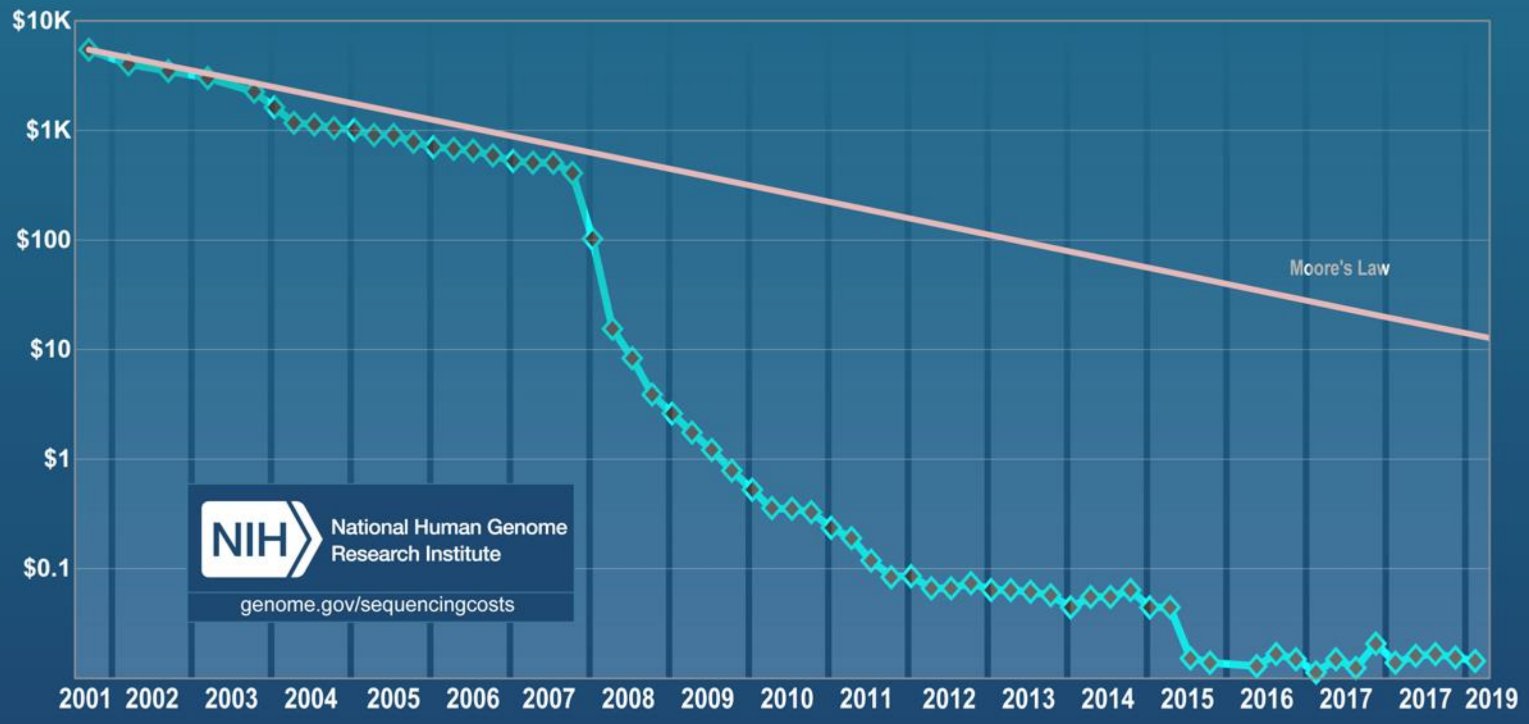
PHASE GENOMICS

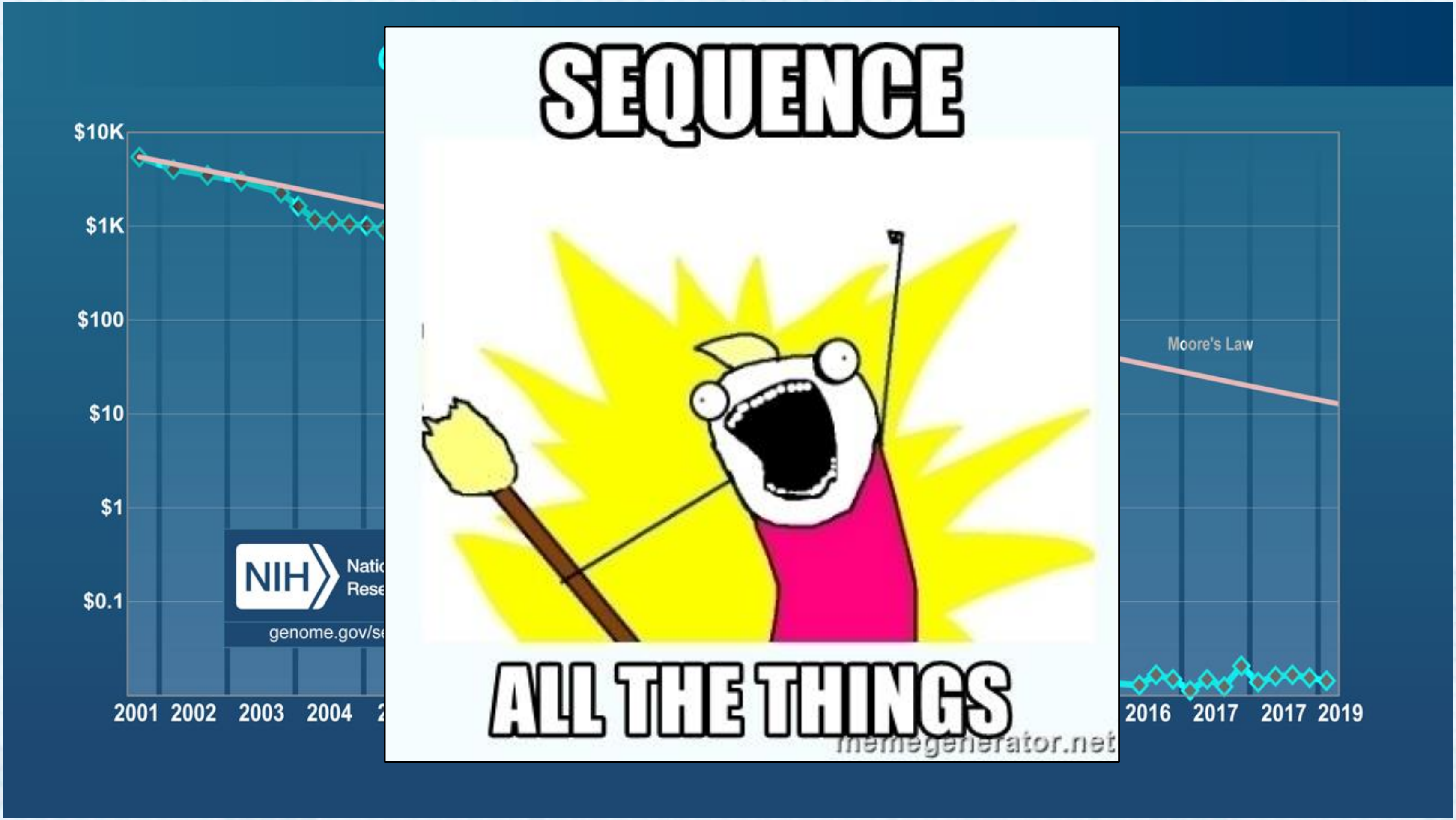
USING PROXIMITY TO FIX ASSEMBLY

Ivan Liachko, Ph.D.
Phase Genomics, Inc.
Seattle, WA, USA

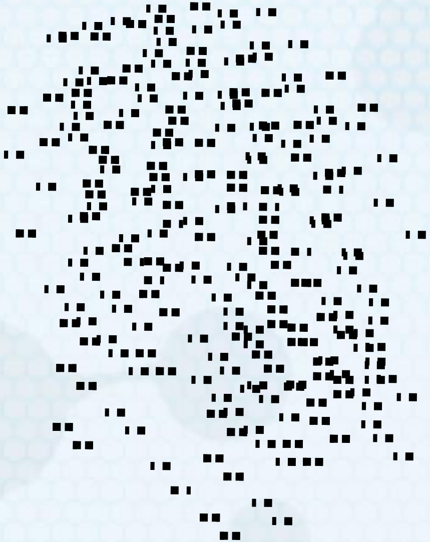
Web: phasegenomics.com
Twitter: @PhaseGenomics

Cost per Raw Megabase of DNA Sequence

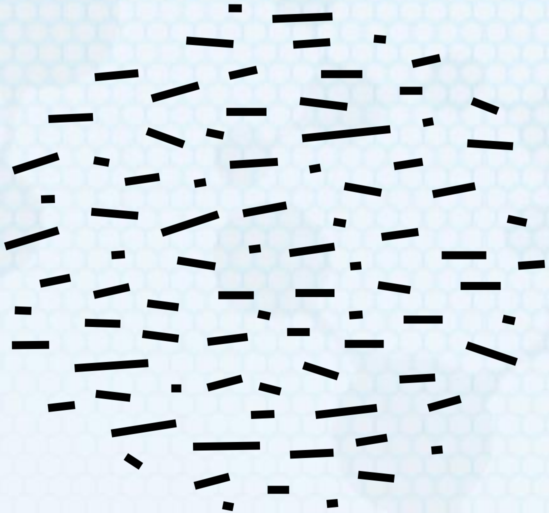



$10K
$1K
$100
$10
$1
$0.1
2001
2002
2003
2004
2016
2017
2017
2019
Moore's Law
