology and Public Health.*

carlos.valiente@uv.es

By definition, at any time point, the pangenome of a bacterial species is a closed set of genes. However, for a limited number of analyzed genomes, the fitted value of Heap's law exponent α [1] is used to classify the pangenome as open (α≤1), if its size (i.e., number of unique gene families) is expected to grow limitlessly, or closed (α>1), if it tends to an asymptotic plateau. We lack a gold standard implementation of this model and, thus, determining the openness state is affected by methodological decisions [2]. These involve: (a) the rarefaction curves (of total or new genes), (b) the processing of genome addition permutations (using centrality indicators or not), (c) the formula (exponential or linearized) used to fit the data and, ultimately, (d) the bioinformatic aspects of the implementation. In this work, we address these points by estimating openness in progressively larger sets of genomes using 6 different approaches for >70 pangenomes, with over 40,000 genome sequences, of pathogenic bacteria and simulated data with predefined α values. Firstly, we found that all pangenomes that are open when analyzing a small number of genomes increase their α when analyzing progressively larger samples. Secondly, we observe that the median increment of new genes is <2 in all pangenomes with >500 strains, despite being open according to α. Thirdly, we found that the different methods introduce errors to some extent in the estimation of α, which can lead to the same pangenome having different alpha values and openness states. Finally, we found systematic deviations from the power law model in rarefaction curves that bias the inference of α, particularly in large datasets and closed pangenomes. The evaluation of 5 additional models shows that most curves of new genes are described accurately by fitting slightly more complex models than Heap's law. These results emphasize the limitations of defining openness using a single parameter, particularly when studying large pangenomes that, although genetically diverse, approach the closed state that defines each bacterial species.

Keywords: comparative genomics, Heap's law, pangenome, bacterial pathogens, openness

References:

1) Tettelin, H. et al. (2008). Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol. 10.1016/j.mib.2008.09.006
2) Deng, X. et al (2010). Probing the pan-genome of Listeria monocytogenes: new insights into intraspecific niche expansion and genomic diversification. BMC Genom. 10.1186/1471-2164-11-500

4.

A large pangenome allows for a broad niche range

F.A. Bastiaan von Meijenfeldt[1,2,*], Paulien Hogeweg[1], and Bas E. Dutilh[1,3]

[1] Theoretical Biology and Bioinformatics, Department of Biology, Science for Life, Utrecht University, Utrecht, the Netherlands. [2] Department of Marine Microbiology and Biogeochemistry, NIOZ Royal Netherlands Institute for Sea Research, Texel, the Netherlands. [3] Institute of Biodiversity, Faculty of Biological Sciences, Cluster of Excellence Balance of the Microverse, Friedrich Schiller University Jena, Jena, Germany

bastiaan.von.meijenfeldt@nioz.nl

Generalists can survive in many environments, whereas specialists are restricted to a single environment. Although a classical concept in ecology, niche breadth has remained challenging to quantify for microorganisms because it depends on an objective definition of the environment. By defining the environment of a microorganism as the community it resides in, we recently integrated information from over 22,000 environmental sequencing samples to derive a quantitative measure of the niche, which we call social niche breadth[1]. At the level of genera, we explored niche range strategies throughout the prokaryotic tree of life. We found that social generalists include opportunists that stochastically dominate local communities, whereas social specialists are stable but low in abundance. We found no global correlation between social niche breadth and genome size. Instead, we observed two distinct evolutionary strategies, whereby specialists have relatively small genomes in habitats with low local diversity, but relatively large genomes in habitats with high local diversity. In contrast, social generalists from any habitat consistently have a more diverse and open pangenome than social specialists. Together, our analysis shines data-driven light on microbial niche range strategies.

References:
1) von Meijenfeldt, F.A.B., Hogeweg, P. & Dutilh, B.E. A social niche breadth score reveals niche range strategies of generalists and specialists. Nat Ecol Evol 7, 768–781 (2023). https://doi.org/10.1038/s41559-023-02027-7.

5.

Improving core genome gene alignments for large-scale evolutionary and transmission studies

Neris García-González, Fernando González-Candelas

[1] *Joint Research Unit Infection and Public Health FISABIO-Univ. Valencia, Institute for Integrative Systems Biology (I2SysBio, CSIC-UV). Valencia, Spain.* [2] *CIBER in Epidemiology and Public Health.*

nerisgg@gmail.com

The two main methodologies used in comparative microbial genomics are mapping sequencing reads to a reference genome and using pangenome tools to identify orthologous genes in de novo assemblies. However, pangenomes are seldom used for further and deeper studies such as transmission, epidemiological and recombination analyses, being mapping the gold standard methodology [1] despite its potential problem [2]. However, we have observed that popular pangenome tools yield inconsistent results, primarily due to the presence of erroneous variant calls in orthologous genes. These variants usually arise when the gene annotation software incorrectly identifies the start or end of a CDS in fragmented contigs, introducing frameshift errors. Frameshift errors lead to an accumulation of false single nucleotide polymorphisms (SNPs), profoundly impacting epidemiological and phylogenetic estimates, such as transmission clusters, dating analysis, or the identification of recombination events. Despite its relevance, this issue has received limited attention, and usual pan-genome analyses fail to address it adequately. Moreover, our attempts to mitigate these errors using various alignment trimming tools, such as Trimmal and CLipKit, did not yield improved results. To tackle this problem, we developed and implemented an algorithm designed to detect and remove affected sequence regions in individual gene alignments, resulting in error-free core genome sequence alignments. The algorithm involves constructing a phylogeny for each gene, identifying unexpectedly long branches, and verifying whether these are caused by SNP runs unique to that sample in the alignment. In our study, we applied this algorithm to investigate transmission groups in a collection of over 1600 clinical isolates, a critical antimicrobial resistant bacterial pathogen. By using pangenomes and our algorithm, we eliminated the bias associated with mapping to a common reference genome, thus facilitating the exploration of transmission dynamics across diverse lineages with higher accuracy and reliability. Our findings underscore the critical importance of addressing and correcting these errors, offering a valuable tool to understand the evolutionary, transmission, and epidemiological aspects of highly variable samples and diverse microorganisms.

Keywords: core-genome, transmission, genome alignment

References:

1) David, Sophia, Sandra Reuter, Simon R. Harris, Corinna Glasner, Theresa Feltwell, Silvia Argimon, Khalil Abudahab, et al. (2019). Epidemic of Carbapenem-Resistant Klebsiella Pneumoniae in Europe Is Driven by Nosocomial Spread. Nature Microbiology 4 (11): 1919–29. DOI: 10.1038/s41564-019-0492-8

2) Valiente-Mullor, Carlos, Beatriz Beamud, Iván Ansari, Carlos Francés-Cuesta, Neris García-González, Lorena Mejía, Paula Ruiz-Hueso, and Fernando González-Candelas. (2021). One Is Not Enough: On the Effects of Reference Genome for the Mapping and Subsequent Analyses of Short-Reads. PLoS Computational Biology 17 (1): e1008678. DOI: 10.1371/journal.pcbi.1008678

6.

What we can expect from pangenome analysis using MAGs

Esther Rubio-Portillo[1], Josefa Antón[1], Borja Aldeguer-Riquelme[12]

1 Dpto. Physiology, Genetics and Microbiology, University of Alicante. 2 School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, US

esther.portillo@ua.es

In a pan-genome, genes are classified into two categories: core and accessory. Core genes are expected to be shared by all or almost strains (>90%) within a species, while accessory genes are present in only a subset of strains[1]. The recovery of metagenome-assembled genomes (MAGs), or genome binning, from metagenome shotgun sequencing reads has become a common practice to recover individual genomes within a natural environment. These MAGs are increasingly utilized in pangenome analysis to investigate the genetic diversity among related microbial strains. Such analyses have significant applications in various study fields, including species evolution, pathogenic mechanisms, niche specialization, and population structures. To which extent accessory and core genes bin together during MAG recovery remains uncertain. Recent research has revealed that MAGs with 95% completeness captured only 77% of core genes and 50% of accessory genes, indicating that the quality of MAGs often falls short of expectations[2].

To assess the biases introduced during MAG recovery, we conducted a study using a natural seawater metagenome mixed with two artificial mock communities. The first mock community comprised three strains from seven *Vibrio* species, while the second mock community consisted of 30 strains from a single *Vibrio* species (*Vibrio coralliilyticus*). Our observations clearly demonstrated that MAGs can recover more than 90% of core genes and between 10% and 90% of accessory genes, depending on intraspecific diversity. Moreover, our results highlight that MAGs can include unique genes from different strains and thus, they may represent ecotypes rather than strains. This finding suggests that strategy of co-binning several metagenomes from different environments, commonly used to increase MAG completeness, also increases the probability of generating artefactual MAGs with single genes from different strains. Consequently, these errors and biases may impact the accuracy and reliability of pan-genome analysis results, as well as downstream functional and phylogenetic analyses.

Keywords: MAG, mock community
References:
1) Conrad, R. E., Viver, T., Gago, J. F., Hatt, J. K., Venter, S. N., Rossello-Mora, R., & Konstantinidis, K. T. (2022). Toward quantifying the adaptive role of bacterial pangenomes during environmental perturbations. The ISME Journal, 16(5), 1222-1234.
2) Mezeti et al., (2021). The reliability of metagenome-assembled genomes (MAGs) in representing natural populations: insights from comparing MAGs against isolate genomes derived from the same fecal sample. Applied and Environmental Microbiology. DOI: 10.1128/AEM.02593-20

7.

Predicting Salinity Preferences from Bacterial Genomes with Machine Learning

Maria Beatriz Walter Costa, Rose Brouns, Bas E. Dutilh

Friedrich Schiller University Jena, Faculty of Biological Sciences, Institute of Biodiversity

maria.beatriz.walter.costa@uni-jena.de

Characterizing niche preferences of bacteria is key to understanding their distribution throughout the Microverse. Preferences for abiotic factors is available online as metadata for bacterial isolates. By associating these with genomic features and/or the pan-genome, we aim to understand what determines organismal niche preferences and at the same time contribute to building predictive models across biomes. Machine learning (ML) is a powerful approach to make predictions and help understand underlying mechanisms. Such models have been extensively used for phenotype-to-genotype association in a microbiological context, such as antimicrobial resistance or growth rate. Few attempts however have been made to predict niche preferences including abiotic factors [1], without deep exploration of the biological meaning of the models.

We collected salinity and other metadata along with genomic data of bacteria from public resources. The two studied groups of salinity preference were: non-halophilic isolates with optimum salinity of 0 practical salinity units (PSU) and halophilic isolates with optimum salinity >35 PSU (for reference, the ocean has salinity of ~35 PSU). With ML models, we predicted salinity preferences using the following genomic features as input: gene families and nucleotide usage patterns (k-mers). We want to expand the study to include more features: DRAM (Distilled and Refined Annotation of Metabolism) categories and Pfam protein domains. As a result, we could successfully classify salinity preference while providing a proof-of-concept of our approach.

Next, we want to study the biological relevance of the genomic features in terms of association with the environment. In this way, we wish to connect the pan-genome to abiotic factors of the environment. With our ML models, we want to expand our understanding of bacterial niche adaptations and contribute to microbiome prediction models.

Keywords: Bacteria, Machine Learning, Salinity, Abiotic Factors, Genome

References:

1) Alneberg et al. (2020). Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. Communications Biology. DOI 10.1038/s42003-020-0856-x.

8.

SynTracker differentiates evolution via mutation or recombination on a per-species basis in complex microbiomes

Hagay Enav, Ruth E Ley

Max Planck Institute for Biology, Tübingen (Germany)

hagayenav@gmail.com

Microbial species diversify into separate strains through mutation, recombination, and gene loss/acquisition. Elucidating mechanisms driving the genomic diversity of species residing within complex microbiomes remains biased towards mutation, because current strain tracking methods have low sensitivity to genomic structural differences. To overcome this and to complement existing strain tracking tools we developed SynTracker, a tool that compares strains using synteny, i.e., the conservation of the order of genomic markers in homologous regions in pairs of metagenomic assemblies or genomes. SynTrackers attributes include low SNP sensitivity, no database requirement, and a high comparative performance. The combined use of SynTracker and SNP-based tools in metagenome analysis allows the identification of species undergoing high rates of recombination with low rates of mutation, or conversely, high rates of structural change with low mutation rates. When used as a standalone, SynTracker can accurately track strains using a fraction of the total genome length, allowing strain tracking in low abundance taxa, plasmids and phages. Taken together, SynTracker, when used alone or combined with other existing tools, provides a novel window into different modes of evolution on a per species-basis in complex microbiomes.

9.

PPanGGOLiN V2: technical enhancement and new features to analyze thousands of prokaryotic genomes

Jean Mainguy[1], Jérôme Arnoux[1], Guillaume Gautreau[2], Adelme Bazin[3], David Vallenet[1] and Alexandra Calteau[1]

1 LABGeM, Génomique Métabolique, CEA, Genoscope, Institut François Jacob, Université d'Évry, Université Paris-Saclay, CNRS. 2 MetaGenoPolis, Paris Saclay University, INRAE, MGP. 3 Global Bioinformatics, Evotec ID (LYON) SAS, Lyon, France

jmainguy@genoscope.cns.fr

For 3 years, PPanGGOLiN[1] (https://github.com/labgem/PPanGGOLiN) has emerged as a solid and widespread solution to construct and analyze prokaryotic pangenome graphs. In contrast to other tools, it is based on a graphical and statistical model to enable efficient and stable partitioning of pangenomes in persistent, shell and cloud genomes. Moreover, using the pangenome graph, PPanGGOLiN is able to identify regions of genomic plasticity (RGPs), their spots of insertions and their segmentation in conserved modules, with the panRGP[2] and panModule methods.

With the ever-increasing number of genomes available in databases, it is essential to optimize the execution and storage capacities of PPanGGOLiN. Here we present the second version of PPanGGOLiN with technical improvements and code revision leading to a more effective data compression of the pangenome file as well as a better user experience. A configuration file can be used for reproducible and tunable analyses. Also, new features have been added in order to enable novel analyses. First, we included a feature allowing to cluster the RGPs based on their family content to investigate the evolutionary dynamics of mobile genetic elements (MGE). Also, the pangenome information (partition, RGPs, spots, modules, …) can be projected on multiple new genomes without recomputing the full pangenome. Furthermore, a set of genes can be searched in the pangenome graph with the aim to extract their conserved genomic context. On top of this, metadata can be added to all pangenome elements and linked to PPanGGOLiN analyzes and outputs, such as functional annotation of gene families.

The aim of this new version of PPanGGOLiN is to continue facilitating the study of the evolutionary dynamics of prokaryotic species in the era of high-throughput genomics, paving the way for comparative pangenomics.

Keywords: Pangenome graph, Bioinformatic software, Microbial Genomics, Mobile genetic elements

References:

1) Gautreau, Guillaume, et al. "PPanGGOLiN: depicting microbial diversity via a partitioned pangenome graph." *PLoS computational biology* 16.3 (2020): e1007732.
2) Bazin, Adelme, et al. "panRGP: a pangenome-based method to predict genomic islands and explore their diversity." *Bioinformatics* 36.Supplement_2 (2020): i651-i658.

10.

PANORAMA: comparative pangenomics tools to explore interspecies diversity of microbial genomes

Jérôme Arnoux, Jean Mainguy, Laura Bry, Quentin Fernandez de Grado, David Vallenet and Alexandra Calteau

LABGeM, Génomique Métabolique, CEA, Genoscope, Institut François Jacob, Université d'Évry, Université Paris-Saclay, CNRS

jarnoux@genoscope.cns.fr

In recent years, to cope with the increase of genomes in databases, comparative genomics studies have focused on the overall gene content of a species, the pangenome, imposing a paradigm shift in the representation of knowledge and in the algorithms used.

We developed PANORAMA, a flexible and open-source bioinformatics toolbox, which exploits multiprocessing, to perform rapid and easy-to-use comparative analysis of pangenomes using thousands of microbial genomes. It benefits from methods for the reconstruction and analysis of pangenome graphs, thanks to the PPanGGOLiN[1] software suite (https://github.com/labgem/PPanGGOLiN). PANORAMA integrates multiple features. It leverages homologous family conservation combined with graph connectivity to allow users to search for a specific genomic context in a set of pangenome graphs. Then, the presence-absence of the context in multiple pangenomes can be studied, and a completeness score is associated. PANORAMA also predicts biological systems, such as conjugation, secretion or defense systems, at the pangenome level. Pangenome gene families must be first associated with functional annotations using HMM profile searches. These HMMs are associated with a system-modeling describing gene presence/absence and synteny rules. PANORAMA detects genomic contexts corresponding to the models and verifies that presence/absence parameters fit with the associated rules. All generated results are associated to pangenome partitions, as well as to regions of genomic plasticity (RGPs), their spot of integration and their segmentation in conserved modules, thanks to panRGP[2] and panModule methods. The module associations can be used to find functional similarities between modules. A score of similarity between modules from different pangenomes is computed, corresponding to the number of common families compared to the total number of families in modules.

PANORAMA aims to help microbiologists to understand the adaptive potential of bacteria and, with the exploration of functional modules in different species, to better understand the evolutionary dynamics behind the metabolic diversity of microorganisms. Future developments will integrate additional models for the annotation of KEGG modules or Polysaccharide Utilization Loci. We are also working on the integration of pangenomes in graph databases, such as Neo4J, to address the challenge of large-scale comparative pangenomics.

Keywords: Pangenome graph, Bioinformatic software, Microbial Genomics, Comparative pangenomics, Biological systems

References:

1) Gautreau, Guillaume, et al. "PPanGGOLiN: depicting microbial diversity via a partitioned pangenome graph." *PLoS computational biology* 16.3 (2020): e1007732.
2) Bazin, Adelme, et al. "panRGP: a pangenome-based method to predict genomic islands and explore their diversity." *Bioinformatics* 36.Supplement_2 (2020): i651-i658.

11.

Population genomics of Archaea

Maistrenko, Oleksandr[1], Wen-Cong Huang[1], Joshua Hamm[1], Dina Castillo, Boukhchtaber[1], Scott Maxson[1], Nina Dombrowski[1], Alejandro Abdala Asbun[1], Bastiaan von Meijenfeldt[1], Jacob Munson-McGee2, Mark Young[2], Rachel Whitaker[3], Henk Bolhuis[1], Ana-Belen Martin-Cuadrado[4], Francisco Rodriguez-Valera[5], Laura Villanueva[1,6], Anja Spang[1,7]

[1] Royal Netherlands Institute for Sea Research (NIOZ), Department of Marine Microbiology & Biogeochemistry, Landsdiep 4, 1797 SZ, 't Horntje (Texel), Netherlands. [2] Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, Montana, USA. [3] Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL, United States. [4] Universidad de Alicante, Departamento de Fisiología, Genética y Microbiología, 03690 San Vicente del Raspeig, Spain. [5] Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, Alicante, Spain. [6] Department of Earth Sciences, Faculty of Geosciences, Utrecht University, Princetonlaan 8a, Utrecht 3584, The Netherlands. [7] Department of Evolutionary & Population Biology, Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, The Netherlands.

oleksandr.maistrenko@nioz.nl

Recent progress in microbiome research has led to the discovery of diverse clades of Archaea, including the superphylum DPANN, which comprises at least eight phylum-level lineages. Many known DPANN representatives are small microorganisms with streamlined genomes. It is hypothesized that most of these microorganisms depend on symbiotic relationships with other archaeal or potentially bacterial groups; however, most of the hosts of DPANN are unknown. Initially, a member of this group of organisms was isolated from an extreme environment: a thermal vent system in the North Atlantic Ocean, near Iceland. Later this group of microorganisms was found prevalent in many other environments, including hydrothermal, hypersaline, freshwater, marine and terrestrial environments, present in the water column and sediments, and was even detected in the human gut. Such a broad range of habitats suggest that DPANN are likely to be ubiquitous and highly important for the evolution of their host organisms and the functioning of ecosystems in general. The mechanisms of genome streamlining at micro- and macroevolutionary scales in Archaea, across extreme and moderate habitats, are not understood in the context of microbe-microbe interactions. In this study, we establish a pipeline for meta-pangenomic analysis of public and new datasets that allows us to construct the pangenomes of archaeal species and investigate global trends of pangenome evolution and co-evolution in Archaea with a focus on DPANN symbiont and their host populations. This will enable us to establish how pangenome diversity across Archaea is associated with biotic and abiotic factors driving symbiont-host evolution.

12.

Evolution of the Patescibacteria symbiotic lifestyle

Bouderka Feriel, David Moreira, Philippe Deschamps, Purificacion Lopez-Garcia, Ludwig Jardillier

Unité d'Ecologie, Systématique et Evolution, Université Paris-Saclay. IDEEV site - 12 rue 128, 91190 Gif-sur-Yvette, France.

feriel.bouderka@universite-paris-saclay.fr

The Patescibacteria has been recently reported as the sister group to the free-living phyla Chloroflexota and Dormibacterota [1]. Yet, based on the numerous metagenomic studies and the few culture representatives available, this phylum is thought to include mainly symbiotic representatives [2]. The evolutionary history of the transition from free-living to a symbiotic lifestyle in Patescibacteria, whether confined to certain representatives or widespread throughout the whole clade, is not fully comprehended. The genomes of these bacteria are characterized by a lack of many genes coding for major metabolic pathways. Consequently, the hypothesis of extensive gene losses, which may have occurred through one or multiple independent events, has been proposed. This hypothesis heavily relies on annotations obtained using similarity-based comparisons with well-characterized representatives. However, about half of the Patescibacteria genes cannot be functionally annotated using this approach, leaving open the possibility that they code for yet unknown pathways or highly derived known pathways that escape detection. Additionally, many patescibacterial proteins appear to be involved in interactions with host cells (enabling cell-to-cell contact, metabolites exchange, etc.), which suggests gene gains and/or repurposing of already available genes. To determine whether the Patescibacteria symbiotic lifestyle primarily arose from gene losses or whether other mechanisms played a role, as well as whether the symbiotic lifestyle evolved in one or multiple independent events, we studied gene content in more than 480 genomes and Metagenome Assembled Genomes (MAGs) of Patescibacteria, Chloroflexota, and Dormibacterota from the Genome Taxonomy DataBase (GTDB). We reconstructed the phylogenetic relationships between all the selected representatives, identified more than 20,000 clusters of orthologous genes, and built their individual phylogenetic trees. Using the annotation-independent gene trees-species tree reconciliation method, we aim to detect patterns of gene gain, loss, and transfers in Patescibacteria and its sister groups, as well as to reconstruct the coding and metabolic potential of Patescibacteria\'s last common ancestor. This information should provide valuable insights into the evolution of symbiotic lifestyles within this intriguing phylum.1. Coleman GA et al. (2021). A rooted phylogeny resolves early bacterial evolution. Science. doi:10.1126/science.abe05112. Castelle CJ et al. (2018). Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. Cell. doi:10.1016/j.cell.2018.02.016

13.

Nitrate-reducing microbes in winter in sediments of large boreal lakes affected by browning and mining

Carlos Palacin-Lizarbe[1], Stefan Bertilsson[2], Moritz Buck[2], Tong Liu[2], Henri J. Siljanen[1], Dhiraj Paul[1], Lukas kolh[1], Mikko Kiljunen[3], Saani L. Aalto[4], Antti J. Rissanen[5], Anssi Vainikka[6], Jukka Pumpanen[1]

[1] Department of Environmental and Biological Sciences, University of Eastern Finland, Kuopio Campus, P.O. Box 1627, FI-79211 Kuopio, Finland. [2] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden. [3] Department of Biological and Environmental Science, University of Jyväskylä, PO Box 35, 40014 Jyväskylä, Finland. [4] DTU Aqua, Section for Aquaculture, The North Sea Research Centre, Technical University of Denmark, P.O. Box 101, 9850 Hirtshals, Denmark. [5] Faculty of Engineering and Natural Sciences, Tampere University, Korkeakoulunkatu 6, 33720 Tampere, Finland. [6] Department of Environmental and Biological Sciences, University of Eastern Finland, Joensuu Campus, P.O. Box 111, 80101 Joensuu, Finland.

carlos.palacin@uef.fi

The ice-covered period of boreal lakes has contrasting environmental conditions respect to the ice-free, with cold temperatures, absence of light, and minor gas exchange between water and atmosphere. Focusing on the nitrogen (N) cycle, winter seems a suitable period for N-transforming prokaryotes with a high availability of reactive-N due to minor assimilation by photoautotrophs. However, there is limited data about winter N cycling rates and the microbes involved on, and about the role of organic matter quality on N cycling.

We studied 2 oligotrophic large boreal lakes in North Karelia, Finland, with clear-water and brown-water sides, and an additional side affected by mining resulting in higher nitrate and sulphate levels in the hypolimnion. During winter of 2021 we sampled at the beginning and at the end of the ice-cover. Using the isotope-pairing technique (IPT) we incubated sediment cores with $^{15}NO_3^-$ and quantified the products of A) complete denitrification (N_2), B) truncated denitrification (N_2O), and C) dissimilatory nitrate reduction to ammonium (DNRA, NH_4^+) to infer the rates of these nitrate-reducing processes. Also, to see the role of organic matter, we did anoxic slurry incubations with $^{15}NO_3^-$ and I) lake water, II) miliQ water, III) algal dissolved organic matter (DOM) extract, or IV) peatland DOM extract. We characterized the DOM using FT-ICR MS. We also explore the genetic potential (DNA) of the sediment microbiome by using several sequencing techniques: A) amplicon (16S rRNA), B) targeted (N-mineral and CH4 functional genes), and C) shotgun.

Preliminary results identify the N-transforming microbes and point to changing nitrate consuming activities and genetic potentials between sites. *METABOLIC* software (1) indicates the metagenomes-assembled genomes (MAGs) present in the sediments have complex heterotrophic capabilities with organic carbon oxidation, fermentation, acetate oxidation, iron reduction and sulfur oxidation as the more common abilities, being present in 100, 86, 62, 62 and 59% of the mOTUS, respectively. Using *NcycDB* (2) we look on their N capabilities, only few of these mOTUs can do nitrogen-fixation while most can do nitrate-reducing processes.

Keywords: sediment microbiome, winter limnology, denitrification, DNRA, N2O

References:

1) Tu, Q., et al. (2019). NCycDB: A curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes. *Bioinformatics*, *35*(6), 1040–1048. https://doi.org/10.1093/bioinformatics/bty741
2) Zhou, Z., et al. (2022). METABOLIC: High-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome*, *10*(1), 33. https://doi.org/10.1186/s40168-021-01213-8

14.

Exploring Microbial Processes and Genomic Islands in Deadwood Decomposition: Integrating Metagenomics and Metatranscriptomics

Etienne Richy[1*], Priscila Thiago Dobbler[1], Vojtěch Tláskal[2], Martina Kyselková[1], Petr Baldrian[1]

[1] Laboratory of Environmental Microbiology, Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic. [2] Institute of Soil Biology and Biogeochemistry, Biology Centre of the Czech Academy of Sciences, Ceske Budejovice, Czech Republic

etiennerichy.art@gmail.com

Deadwood constitutes a significant reservoir of organic matter in forests and plays a role in soil formation. It provides diverse niches and harbors a wide array of microorganisms, including bacteria and fungi, which collaborate to decompose the wood biomass. However, the microbial processes driving deadwood decomposition are not yet fully understood. In this study, we employed a combination of long-read sequencing on a Pacific Biosciences platform and short-read metatranscriptomics to reconstruct 71 bacterial metagenome-assembled genomes (MAGs) and investigate the functional roles of these bacteria during different wood decomposition stages. By utilizing pangenome analysis, we identified the core genes involved in the principal deadwood decomposition pathways and additional genes associated with alternative metabolic capabilities. We characterized genomic islands within the bacterial genomes and evaluated their expression patterns. The MAGs was dominated by Proteobacteria, Bacteroidota, Verrucomicrobiota, and Patescibacteria. Fourteen MAGs exhibited completeness above 90% and a third of the MAGs consisted of fewer than 10 contigs, including seven represented by a single contig. Our findings revealed that bacteria commonly utilized cellulose, xyloglucan, glucan, and arabinan during wood decomposition, while acetate, alcohol and pyruvate were frequently transformed to generate energy. Furthermore, we detected genomic islands in varying proportions across all bacteria. Although viral origin in the genomic islands was generally low, one MAG assigned to the *Methylocella* genus exhibited all 12 genomic islands of putative viral origin. This particular MAG drew our attention due to its high completeness (99.82%) within one contig, very low contamination (0.04%), lack of methane oxidation capacity, and high abundance, particularly in the wood at late stage of decomposition after 20 years since downing. Ongoing research aims to further elucidate the expression patterns of these MAGs, particularly the genes encoded within the genomic islands.

Keywords: Deadwood, genomic island, decomposition, bacteria, pangenome

15.

Pangenomics to understand the emergence of a new *Pectobacterium brasiliense* pathovar

Lakhansing A. Pardeshi[1]*, Eef M. Jonkheer[1], Michiel J.C. Pel[2], Inge van Duivenbode[3], Dick de Ridder[1], Sandra Smit[1], Theo A. J. van der Lee[1]

[1]Wageningen University and Research, The Netherlands. [2]Netherlands Food and Consumer Product Safety Authority, The Netherlands. [3]Nederlandse Algemene Keuringsdienst, The Netherlands.

lakhansing.pardeshi@wur.nl

Soft rot disease is mainly caused by bacteria from the *Pectobacterium* and *Dickeya* species. The genetically highly variable *P. brasiliense* (Pbr) species first appeared in 2013 in the Netherlands and subsequently became dominant, detected in almost 90% of the field samples showing potato blackleg. Recently, a new lineage of Pbr was identified that escaped a PCR assay for the detection of blackleg causing Pbr strains. Here, we applied a pangenomic approach to understand the evolution and establishment of a new virulent lineage. *Pectobacterium* isolates sequenced as a part of routine inspection in the Netherlands were combined with *Pectobacterium* genomes available in NCBI, to build a pangenome using PanTools[1].

ANI clustering of genomes revealed a distinct clade of newly emerged Pbr isolates that were false-negative for virulence assays. To explain these false-negative results, PCR primers were mapped against genomes, which resulted in no match for the isolates that showed false negative results while a perfect match for the true-positive isolates. Pangenome construction and subsequent grouping of genes resulted in the identification of 30,156 homology groups for the *Pectobacterium* spp. Accessory groups in this pangenome had virus integration-related processes enriched. Association analysis of homology groups identified 70 groups that were specific to the new lineage with respect to other Pbr genomes. Based on the uniqueness and negative cross-species match, we selected 5 candidates to design a novel PCR assay for the identification of this lineage. Mapping these homology groups to the reference genomes of the newly detected Pbr isolate showed a 25kb region with bacteriophage origin genes arranged in tandem, indicating a prophage sequence. This prophage was also present in other Pectobacterium species in the pangenome.

Although a *P. brasiliense* specific pangenome could have detected the homology groups for the new virulent clade, a genus-scale pangenome helped to detect shared genomic dynamics at genus level, where we also detected a bacteriophage insertion in other *Pectobacterium* species. Thus, the current analysis highlights the importance of pangenomic analysis to understand the evolutionary dynamics of evolving pathogens. Additionally, pangenome analysis allowed us to select lineage-specific target genes to design new PCR assays.

Keywords: Pectobacterium, pathogen, virulence, pangenome

References:

1) Jonkheer, E. M. et al. (2022) PanTools v3: functional annotation, classification and phylogenomics. Bioinformatics 38, 4403–4405.

16.

Comparative enomics of the *Pseudomonas* pangenome reveals host- and environment-specific evolution

Zaki Saati-Santamaría[1,2,3], Ricardo Baroncelli[4], Raúl Rivas[1,2], **Paula García-Fraile**[1,2,5*]

[1] Departamento de Microbiología y Genética, Universidad de Salamanca, Salamanca, Spain. [2] Institute for Agribiotechnology Research (CIALE), Villamayor, Salamanca, Spain. [3] Institute of Microbiology of the Czech Academy of Sciences, Vídeňská, Prague, Czech Republic. [4]Department of Agricultural and Food Sciences (DISTAL), University of Bologna, Bologna, Italy. [5]Associated Research Unit of Plant-Microorganism Interaction, USAL-CSIC (IRNASA), Salamanca, Spain

paulagarciafraile@usal.es

Each Earth ecosystem has unique microbial communities. *Pseudomonas* bacteria have evolved to occupy a plethora of different ecological niches, including living hosts, such as animals and plants. Many genes necessary for the *Pseudomonas*-niche interaction and their encoded functions remain unknown. Here[1], we describe a comparative genomic study of 3,274 genomes with 19,056,667 protein-coding sequences from *Pseudomonas* strains isolated from diverse environments.

In summary, we show the genomic adaptability patterns of *Pseudomonas* strains to different lifestyles. For example, plant-associated *Pseudomonas* strains dedicate the largest number of genes to the metabolism of carbohydrates, but the involved proteins are likely located in the cytosol, in contrast to other strains that present a higher proportion of CAZys in the outer envelope or excrete them. Additionally, contrary to plants, the human/animal environment seems to add pressure to resist stresses, although this issue may be biased toward the better understanding of clinically relevant genes. Furthermore, the association of Pseudomonas with higher hosts increases the probability of gene exchange through horizontal transfer. Overall, our results will facilitate studies focusing on the evolutionary dynamics, ecology, biotechnology, and clinical relevance of bacteria. New insights into genes or functions associated with isolation niches can inspire scientific applications in infectious disease diagnosis and treatment or even the development of engineered strains with biotechnological uses.

Keywords: *Pseudomonas* pangenome; GWAS; Comparative genomics; Host-microbe interactions

References:

1) Saati-Santamaría et al., (2022). Comparative Genomics of the Genus Pseudomonas Reveals Host-and Environment-Specific Evolution. *Microbiology spectrum*, https://doi.org/10.1128/spectrum.02370-22

17.

Analysis of the pangenome in Thaumarchaeota from the western Mediterranean Sea in seasons with mixed and stratified water column

Pablo Suárez-Moo*, Carmen Molina Pardines*, Jose M. Haro-Moreno, Mario López-Pérez; Francisco Rodriguez-Valera

*Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, 03550 Alicante, Spain. *Equal contribution.*

psuarez@umh.es

Thaumarchaeota or ammonia-oxidizing archaea (AOA) represents an important archaea group associated with the nitrogen cycle, contributing to ammonia oxidation in oligotrophic marine environments [1]. The goal of the present study was to describe the diversity and patterns of variation of the pangenome (core and flexible) in a population of Thaumarchaeota marine microbes from the western Mediterranean Sea in two seasons, using a long read and genome recruitment approach. A high phylogenetic diversity of the Thaumarcheota population in the stratified and mixed column water seasons using 16S long-read (PacBio CCS) was found. Some Metagenome assembled genomes (MAGs) were recovered using a hybrid contig assembly of short (Illumina) and long reads (PacBio CCS). Island regions (IR) were found in the genome recruitment of a reference genome *Ca.* Nitrosopelagicus brevis CN25 (GCA_000812185.1) against previous metagenomes from the western Mediterranean Sea. These IRs were aligned against the long reads from mixed and stratified water column samples, and protein coding sequence (CDS) that as expected encode predicted proteins associated with glycosylation of cell surface components were found. However, several under-recruiting islands were detected with diverse gene content, and ammonia and phosphonate transporters.

18.

A global pangenome of marine nitrite-oxidizing bacteria highlights differences in nutrient acquisition strategies

Justine B. Albers[1], Kiera Sullivan[1], Barbara Bayer[2], Alyson E. Santoro[1]

[1] *Department of Ecology, Evolution & Marine Biology, University of California, Santa Barbara, California, USA.* [2] *Centre for Microbiology and Environmental Systems Science, University of Vienna, Austria*

justinealbers@ucsb.edu

Nitrospinae are nitrite-oxidizing bacteria (NOB) adapted to nutrient-poor marine environments. NOB make up only a small fraction of the total marine microbial community; however, nitrite oxidoreductase (Nxr) – an iron (Fe)-containing enzyme that mediates nitrite oxidation – is one of the most abundant proteins found in the mesopelagic[1]. If NOB have a high requirement for Fe, the availability of which varies by depth and ocean basin, then populations of NOB may have developed different strategies for accessing Fe depending on its availability. There is preliminary evidence for trace metal-based niche differentiation in the genomes of isolated *Nitrospinae*; a Pacific Ocean strain (Nb-3) contains genes to produce a ferric iron chelator (siderophore) while a closely related Atlantic Ocean strain (Nb-211) does not[2]. Expanding this analysis to include environmental metagenome-assembled genomes may provide additional support for this hypothesis at the population level. Therefore, we are using pangenomics to determine if NOB are adapted to different gradients of iron and nutrient availability. Here we present a global pangenome of *Nitrospinae* and investigate the placement of nutrient acquisition and transport genes in the core vs flexible pangenome, including genes for iron acquisition and siderophore biosynthesis and transport. We show that marine NOB are grouped into at least two distinct clades that differ in their biosynthetic capabilities. We also highlight environmental factors such as depth, ocean basin, and nutrient availability that may be driving differentiation in the observed clades.

Keywords: nitrite-oxidizing bacteria, pangenome, iron

References:
1) Saito et al. (2020). Abundant nitrite-oxidizing metalloenzymes in the mesopelagic zone of the tropical Pacific Ocean. *Nature Geoscience.* DOI:10.1038/s41561-020-0565-6.
2) Bayer et al. (2022). Complete Genome Sequences of Two Phylogenetically Distinct Nitrospina Strains Isolated from the Atlantic and Pacific Oceans. Microbiology Resource Announcements. DOI: 10.1128/mra.00100-22.

19.

Uncovering Methane Sink Potential and Habitat Preferences of Methanotrophs in Denmark

AKS Knudsen, TBN Jensen, M Sereika, PH Nielsen, M Albertsen & C Singleton

Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark

aksk@bio.aau.dk

Methane is a potent greenhouse gas with a global warming potential 80 times higher than CO_2 on a 20-year time scale [1]. Mitigation of methane is essential to limit increase in global surface temperatures below the critical threshold of 1.5-2°C [1]. To address this issue, it is essential to understand the diversity and biogeography of methanotrophs, the microorganisms responsible for lowering emissions by oxidising methane. The current understanding of methanotrophs relies largely on isolates. Considering the diversity discovered within the functional genes for methane oxidation, it is evident that several groups have not yet been described, lacking both isolates and genome representatives. This is especially the case for atmospheric methane oxidisers, being of particular interest from a methane mitigation perspective. Historically, methanotrophs have been identified by the marker genes *pmoA* and *mmoX*, encoding two isoforms of methane monooxygenase. However, *pmoA* and *mmoX* share sequence similarity to monooxygenases oxidising non-methane compounds. This complicates identification based solely on marker genes and requires detailed information on gene phylogeny and metabolic potential. Furthermore, methanotrophs display species-level variation in substrate affinity and pathways employed for assimilation of compounds such as carbon, nitrogen, and sulphur. Defining the core genome within novel species and investigating the ecological relevance of the accessory genome could prove pivotal in understanding this important group of GHG-mitigators. A comprehensive dataset of 10,000 Danish shallow metagenomes (the Microflora Danica project) has set the stage for selecting soil samples for long-read sequencing, based on the presence of methanotrophy marker genes. We managed to populate poorly described parts of the methanotroph tree of life with HQ MAGs, including, but not restricted to novel species of atmospheric methane oxidisers. In combination with the extensive dataset of 10,000 shallow metagenomes, we aim to establish an association between novel methanotrophs, genomic variation, habitat preferences, and geographical distribution on a national scale.

Keywords: Methanotrophs, metagenome-assembled genomes, shallow metagenomes, gene phylogeny

References:

1) IPCC. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. 10.1017/9781009157896.

20.

Genomic diversity of *Zoogloea*: Exploring secrets of biopolymer production in the world of activated sludge

Sofie Zacho Vestergaard[1], Mojtaba Azari-Anpar[2], Amrita Bhattacharya[2], Thomas William Seviour[2], Morten Kam Dahl Dueholm[1], Per Halkjær Nielsen[1]

[1] *Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark.*
[2] *Department of Biological and Chemical Engineering, Aarhus University, Aarhus, Denmark.*

szv@bio.aau.dk

Why care about extracellular polymeric substances (EPS) from activated sludge (AS) wastewater treatment systems? Wastewater is one of the largest waste streams worldwide and is mostly treated through the AS process. A large fraction of this biomass is constituted by EPS secreted by bacteria enabling them to form microbial flocs. In the REThiNk project we want to utilize AS-EPS by extracting and recycling these into new alternative biopolymers replacing oil-based polymers, which will implement a novel and sustainable approach to resource recovery in wastewater treatment. However, we lack fundamental knowledge on EPS composition and formation and the identity of the microbial producers in AS systems, all essential questions that needs to be addressed before AS-EPS can be introduced as a commercial resource. Members of the genus *Zoogloea* belong to one of the few known EPS producers in AS, known to be important for floc formation and thereby for the wastewater treatment efficiency.

Here, we explored and compared the molecular machinery in *Zoogloea* to gain fundamental knowledge on EPS production along with functional and evolutionary relationship within the genus. Publicly available whole genome sequences belonging to the *Zoogloea* genus (GTDB r214) and three in-house isolates (*Z. caeni*, *Z. resiniphila* and *Z. oleivorans*) formed the basis for this comparative genomic analysis. Two extracellular polysaccharide (exoPS) gene clusters, constituted by 7 and 28 genes respectively, have recently been found to be required for floc formation in *Zoogloea resiniphila* [1,2]. Additionally, these genetic elements were encoded in other AS-relevant bacteria [1,2]. Based on this, we wanted to examine the conservation of these exoPS biosynthetic pathways within the *Zoogloea* genus as well as map the full distribution in the entire bacterial kingdom to evaluate to which taxonomic level the exoPS production is conserved. Multiple sequences alignments revealed two clades of the 7-gene gene cluster within the *Zoogloea* genus, suggesting differences in exoPS biosynthetic pathways even within the genus. To further validate the structure and conservation of the exoPS production, we have extracted EPS from three *Zoogloea* isolates which are to be used for chemical and structural analyses to verify genomic findings.

Keywords: Comparative genomics, metabolic reconstruction, phylogenetics, biopolymer, genome-mining

References:

1) An et al. (2016). Comparative genomics analyses on EPS biosynthesis genes required for floc formation of Zoogloea resiniphila and other activated sludge bacteria. Water Research, Volume 102. https://doi.org/10.1016/j.watres.2016.06.058.
2) Gao et al. (2016). Both widespread PEP-CTERM proteins and exopolysaccharides are required for floc formation of Zoogloea resiniphila and other activated sludge bacteria. Environmental Microbiology. 10.1111/1462-2920.14080.

21.

Exopolysaccharide biosynthesis genes clusters in *Alteromonas*: a role in carbon remineralization and environmental adaptation

R. Liébana, C. Perez-Cruz, U. Arrizabalaga, E. Bilbao, L. Alonso-Sáez

AZTI, Marine Research Division, Basque Research Technology Alliance (BRTA), Sukarrieta, Spain

rliebana@azti.es

Alteromonas are marine widespread copiotrophs which play an important role in the biogeochemical cycling of organic matter in the ocean, by e.g. degrading polymers generated during algal blooms[1]. Members of this genus are considered r-strategists which can rapidly proliferate in response to pulses of organic matter. The secretion of an extracellular degradative matrix may represent a relevant, yet poorly understood mechanism of some marine taxa like *Alteromonas* for efficiently degrading marine polymers. Exopolysaccharides (EPS) represent a major component of this extracellular matrix, with a key role not only in nutrient acquisition, but secretion of EPS by members of *Alteromonas* in a biotechnological context[2], but their potential role in natural habitats has not been addressed. Here, we studied the presence of marker genes of EPS synthesis in the genome of an *Alteromonas* 805 strain isolated from deep oceanic waters. We identified four differentiated EPS biosynthesis gene clusters and analysed their transcriptional patterns under growth on different carbon sources (i. e. alginate, laminarin and glucose). Additionally, we analysed the presence of EPS biosynthesis gene clusters in phylogenetically related and unrelated marine bacteria, observing that these genes are highly conserved within the *Alteromonas* genus. The distribution of Alteromonas' EPS synthesis genes in a dataset containing metagenomes and metratranscriptomes from the global ocean was further analysed to address the potential ecological relevance of this mechanism. Our results aim at expanding our understanding of the role of EPS secretion in *Alteromonas* in the remineralization of carbon produced by phytoplankton and their adaptation to different environmental conditions.

Keywords: Alteromonas, extracellular polysaccharides (EPS), EPS biosynthesis gene clusters, carbon remineralization

References:

1. Koch, H. et al. Biphasic cellular adaptations and ecological implications of *Alteromonas macleodii* degrading a mixture of algal polysaccharides. ISME J. 13, 92–103 (2019).
2. Poli, A., Anzelmo, G. & Nicolaus, B. Bacterial exopolysaccharides from extreme marine habitats: Production, characterization and biological activities. Marine Drugs vol. 8 1779–1802 (2010).

22.

A novel machine learning prediction tool for extracellular proteins doubles current size of the marine secretome

Asier Zaragoza-Solas, Federico Baltar

Fungal and Biogeochemical Oceanography Group, Department of Functional and Evolutionary Ecology, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria

asier.zaragoza.solas@univie.ac.at

In order to interact with their environment and the community, bacteria secrete proteins into the environment (known collectively as the secretome), where they remain active for prolonged periods of time. Despite their importance and the vast quantities of marine prokaryotic sequences available, our knowledge of the marine secretome is limited by the *in silico* methods available for the study of secreted proteins, which are not suited to their use in large datasets. A possible solution is to exploit the evolutionary adaptation changes in the proteome for adaptation to the marine environment, which affect secreted proteins more intensely. In this study, we uncover the specific adaptations of marine extracellular proteins. We found that these adaptations vary between phyla, resulting in differences in ATP costs, amino acid composition and nitrogen and sulphur content. With these adaptations in mind, we have developed Ayu, a machine prediction tool that does not employ homology-based predictors and achieves better performance than current state-of-the-art software at a fraction of the time. When applied to the Tara Oceans dataset, our method was able to recover more than double the proteins compared to only using the signal peptide. Thus, this tool allows to better characterize the role of the secretome in marine ecosystems, and exposes the large fraction of secreted proteins that are presently unconsidered in current secretome analysis.

Keywords: Secretome, Machine learning, Metagenome

17.00-19.30	FLASH TALKS I. 5 min.	Chair: Alex Mira
23	Mascarós Núñez, Patricia	*Staphylococcus aureus*: Antibiotic Resistance and Virulence - A Comparative Approach
24	Díaz Méndez, José Francisco	Exploring *Staphylococcus aureus*: Pangenome, Phylogeny and Comparative Analysis of Mobile Genetic Elements among Clonal Complexes
25	Sabnis, Akshay	Low plasmid mobility benefits cheaters and prevents eradication in nature
26	Rodriguez Mestre, Mario	Unraveling the role of plasmids in disseminating prokaryotic immune systems across different ecosystems
27	Figueroa Chavez, Wendy	Lateral transduction is the main route of horizontal gene transfer in immunocompetent bacteria
28	Robinson, Christopher	Longitudinal tracking of the ecology and evolution of mobile genes in the honey bee microbiome
	TIME FOR QUESTIONS (10 min)	
29	Priest, Taylor	Circularised mobile genetic elements and their eco-evolutionary patterns at ecosystem and global scales
30	Romero Picazo, Devani	The evolution of domesticated plasmids coincides with species divergence in the plant-associated genus Pantoea
31	Mehrshad, Maliheh	Phage co-infection; a mystifying shuffling trick for the pangenome deck
32	Gonzalez-Serrano, Rafael	Homologous host recognition modules in distant families of *Alteromonas* phages
33	Ashkenazy, Haim	The evolution of *Arabidopsis thaliana*-associated Pseudomonas
34	Lomeli Ortega, Carlos Omar	Modulation of shrimp microbiota using phage therapy during a controlled infection
35	Guridi Fernández, Pablo	Exploring DNA transfer and integration in human cells upon infection with intracellular bacterial pathogens
	TIME FOR QUESTIONS (10 min)	
36	Sánchez-Nieto, Esperanza	Prokaryotic reverse transcriptases. Retroelements converted into specialized defense systems
37	Nadal Molero, Francisco	MITES as a tool to Associate Virus to Microbial Host
38	Koszucki, Janusz	Using genome-wide associations studies to identify putative genetic determinants of phage host-range in *Klebsiella*
39	Rosselli, Riccardo	Exploring the potential of CRISPR-Cas module transfers between bacteria lineages to overcome Anti-CRISPR inhibition
40	Smug, Bogna	How phages play with LEGO: studying protein-level mosaicism in phages and its evolutionary implications
41	Pinilla Redondo, Rafael	Phages suppress CRISPR-Cas immunity via RNA-based anti-CRISPRs
42	López Beltrán, Adrián	Virus-Host Ecological Interactions in the Gut Microbiome Through the Analysis of CRISPR Arrays
43	Pimenoff, Ville	Ecology of oncogenic human papillomaviruses after the eradication of vaccine-targeted types
44	Martinez-Martinez, Daniel	Navigating the *E. coli* Pangenome to Identify Metformin Drivers
	TIME FOR QUESTIONS (10 min)	

23.

Staphylococcus aureus: Antibiotic Resistance and Virulence - A Comparative Approach

Mascarós P., *Díaz-Mendez J.F., Martínez-Seijas C., Arnau A., Selva L., Corpa J.M, Viana D.*

Grupo de investigación "Patología y Sanidad Animal", PASAPTA, Instituto de Ciencias Biomédicas, Facultad de Veterinaria, Universidad Cardenal Herrera-CEU, CEU Universities, c/ Assegadors nº 2, 46115 Alfara del Patriarca, Valencia, Spain.

patricia.mascarosnunez@uchceu.es

Staphylococcus aureus, an opportunistic microbe known for its pathogenicity and adaptability, is a concern for both human and animal medicine. In recent years, it has caused a significant increase in infections in rabbits, resulting in substantial economic losses for farmers. In this study, the analysis of 90 identified strains of *Staphylococcus aureus* isolated from rabbits exhibiting lesions in Spain and Portugal was conducted. These samples were subjected to sequencing, resulting in a total of 85 samples suitable for this investigation. Among the sampled strains, the most prevalent clonal complexes were 121 and 96, a finding consistent previous works, as these strains have traditionally been the most frequently isolated from rabbit farms.

A comparative study of clonal complexes was undertaken, focusing on antibiotic resistance profiles and virulence factors. Resistance genes were analysed using the 'AMRfinder' bioinformatic tool, whereas virulence genes were identified using the 'vfdb' tool.

Regarding antibiotic resistance, CC121 and CC398 exhibited the highest diversity of resistance genes, while CC 45, CC1, and CC130 demonstrated the lowest resistance profiles. Correlation was observed between the number of resistance genes and the number of mobile genetic elements present in each clonal complex, as previously observed in other studies. [1]

In contrast, virulence factor distribution exhibited a similar pattern across different clonal complexes, observing the same pattern in all of the isolates. Interestingly, when considering the origin of the samples, more invasive lesions correlated with a higher prevalence of virulence genes belonging to the "effector delivery system" group. The *Staphylococcus aureus* type VII secretion system (T7SS) belongs to this group. The T7SS displays modularity and heterogeneity in expression between different strains and strains lacking any of these components are less virulent in infection models[2]. Therefore, the "effector delivery system" group could be related to the virulence of the most severe strains and lesions.

The association between clonal complexes, antibiotic resistance patterns, and virulence factor distribution sheds light on the pathogenic potential and adaptability of these strains. These findings contribute to our understanding of the epidemiology and virulence mechanisms of *Staphylococcus aureus* in rabbit populations, with potential implications for disease management.

Keywords: Staphylococcus aureus, antibiotic resistance, virulence factors

References:

1.Wang W, Li FQ. Overview of *Staphylococcus aureus* mobile genetic elements and horizontal gene transfer of antimicrobial resistance genes. Zhonghua Yu Fang Yi Xue Za Zhi. 2018 Oct 6;52(10):1067-1071. Chinese. doi: 10.3760/cma.j.issn.0253-9624.2018.10.020. PMID: 30392330.

2.Kengmo Tchoupa, A., Watkins, K. E., Jones, R. A., Kuroki, A., Alam, M. T., Perrier, S., Chen, Y., & Unnikrishnan, M. (2020). The type VII secretion system protects *Staphylococcus aureus* against antimicrobial host fatty acids. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-71653-z.

24.

Exploring *Staphylococcus aureus*: Pangenome, Phylogeny and Comparative Analysis of Mobile Genetic Elements among Clonal Complexes

Díaz-Méndez J.F. [1*], Mascarós P. [2*], Martínez-Seijas C[2*]., Arnau A. [2*], Selva L. [2*], CorpaJ.M. [2*],Viana D. [2*]

[1 2] Grupo de investigación "Patología y Sanidad Animal", PASAPTA, Instituto de Ciencias Biomédicas, Facultad de Veterinaria, Universidad Cardenal Herrera-CEU, CEU Universities, c/ Assegadors nº 2, 46115 Alfara del Patriarca, Valencia, Spain

jose.diazmendez@uchceu.es

Staphylococcus aureus, a Gram-positive bacterium known for its pathogenicity and clinical significance, poses a versatile threat as both a human and animal pathogen[1], capable of causing a wide range of infections. In recent years, there has been a concerning increase in staphylococcal infections in rabbits, leading to substantial losses. This underscores the urgency of understanding the genetic basis of *S. aureus* adaptation and pathogenicity in this host.

In this preliminary study, a comprehensive analysis was conducted of 90 *S. aureus* strains isolated from rabbit farms in the Iberian Peninsula. This analysis revealed a pangenome comprising 3,555 genes, neatly categorized into 2,123 core genes, 94 soft core genes, 626 shell genes, and 712 cloud genes. This classification provided a comprehensive insight into the *S. aureus* genomic diversity.

Phylogenetic analysis yielded valuable insights into the evolutionary lineage of these strains, with clonal complex (CC) CC5 identified as a common ancestor, serving as the origin for multiple strains of other CCs. This discovery underscores the significance of shared ancestry and gradual divergence in the evolution of these strains.

Examination of mobile genetic elements (MGEs) uncovered four prophages from the Siphoviridae family, consistently present across CCs. Among the 90 strains, 67 contained plasmids, with the Rep7a type plasmid being the most frequent and CC5 like the CC with a greater number of plasmids. A total of 30 distinct insertion sequence (IS) types were identified, with ISL3 and IS1182 being common, especially in CC398, which exhibited significant differences. All strains carried integrative and conjugative elements (ICEs), with Tn6098 being the most frequent. Remarkably, CC5, CC96, and CC398 exhibited the highest ICE counts among the strains, highlighting significant differences in terms of the total number of EGMs.

This comprehensive genomic and phylogenetic analysis of *S. aureus* strains illuminates their extensive genetic diversity and provides critical insights into their evolutionary trajectory. Mobile genetic elements emerge as pivotal determinants of genetic variability within these strains, potentially influencing antibiotic resistance and pathogenicity[2]. These findings significantly contribute to our understanding of the biology and epidemiology of *S. aureus*, offering indispensable insights for future research.

Keywords: Staphylococcus aureus, MGEs, evolution and pangenome

References:

1) Haag AF et al. (2019). *Staphylococcus aureus* in Animals. Microbiol Spectr.. doi: 10.1128/microbiolspec.
2) Malachowa N & DeLeo FR (2010). Mobile genetic elements of *Staphylococcus aureus*. Cell Mol Life Sci.;67(18):3057-71. doi: 10.1007/s00018-010-0389-4.

25.

Low plasmid mobility benefits cheaters and prevents eradication in nature

Akshay Sabnis[1*], Wendy Figueroa[1], Alfonso Santos-Lopez[2], Alvaro San Millan[2], José R. Penadés[1]

[1] Centre for Bacterial Resistance Biology, Department of Infectious Disease, Flowers Building, Imperial College London, Exhibition Road, London. [2] Centro Nacional de Biotecnología–CSIC, Madrid, Spain

akshay.sabnis13@imperial.ac.uk

Plasmids have classically been considered to be highly-mobile contributors to horizontal gene transfer in bacteria, readily transferring antimicrobial resistance determinants and virulence factors that compromise treatment of pathogens. However, the transfer efficiency of plasmids by phage transduction – one important mechanism of plasmid mobility – is markedly lower than that of other mobile genetic elements, and the evolutionary basis for this remains unclear.

Here, we show using *Staphylococcus aureus* as a model organism that plasmids have evolved low mobility because they lack phage packaging sites, although the addition of these "*pac*" or "*cos*" sites increased transfer efficiency with no impact on plasmid fitness. Moreover, we demonstrate high plasmid transduction can be easily evolved in the laboratory through serial passaging in the presence of a phage. Sequencing revealed the mechanism of increased transferability to be incorporation of phage DNA into the plasmids, which in some cases dramatically impacted the capacity of the phage to be induced. Crucially, there was no negative evolutionary effect of highly-transducible evolved plasmids in terms of their stability or burden on host bacterial cells, suggesting an alternative reason why natural *S. aureus* plasmids restrict their transfer efficiency.

We propose that in mixed bacterial communities, the sharing of "common goods" encoded on plasmids can provide protection from external threats to the whole population even in the absence of plasmid mobility, and provide experimental proof of this concept using anti-phage systems and antibiotic resistance genes carried by plasmids. Furthermore, we highlight that this community-level protection works optimally when plasmids are present in the population at equal proportions, and is compromised in the event that a single dominant plasmid takes over the community. Thus, we provide an ecological scenario to explain the benefits of plasmid mobility by phages being low, showing that reduced transfer is key to maintain variability and prevent plasmid eradication.

Keywords: plasmids, evolution, MGEs, transduction

26.

Unraveling the role of plasmids in disseminating prokaryotic immune systems across different ecosystems

Mario R. Mestre, Haotian Zheng, Rafael Pinilla-Redondo, Jakob Russel, Søren Johannes Sørensen

Section of Microbiology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark

mario.mestre@bio.ku.dk

Plasmids play a crucial role as facilitators of gene and function dissemination in bacterial and archaeal ecosystems, exerting significant influence on their evolutionary and ecological dynamics. While extensive research has focused on their role as carriers of antibiotic resistance or virulence genes, their potential contribution to the transfer of other essential functions remains largely unexplored. One of the main ecological stresses faced by prokaryotic organisms in different ecosystems is predation by viruses. Recent work has revealed a substantial fraction of accessory genes in bacterial genomes is dedicated to immune functions[1], and a growing body of research is showing that defense genes are frequently mobilized by mobile genetic elements[2]. However, the role of plasmids in this phenomenon remains elusive. In this study, we conducted a comprehensive analysis of bacterial defense system components across publicly available genomic and metagenomic plasmid databases, focusing on characterizing their prevalence, diversity, and distribution. Our findings reveal a substantial proportion of plasmids encoding defense systems, emphasizing the evolutionary and ecological importance of plasmids in disseminating defense mechanisms across prokaryotic communities.

Keywords: Plasmids, prokaryotic immune systems, metagenomics, comparative genomics

References:

1) Hochhauser, Dina et al. (2023) "The defense island repertoire of the Escherichia coli pan-genome." *PLoS genetics* https://doi.org/10.1371/journal.pgen.1010694

2) Chevallereau, A., & Westra, E. R. (2022) “Bacterial immunity: Mobile genetic elements are hotspots for defence systems” *Current Biology* https://doi.org/10.1016/j.cub.2022.07.075

27.

Lateral transduction is the main route of horizontal gene transfer in immunocompetent bacteria

Wendy Figueroa, Akshay Sabnis, Jose R. Penadés

CBRB, Imperial College London, Exhibition Rd, South Kensington, London SW7 2BX

w.figueroa-chavez@imperial.ac.uk

It has been classically assumed that horizontal gene transfer (HGT) is a process primarily mediated by mobile genetic elements (MGEs) such as plasmids, ICEs, bacteriophages or phage satellites. HGT significantly impacts microbial evolution by expanding and redistributing the gene pool, thus contributing to diversity within populations. Importantly, it has been recently reported that the genes exchanged *via* HGT are not limited to those encoded by MGEs. In fact, it has been demonstrated that the mobilisation rate of chromosomal genes via lateral transduction (LT) can even surpass that of certain MGEs. However, this observation primarily applies to laboratory strains where donor and recipient were genetically identical. In this study, we addressed whether this phenomenon holds true for clinical and environmental strains, which typically harbour an array of mechanisms that impede the entry of invading DNA. We first measured the efficiency of different HGT mechanisms, such as phage lysogeny, conjugation and LT in diverse *Staphylococcus aureus* strains. We demonstrate that LT is the most efficient way of HGT. Conversely, MGEs such as phages and plasmids are heavily targeted by bacterial immune systems, resulting in infrequent transfer. We then investigated the impact of the most abundant immune system in *S. aureus* on HGT, showing that the type 1 Restriction-Modification (RM) system plays a crucial role in blocking the transfer of MGE between strains from different clonal complexes. Moreover, we revealed the presence of natural mutants of this system, which, intriguingly, are not uncommon in *S. aureus*, occurring in approximately 4% of complete genomes. Finally, we explored the evolutionary advantages conferred by mutations in the RM system. We show that they can function as genetic “gateways”, expanding the boundaries of HGT by enabling the transfer of MGEs between different clonal complexes. Altogether, our study reveals that gene exchange is more efficient than the transfer of MGEs, particularly *via* LT, and that bacterial immune systems play a key role in modulating this by blocking MGEs while allowing the transfer of genes.

Keywords: HGT, MGE, Phages, Transduction, RM system

28.

Longitudinal tracking of the ecology and evolution of mobile genes in the honey bee microbiome

Christopher Robinson

Indiana University

robinch@iu.edu

Host-associated microbial communities are influenced by a combination of ecological and evolutionary factors. These factors result in distinct clusters of genotypic and ecological similarity between microbial community members. The observed set of genes carried within each cluster, its pangenome, are known to be impacted by drift, selection, migration, and horizontal gene transfer (HGT). Many bacterial pangenomes show a variation in gene content, and recent debate has centered on the relative importance of evolutionary forces in generating this observed variation. Mobile genes, trafficked through homologoous and non-homologous recombination, often make up a large portion of pangenomic gene gain and loss. Here, we seek to investigate how mobile genes affiliated with one microbial cluster might impact other microbial clusters within the context of a complex bacterial community.Our work utilizes the honey bee gut microbiome, emerging as a model system in microbial community ecology, due to the relatively simple and stable associated bacterial community. We have tracked 12 honey bee colonies from 3 isolated apiaries over 10 months of sampling. Combining bee metagenomic data with hundreds of isolate genomes from several bee-associated bacterial species, we will first utilize a reverse ecology approach to delineate microbial clusters, or populations, within honey bee colonies and study how these clusters change over the course of a single honey bee colony generation. We predict that bacterial populations will undergo niche differentiation, resulting in a wide variation in pangenomic gene content, which will be driven largely by the transmission and association of mobile genes within and between populations. The role of selection acting on these genes will be dictated by several factors, such as effective population size and the degree of host-association, which may further be influenced by ecological forces over time. Further, we will evaluate how pangenomic diversity correlates with ecological diversity over time, contrasting three models of ecological community assembly.

29.

Circularised mobile genetic elements and their eco-evolutionary patterns at ecosystem and global scales

Taylor Priest, Hans-Joachim Ruscheweyh, Marie Schloemerich, Shinichi Sunagawa

[1]Institute of Microbiology, ETH Zurich, Zurich, Switzerland. [2] Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland

tpriest@ethz.ch

The extant biodiversity of life on Earth is a product of evolution acting over billions of years. This evolutionary journey began from a primordial pool of genetic elements. Self-replicating RNA and DNA elements provided the source material for cellular life to evolve, and such elements have continued to act as engines of evolution ever since, through their role as agents of genetic exchange[1, 2]. Of particular significance are circularized mobile genetic elements (cMGEs), which constitute self-replicating entities that often integrate into and excise material from chromosomes. Currently, our understanding of cMGEs is derived from research targeted towards specific environments, host organisms or element types. However, these approaches capture only a fraction of the cMGE pool within the environment. Here, we employ an unsupervised screening of >100,000 metagenomes to recover and characterise the global diversity of cMGEs, the genetic reservoir they encode and their eco-evolutionary patterns. The resource encompasses metagenomes from diverse ecosystem types, from oceans to human gut and extreme environments (e.g. hydrothermal vents). We present here the framework that we are employing along with preliminary insights obtained thus far. To illustrate the scope of our analyses, we recover >220,000 cMGEs from 12,000 ocean metagenomes, with an average length of 14 kbp. Using the recovered circular sequences, we delineate discrete cMGE units of evolution and elucidate the genetic repertoire they possess, employing *de novo* protein structural prediction and distant homology-based approaches to extend current limitations of sequence identity-based annotation. Subsequently, we examine the eco-evolutionary patterns of cMGEs and their encoded functions by assessing their distribution across ecosystems, different axes of variability and predicted host organisms. This work will provide unprecedented insights into the ecological and evolutionary patterns of cMGEs and the mechanisms that underpin them at an ecosystem- and global scale.

Keywords: mobile genetic elements, circular elements, metagenomics, evolution, ecology

References:.

1. Koonin EV, Martin W. On the origin of genomes and cells within inorganic compartments. Trends Genet. 2005;21:647–654.
2. Kazazian HH. Mobile Elements: Drivers of Genome Evolution. Science. 2004;303:1626–1632.

30.

The evolution of domesticated plasmids coincides with species divergence in the plant-associated genus *Pantoea*

Devani Romero Picazo, Tal Dagan

Evolutionary Microbiology Group. Institute for General Microbiology, Kiel (Germany)

dpicazo@ifam.uni-kiel.de

The evolution of vertically inherited secondary replicons via plasmid domestication is common among plant-associated bacteria. Domesticated plasmids encode functions that are beneficial or essential in the plant environment, e.g., functions required for bacterial nodulation and nitrogen fixation in Rhizobia. Consequently, domesticated plasmids become an integral part of bacterial core genomes. However, the evolution and integration of domesticated plasmids within their host lineage remains poorly understood. Here we show that the introduction of domesticated plasmids in *Pantoea* involves the massive acquisition of genes, coinciding with species divergence, and is followed by differential gene loss. Reconstructing the evolution of two plasmids, we infer the Large Pantoea Plasmid (LPP) origin to the genus ancestor, while the origin of the smaller plasmid, pPag1, can be traced back to the ancestor of four plant growth-promoting species. We show that both plasmids evolve vertically and thus they are domesticated plasmids. The LPP plasmid largely contributes to the core genome of the genus, where the allocation of core genes alternates between chromosome and plasmid depending on the species. Such genes primarily belong to the transcriptional functional category. In most replicon-shared gene families, we identified an ancient split between plasmid and chromosome, suggesting a different origin for the plasmid and chromosomal gene copies. Altogether, these results suggest that the massive introduction of xenologous genes after plasmid acquisition, coupled with the differential loss of regulatory genes, is one route for plasmid genome integration into the hosts core pangenome, and thus plasmid domestication.

Keywords: Plasmids, plant-association, diversification, xenologs, core-genome.

31.

Phage co-infection; a mystifying shuffling trick for the pangenome deck

Maliheh Mehrshad

Swedish University of Agricultural Sciences

maliheh.mehrshad@slu.se

The evolutionary trajectory of phages, as obligate intracellular parasites, is locked in a perpetual arms race with their bacterial hosts and the gene flow associated with this further contributes to both phage and bacterial genome complexity. In natural settings, phages outnumber their bacterial hosts; consequently, bacteria are most likely exposed to multiple phage lineages simultaneously. Accumulating evidence, including from my own research, suggest that co-infection that is the simultaneous infection of bacterial host by multiple phages occurs in aquatic ecosystems. However, our whole knowledge of phage ecology and evolution almost exclusively stems from studies that target cells infected by a single phage. This bias further constrains even our existing theoretical frameworks for understanding phage ecology and evolution. Phage co-infection likely plays a significant role in structuring phage/host populations and genomic architecture by providing a greater deck of genes to each partner and a new trick to shuffle and pick from them for horizontal gene transfer. Nevertheless, the prevalence of co-infection in the natural setting is still unknown as well as infection strategies and dynamics that enable and support phage co-infection. Metagenomics have revolutionized our view of microbial ecology and evolution and recovered thousands of phage and bacterial genomes. However, current methods are inefficient is assigning phage genomes to their respective host genomes, thus fail in resolving phage co-infection. I am leveraging my novel and unique cultured model systems where two phages are co-infecting the same host populations to reveal co-infection strategies. I will present a novel perspective on phage-host co-evolution and highlight possibilities for horizontal gene transfer that are facilitated via phage co-infection, as an overlooked avenue for gene flow in the extended pangenome of bacteria.

32.

Homologous host recognition modules in distant families of *Alteromonas* phages

Rafael Gonzalez-Serrano[1], Rosselli, Riccardo[3], Roda-Garcia, Juan J.[2], Martin-Cuadrado, Ana-Belen[3], Dunne, Matthew[4], Rodriguez-Valera, Francisco[2]

[1] Centro de Biología Molecular Severo Ochoa. [2] Universidad Miguel Hernández. [3] Universidad de Alicante. 4 ETH Zurich

r.gonzalez@cbm.csic.es

The abundance and diversity of phages in nature is an indication of the evolutionary success of these microorganisms. Random illegitimate recombination generates novel phage genomes over time and HGT plays a key role in phage evolution. Here we have identified the new *Alteromonas* podovirus A5, with unexpected similarity in its host recognition module with others found in *Alteromonas* phages with different morphologies and belonging to distant viral families. The genomic comparison among these phages revealed a conserved module including a homologous receptor binding protein (RBP) and tail fiber chaperone. In contrast, the rest of the genome was completely different. Binding assays with A5 demonstrated that gp8 is the RBP in this podovirus and is required for host recognition, but gp9, identified as a tail fiber chaperone, is not essential for host binding. In contrast, in the case of the *Alteromonas* myovirus V22, which contains the same recognition module, the tail fiber chaperone is required for RBP maturation and thus for proper host recognition. The results obtained here suggest that recombination of host recognition modules among unrelated phages may be more frequent among the entire community of tailed phages than previously thought and a major mechanism of adaptation to host evasion by target import by horizontal gene transfer. To our knowledge, this is the first time that this level of conservation in phage-host recognition proteins has been detected at the same time in marine phages belonging to the Myoviridae, Podoviridae, and Siphoviridae families, with no other closely related gene in the rest of the genome.

33.

The evolution of *Arabidopsis thaliana*-associated *Pseudomonas*

Ashkenazy H, Shalev O, Karasov T, Neumann M, Weigel D

Max Planck Institute for Biology – Tuebingen, Department of Molecular Biology, Tuebingen, Germany

haim.ashkenazy@tuebingen.mpg.de

Members of the *Pseudomonas* genus of gram-negative bacteria are often highly abundant in the plant microbiota. While some species or isolates have been described as pathogenic, others are beneficial to host plants or merely commensals. In this study, we focus on the evolution of a focused collection of plant-associated *Pseudomonas* strains, to better characterize the factors governing plant colonization by members of this taxon.

We collected 1,524 *Pseudomonas* isolates from wild *Arabidopsis thaliana* plants, over several years and sites near Tübingen, Germany. We sequenced and assembled the genomes of these bacterial isolates and inferred their pan-genome, based on ortholog groups. Out of the 72,397 ortholog groups specified, only 1.3% belonged to the bacterial 'core genome', while 36.3% were unique to individual isolates - suggesting high variability in isolates' gene content.

Next, we grouped closely-related strains (determined by their genome-wide average nucleotide identity, ANI) and compared the gene content within and between groups. By comparing these closely-related isolates, we discovered genomic islands that varied among isolates and we characterized their function and evolutionary dynamics. We found that pathogenicity levels can vary even among closely-related strains, allowing us to further associate genomic islands with strain pathogenicity. Finally, we describe our strategy for functional tests of genomic islands that are associated with microbe-microbe interactions.

Taken together, our results demonstrate that focusing on the evolution of closely-related strains, which are co-colonizing the same plant population, can shed light on the complex host-microbe and microbe-microbe interactions shaping the plant microbiome.

34.

Modulation of shrimp microbiota using phage therapy during controlled infection model

Carlos Omar Lomeli-Ortega[1,2*], José L. Balcázar[1,2], Eduardo Quiroz-Guzman[3]

[1] Catalan Institute for Water Research (ICRA), 17003 Girona, Spain. [2] University of Girona, 17004 Girona, Spain. [3] Centro de Investigaciones Biológicas del Noroeste S.C. Av. Instituto Politécnico Nacional 195. Col. Playa Palo de Santa Rita Sur, 23096. La Paz, Baja California Sur, Mexico.

olomeli@icra.cat

Aquaculture activities provide protein of high value for human consumption, but due to disease outbreaks an important loss of cultivated organisms has been observed in penaeid shrimp, several fish species, and mollusks farming [1]. Specifically, bacterial infections represent an important problem to aquaculture industry, and considering the growing and global threat of antibiotic resistance, phage therapy represents a suitable alternative against bacterial infections[2]. Here, we evaluate the protective effect of phage therapy in shrimps exposed to vibrio pathogenic mix during *in vivo* trial and the effect of this treatment in the associated microbiota. Shrimp postlarvae treated with phages had the higher survival rate, reduce histological damage and lower total *Vibrio* spp. count. Using high-throughput sequencing analysis of 16S rRNA genes a significantly difference in community structure was found in organisms treated with phages and, the highest bacterial diversity and richness were found in this group too. Taxonomic analyses showed a drastic reduction in members belonging to family *Vibrionaceae* in the phage-treated groups, meanwhile a high abundance was found in untreated groups. Interestingly, members belonging to *Rhodobacteraceae* were the dominant family, regardless of the applied treatment; but a significant increase was observed in groups treated with phages possibly due to *Vibrionaceae* reduction. In conclusion, phage therapy efficiently and selectively inhibited the *Vibrionaceae* populations in shrimp postlarvae, without disrupting other bacterial populations.

Keywords: Microbiota, vibrio species, pathogens, phage therapy.

References:

1) FAO. 2022. The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation. Rome, FAO. https://doi.org/10.4060/cc0461en
2) Pérez-Sánchez, T., et al. (2018). Biological approaches for disease control in aquaculture: advantages, limitations and challenges. Trends in microbiology. https://doi.org/10.1016/j.tim.2018.05.002

35.

Exploring DNA transfer and integration in human cells upon infection with intracellular bacterial pathogens

Sara Samperio, **Pablo Guridi-Fernández**, Ignacio Varela, Matxalen Llosa

IBBTEC, University of Cantabria, Santander, Spain

pablo.guridi@unican.es

Transkingdom horizontal gene transfer and integration from bacteria to plant cells occur naturally from *Agrobaterium tumefaciens*, which sends its DNA through a Type IV Secretion System (T4SS). No such biological example exists with recipient mammalian cells. There are however some hints indicating that DNA transfer from bacteria to human cells could happen and play a role in, for instance, tumoral development [1].

Our lab showed that DNA is transferred from the zoonotic pathogen *Bartonella henselae* to human epithelial cells via its T4SS [2]. Seeking for the fate of this transferred DNA, we performed long-time infections of HeLa cells with *B. henselae*, comparing bacterial strains with or without a functional T4SS, and looked for integration of bacterial DNA into the human genome. To do so, we extracted the genomic DNA, degraded free bacterial DNA, enriched the genomic sample in bacterial DNA by hybridization with a bacterial genomic probe linked to magnetic beads, and then sequenced the resulting genomic sample with Illumina technology.

Upon analysis of the DNA sequence using an in house written java software, called DisPair (DIScordant PAIRs), aimed at detecting breakpoints, we found a much higher number of breakpoints between human and bacterial DNA than between human-human DNA. Approximately the same number of breakpoints (ca. 100) were found with and without the functional T4SS, ruling out a putative role of the T4SS in DNA transfer and integration.

Interestingly, however, we observed a bias in the distribution of the detected DNA insertions. In the human genome, most of the integrations occur in the mitochondrial genome. Regarding *Bartonella* genome, the region encoding its gene transfer agent (BaGTA) and the prophage (BAP) are clearly overrepresented in the integrations, even after correcting by the expected copy number of these bacterial genomic loci. This result opens up the intriguing possibility that these phage related particles could have a role in DNA transfer between *Bartonella henselae* and infected human cells. Literature shows mixed arguments with hints that could support or discard the DNA transfer by those agents. Further research should be carried out to prove the transfer and unravel whether it has any biological role.

Keywords: *Bartonella*, infection, horizonal DNA transmission, gene transfer agent, transkingdom transduction

References:

1) Riley et al. (2013). Bacteria-Human Somatic Cell Lateral Gene Transfer Is Enriched in Cancer Samples. PLoS Comput. Biol. DOI 10.1371/journal.pcbi.1003107.
2) Fernández-González et al. (2011). Transfer of R388 derivatives by a pathogenesis-associated type IV secretion system into both bacteria and human cells. J. Bacteriol. DOI 10.1128/JB.05905-11

36.

Prokaryotic reverse transcriptases. Retroelements converted into specialized defense systems

Sánchez-Nieto E, García-Rodríguez FM, Molina-Sánchez MD, Millán V, Toro N, **Martínez-Abarca F**

Department of Plant and Soil Microbiology, Estacion Experimental del Zaidin, Consejo Superior de Investigaciones Cientıficas, Structure, Dynamics and Function of Rhizobacterial Genomes, Grupo de Ecologia Genetica de la Rizosfera, Spain.

fmabarca@eez.csic.es; esperanza.sanchez@eez.csic.es

Reverse transcriptases (RTs) catalyze the polymerization of DNA from an RNA template. These enzymes were first discovered in RNA tumor viruses in 1970, but it was not until 1989 that they were found in prokaryotes as a key component of retrons. Apart from RTs encoded by the 'selfish' mobile retroelements known as group II introns, prokaryotic RTs are extraordinarily diverse, but their function in the host has remained elusive. However, recent studies have revealed that different lineages of prokaryotic RTs are key components of different lines of defense against phages and other mobile genetic elements [1,2,3]. Examples (presented in this work) of these adapted RTs are found as: (i) components of the so-called retrons with a high variety of associated key genes involved in their antiphage defense mechanism [2]; (ii) associated with CRISPR-Cas systems involved in the adaptation machinery; (iii) Abi-like RTs and other yet uncharacterized RTs.

The diversity of these prokaryotic enzymes are attracting considerable attention, both for use in cutting-edge technologies, such as genome editing, and as an emerging research topic with an important biotechnological potential.

Grants: PID2020-120619RB-I00 of MCIN (Ministerio de Ciencia e Innovación de España), PAIDI -P20_0047 (Junta de Andalucia) with funding from European Union NextGeneration EU (PRTR-C17.I1).

Keywords: reverse transcriptases; retrons; CRISPR-Cas; UG-Abi; Antiviral defense systems.

References:

1) González-Delgado et al. (2021) Prokaryotic reverse transcriptases: from retroelements to specialized defense systems. FEMS Microbiol Rev. DOI: 10.1093/femsre/fuab025.
2) Mestre et al. (2020) Systematic prediction of genes functionally associated with bacterial retrons and classification of the encoded tripartite systems. Nucleic Acids Res. DOI: 10.1093/nar/gkaa1149.
3) Mestre et al. (2022). UG/Abi: a highly diverse family of prokaryotic reverse transcriptases associated with defense functions. Nucleic Acids Res. DOI: 10.1093/nar/gkac467.

37.

"MITES" as a tool to associate microbial host to virus

Francisco Nadal-Molero*, Alicia Campos-Lopez*, Riccardo Rosselli, Silvia García-Juan, Ana-Belen Martin-Cuadrado

*Dpto. Fisiología, Genética y Microbiología. Universidad de Alicante (Spain). *Equal contribution*

francisco.nadal@ua.es, alirousa99@gmail.com

Mobile genetic elements (MGEs) are relevant agents in organism adaptation and evolutionary diversification. They are associated with all domains of life and include elements with diverse replication and mobility strategies. Miniature Inverted-repeat Transposable Elements, MITEs, are transposable elements found in genomes across all domains of life that lack their own transposase and use others present in the genome for their own mobility. As non-completely autonomous elements, we analysed the mobilization boundaries of MITEs across different taxa. Using MITETracker [1], we identified more than 1.4 million MITEs within 58.4% of the prokaryotic genomes examined (n=263 688). Through clustering analyses, it was confirmed that MITE specificity was preserved at the genus level, with only 0.03% of clusters extending beyond these taxonomic boundaries (the MITEsphere frontiers).

Thus, MITE specificity might serve as a valuable taxonomical link among sequences sharing similar MITEs [2]; i.e, the pair Bacteria/Archaea-virus. We identified 1 726 viral MITEs among the databases (Genebank and IMG/VR v4). When the cellular MITEs were used as a query against the viral databases we expanded this group to 5844 MITEs. Clustering of all the cellular and viral MITEs detected allowed to identified 1 245 clusters where a bacterial genome shared the same MITE with viral sequences (encompassing 935 447 MITEs). Finally, it was possible to assign 1868 viral contigs to their hosts. Among them, 1490 cases corresponded to previously known host-virus associations. We conducted in-depth investigations into several novel associations involving viral sequences belonging to putative viruses of important clinical genera, such as *Neisseria, Bacteroides* or *Enterococcus*.

These findings confirm that MITEs are mobilized during the viral infection cycle (favoured in lysogeny). We also conclude that, when MITEs are present, they serve as valuable markers for establishing associations between the biological pair host-virus, expanding the prokaryotic pangenomes. The directionality of these transfers and the roles of the sequences embedded within the MITEs will be investigated.

Keywords: MITEs, MGEs, virus, host

References:

1.Crescente, J. M. et al. (2018). https://doi.org/10.1186/s12859-018-2376-y
2.Zhang et al. 2018 https://doi.org/10.1186/s13100-018-0125-4

38.

Using genome-wide associations studies to identify putative genetic determinants of phage host-range in *Klebsiella*

Janusz Koszucki[1,2*], Vyshakh R. Panicker[1,2], Bogna Smug[1], Aleksandra Otwinowska[3], Sebastian Olejniczak[3], Louise Judd[4], Kathryn E. Holt[4], Edward J. Feil[5], Zuzanna Drulis-Kawa[3], Eduardo P. C. Rocha[6] & Rafal J. Mostowy[1]

1. Malopolska Centre of Biotechnology, Jagiellonian University, Kraków, Poland. 2. Doctoral School of Exact and Natural Sciences, Jagiellonian University, Kraków, Poland. 3. Department of Pathogen Biology and Immunology, Institute of Genetics and Microbiology, University of Wrocław, Wrocław, Poland. 4. Department of Infectious Diseases, Monash University, Melbourne, Victoria, Australia. 5. The Milner Centre for Evolution, Department of Life Sciences, University of Bath, Bath, UK. 6. Institut Pasteur, Université Paris Cité, CNRS, UMR3525, Microbial Evolutionary Genomics, Paris, France

j.koszucki@uj.edu.pl

Phages that infect *Klebsiella pneumoniae* bacteria are known to often rely on a binding between the bacterial surface polysaccharide, such as the capsule, and the viral receptor-binding protein (RBP), such as tail fiber/spike, as the first stage of infection [1][2]. Previous studies have identified that the success of this interaction is an important determinant of the phage host range, both in lytic and temperate phages. However, due to the considerable diversity of both capsular polysaccharides (K-types) and *Klebsiella*-infecting phages, only a small number of sugar-specific phage proteins have been characterised, which hampers our understanding of the evolution of phage host range in *Klebsiella*. To address this knowledge gap, here we leveraged a diverse and representative collection of 3,911 bacterial isolates of the *Klebsiella* genus to study the association between the prophage pangenome and bacterial K-types. We then applied a genome-wide association study approach (linear mixed model of fixed and random effects) to identify protein clusters in prophage regions associated with the most prevalent K-types which might represent novel, previously unidentified RBPs. Finally, we used sensitive prophage genome annotation (HMM-HMM) and AlphaFold2 to verify the predicted targets and, where possible, to compare them to RBPs with experimentally determined specificity. Our results point to three classes of results. First, we found clear examples of associations with proteins which we predicted as a tail fiber (e.g., K27 capsule) or with a tail fiber and another protein with a carbohydrate-binding domain (e.g., K64 capsule). Second, we found examples of associations with short proteins which tend to co-occur with RBPs, like invertases. In some cases, these proteins helped us identify families of mosaic RBPs with homologous C-terminals, sometimes with high structural similarity to experimentally confirmed RBPs (e.g., K25 capsule). Finally, in some cases we curiously failed to find any association (e.g., K21 capsule). Interestingly, some of the predicted RBPs were found within single phage variants with high variation and multiple rearrangements in RBP regions. Overall, our results provide an example of using genome-wide association studies to predict genetic determinants of phage host range in bacteria and to study its evolution in bacterial populations.

Keywords: Klebsiella, phages, RBPs, polysaccharide capsule

References:

1) Latka et al. (2019). Modeling the Architecture of Depolymerase-Containing Receptor Binding Proteins in *Klebsiella* Phages. Front. Microbiol. DOI: 10.3389/fmicb.2019.02649
2) Haudiquet et al. (2021). Interplay between the cell envelope and mobile genetic elements shapes gene flow in populations of the nosocomial pathogen *Klebsiella pneumoniae*. PLoS Biol. DOI: 10.1371/journal.pbio.3001276

39.

Exploring the potential of CRISPR-Cas module transfers between bacteria lineages to overcome Anti-CRISPR inhibition

Riccardo Rosselli, Noemí Marco Guzmán, Javier Espinosa, Francisco Mojica

Dept. of Physiology, Genetics and Microbiology, University of Alicante. Carretera San Vicente del Raspeig s/n. 03690 San Vicente del Raspeig - Alicante, Spain

riccardo.rosselli@gmail.com

CRISPR-Cas systems confer prokaryotes with an adaptive immune response against mobile genetic elements, including viruses and plasmids [1]. These systems typically encompass effector and adaptation modules associated to a CRISPR-array composed of repeated sequences (repeats) interspersed with foreign-derived sequences (spacers). The adaptation module facilitates the integration of spacers into the array. The process, called spacer acquisition or adaptation, creates a memory of past encounters with foreign nucleic acids. Over time, the array expands accumulating a diverse repertoire of spacers that reflects the unique historical interaction between an individual cell and mobile genetic elements. The effector module, on the other hand, detects and neutralizes foreign nucleic acids through base pairing with a CRISPR-RNA (crRNA) derived from the CRISPR-array.

CRISPR-Cas systems are subjected to selective pressure due to the ongoing host-parasite interaction between bacteria and infectious genetic elements. Viral genomes can encode inhibitory molecules, such as Anti-CRISPR proteins, which can interfere with the microbial Cas proteins potentially compromising the effectiveness CRISPR-Cas mediated host defense systems [2].

Our study focuses into the evolution of subtype I-E CRISPR-Cas systems aiming to unveil coevolutionary dynamics between hosts and parasites. We have found phylogenetic discrepancies between adaptation and effector modules and their corresponding coding genomes, indicating that individual module transfers took place throughout evolution spanning across different strains within the same species. This finding supports the hypothesis that modular transfer processes can reshape the layout of CRISPR-Cas subtype I-E enabling evasion of anti-CRISPR mechanisms. Additionally, we expressed *S. enterica* and *E. coli* I-E adaptation modules using both microbes as recipient cells. The data showed that the *S. enterica* module was non-functional in its native microbial strain. However, it induced spacer acquisition in *E. coli*. This outcome highlights a strong negative regulation of *S. enterica* module mediated by inhibitory factor, including potential anti-CRISPR proteins, and suggests that the acquisition of functional *cas* modules might be sufficient to counteract the inhibition.

Currently, our ongoing research aims to further validate whether the transfer of adaptation and effector modules between different bacterial clades represents an evolutionary mechanism that allows microbial cells to overcome CRISPR-Cas evasion strategies.

Keywords: CRISPR-Cas systems, bacteria-phage coevolution, lateral gene transfer

References:

1) Guzmán NM., et al. Digging into the lesser-known aspects of CRISPR biology. International Microbiology volume 24, pages 473–498 (2021).
2) Pawluk A., et al. Anti-CRISPR: discovery, mechanism and function. Nature Reviews Microbiology volume 16, pages 12–17 (2018).

40.

How phages play with LEGO: studying protein-level mosaicism in phages and its evolutionary implications

Bogna J. Smug[1*], Krzysztof Szczepaniak[1*], Stanisław Dunin-Horkawicz[2,3], Eduardo P.C. Rocha[4], Rafał J. Mostowy[1]

[1] Małopolska Centre of Biotechnology, Jagiellonian University in Krakow, Poland. [2] Institute of Evolutionary Biology, Faculty of Biology & Biological and Chemical Research Centre, University of Warsaw, Warsaw, Poland. [3] Department of Protein Evolution, Max Planck Institute for Developmental Biology, Tübingen, Germany. [4] Institut Pasteur, Université Paris Cité, CNRS UMR3525, Microbial Evolutionary Genomics, Paris, France

bogna.smug@uj.edu.pl

Phages exhibit remarkable genetic modularity, which allows their genomes to evolve independently and combine, resulting in astounding diversity [1]. While genome mosaicism in phage populations has been studied, less is known about protein modularity and its impact on viral evolution. To fill this knowledge gap, here we quantified such modularity by detecting instances of protein mosaicism, defined as a homologous fragment between two otherwise unrelated proteins. We used highly sensitive homology detection [2] to quantify protein mosaicism between pairs of 133,574 representative phage proteins and to understand its relationship with functional diversity in phage genomes. We found that diverse functional classes often shared homologous domains which was often linked to extensive protein mosaicism, particularly in receptor-binding proteins, endolysins, and DNA polymerases. We also identified multiple instances of recent diversification via exchange and gain/loss of domains in receptor-binding proteins, neck passage structures, endolysins and some members of the core replication machinery.

To further investigate protein mosaicism in phages, we analyzed the coverage of various protein functions in domain databases such as ECOD and PFAM. We found underrepresentation of some functional groups, particularly structural proteins like tail fibers and tape measure proteins. To address this issue, we created a database of evolutionary conserved fragments (ECFs) of phage proteins. This database contains fragments not covered by known domain databases such as PFAM or ECOD, enabling better examination of the structural architecture and mosaicism of diverse phage proteins, including receptor-binding proteins, and enhancing understanding of rare accessory proteins and underrepresented proteins due to their diversity.

Our study reveals ongoing diversification via shuffling of protein domains, indicative of co-evolutionary arms races and diversifying selection towards circumventing bacterial resistance. Given the extent of protein modularity within phage proteins, we propose representing the phage pangenome as a set of ECFs, which recombine in various ways to create new proteins and genomes. This integrated understanding of protein modularity will provide novel insights into phage evolution and complex interactions with bacteria.

Keywords: bacteriophages, phage evolution, modularity, coevolution

References:

1) Dion, M. B. et.al. (2020) Phage diversity, genomics and phylogeny. Nat Rev Microbiol, DOI: 10.1038/s41579-019-0311-5

2) Steinegger, M. et.al. (2019) HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics, DOI: 10.1186/s12859-019-3019-7

41.

Phages suppress CRISPR-Cas immunity via RNA-based anti-CRISPRs

Sarah Camara-Wilpert[1], David Mayo-Muñoz[2,3,4], Jakob Russel[1], Robert D. Fagerlund[2,3,4,5], Jonas S. Madsen[1,S], Peter C. Fineran[2,3,4,5], Søren J. Sørensen[1], **Rafael Pinilla-Redondo**[1,2,*]

[1]Section of Microbiology, University of Copenhagen, Copenhagen, Denmark. [2]Department of Microbiology and Immunology, University of Otago, Dunedin 9054, New Zealand. [3]Genetics Otago, University of Otago, Dunedin 9054, New Zealand. [4]Maurice Wilkins Centre for Molecular Biodiscovery, University of Otago, Dunedin 9054, New Zealand. [5]Bioprotection Aotearoa, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand.

rafael.pinilla@bio.ku.dk

Many bacteria employ CRISPR–Cas systems to combat mobile genetic elements (MGEs), such as phages and plasmids. In turn, these elements have evolved anti-CRISPR (Acr) proteins to block host immunity. Here, we demonstrate a distinct CRISPR–Cas inhibition strategy based on the deployment of small non-coding RNA acrs (Racrs). Racrs mimic the repeats naturally found in CRISPR arrays and occur in viral genomes as solitary repeat units (SRUs). We show that a prophage-encoded Racr strongly inhibits the type I-F CRISPR–Cas system by interacting specifically with Cas components and resulting in the production of aberrant Cas interference complexes. We identified repeat mimics of almost all CRISPR–Cas types encoded by diverse viruses and plasmids, often in the genetic context of other *acr* genes. Functional testing of Racr candidates spanning the two CRISPR classes confirmed their strong immune inhibitory function. Our results demonstrate that molecular mimicry of CRISPR repeats is a widespread anti-CRISPR strategy, opening the door to potential biotechnological applications.

Keywords: anti-CRISPR, bacterial immune evasion, CRISPR-Cas

42.

Virus-Host Ecological Interactions in the Gut Microbiome Through the Analysis of CRISPR Arrays

López Beltrán, Adrián[1], Iranzo Jaime[2]

[1 2] CBGP - Centro de Biotecnología y Genómica de Plantas (UPM/INIA-CSIC

adrian.lopez.beltran@upm.es

Interactions between viruses and microbes are amongst the most relevant and, paradoxically, least well-understood forces driving microbiome dynamics. Such interactions can be monitored through the study of CRISPR arrays, the part of the CRISPR-Cas system that stores short fragments of foreign DNA (mainly from viruses), serving as a record of encounters between prokaryotes and their viruses. Following that idea, we characterized the dynamics of phage-bacteria interactions and the effects of CRISPR-Cas immunity on virome composition in a large-scale longitudinal study of the human gut microbiome (132 subjects sampled every 2 weeks for 1 year). We developed a fully-automated computational workflow to assemble around 80,000 CRISPR-Cas loci from metagenomic samples, identify and annotate viral targets, and quantify the fluctuations in their abundances. We then studied how phages affect the microdiversity of host populations and how the phage abundances respond to the acquisition of matching sequences in their hosts\' CRISPR arrays. Our analyses reveal that phages impose a selective pressure that favors host lineages with protective CRISPR arrays, although acquisition of immunity does not produce complete selective sweeps in the population of susceptible hosts. On the other side, CRISPR immunity at the whole-microbiome level significantly decreases the abundance of lytic phages in the gut, although that rarely leads to the complete disappearance of the targeted phages. A much smaller effect is observed for lysogenic phages. Notably, we found that spacers matching local phages are concentrated at both ends of the CRISPR array: The proximal end is typically enriched with spacers matching lytic phages, whereas the distal end harbors spacers that match lysogenic phages. This suggests that selection shapes the content of CRISPR arrays, so that the same array can be effective against recent infections by lytic phages and long-term recurrent infections by lysogenic phages. All in all, our study reveals a complex scenario in which CRISPR-Cas immunity modulates the composition of the human gut microbiome, although it does not compromise the long-term coexistence of phages and bacterial subpopulations with diverse immunity profiles.

43.

Ecology of oncogenic human papillomaviruses after the eradication of vaccine-targeted types

Ville N. Pimenoff[1], Penelope Gray[1], Karolina Louvanto[2,3], Tiina Eriksson[3], Camilla Lagheden[1], Anna Söderlund-Strand[4], Joakim Dillner[1] and Matti Lehtinen[1,2]

[1] Karolinska Institutet, Stockholm, Sweden. [2] Tampere University, Faculty of Medicine and Health Technology, Tampere, Finland. [3] Tampere University Hospital, Department of Obstetrics and Gynecology, Tampere, Finland. [4] Department of Clinical Microbiology, Skåne University Hospital, Lund, Sweden.

ville.pimenoff@ki.se

Introduction. Comprehensive human papillomavirus (HPV) vaccine implementation will change the ecological conditions of the virus-host interaction. The long-term effects of HPV vaccine programs can only be estimated with long-term follow-up of randomized trials.

Methods. We performed a long-term follow-up of 33 communities randomized to either gender-neutral HPV16/18 vaccination, girls-only HPV16/18 vaccination and hepatitis B-virus (HBV) vaccination of boys, and control vaccination. In the 1992-94 birth cohorts, 8,618/31,117 eligible boys and 15,615/30,139 eligible girls were vaccinated. Follow-up visits for cervico-vaginal sampling at ages 18 and 22 years were attended by 8,782 and 4,273 participants, respectively. Difference in prevalence of HPV types 6/11/16/18/31/33/35/39/45/51/52/56/58/59/66/68 was assessed and further modeled from the observed data.

Results. Significant decrease of vaccine targeted HPV types 16/18/31/45 and significant increase of low oncogenicity HPV52 and 66 was observed for the gender-neutral vaccination communities eight years post-vaccination compared to control and girls-only vaccinated. Moreover, gender-neutral vaccination associated with increased HPV type-level diversity and likely pangenome diversity difference from four to eight years post-vaccination unlike among the control and girls-only communities.

***Conclusions*.** Eight years post moderate coverage HPV vaccination enabled the detection of a vaccination strategy specific HPVs ecological response. The HPV types increasing after elimination of vaccine-targeted types have low oncogenicity and should thus not challenge the WHO milestone to eliminate cervical cancer. However, the post-vaccination diversity of remaining HPV types will affect future cervical cancer screening programs.

44.

Navigating the *E. coli* Pangenome to Identify Metformin Drivers

Daniel Martinez-Martinez[1,2], Jennifer Van der Laan[3], Filipe Cabreiro[1,2,3]

[1] MRC London Institute of Medical Sciences, Du Cane Road, London W12 0NN, UK. [2] Institute of Clinical Sciences, Imperial College London. [3] CECAD, Cologne Germany

d.martinez@lms.mrc.ac.uk

Metformin is a first-line therapy for type 2 diabetes and a promising anti-aging drug. The microbiome is intimately linked to drug response in the host, and microbes interact with xenobiotics in a variety of ways, including altering their chemical structure and bioactivity, or affecting their bioavailability. This diversity of microbial mechanisms regulating drug action is directly linked to the vast uniqueness of metabolic pathways and the genes present in phylogenetically distinct human gut microbes. This diversity is not only observed between bacterial species, but also between strains of the same species, in what is called a species pangenome. While many studies have now uncovered the role of bacterial species in the mediation of drug effects on the host, very little is known on how accessory genomes present in pan-genomes of a dominant member of gut microbiota affect drug action. To explore the genetic determinants within a pangenome that drive drug activity, we set up a high-throughput platform where we examined more than 750 different strains of *Escherichia coli* and their role in mediating metformin effects on host physiology, using the nematode *Caenorhabditis elegans* as an animal model and drug biosensor. We observed that both host response and bacterial sensitivity to metformin are strain-dependent. Through the development of a novel microbe-host-drug GWAS analysis, we uncovered that drug effects on the host are linked to both the presence/absence of genes or gene variants that are enriched to microbial metabolic pathways such as ATP synthesis and iron metabolism, drug transporters, metabolic signalling pathways such as the phosphotransferase system, or heterologous pathogenic islands. Our high-throughput screening platform paves the way to navigate the complex host-microbe genetic landscape for drug activity, thus improving our ability to predict and act on potential targets in the future.

Keywords: E. coli, metabolism, *C. elegans*

S

V

W

Z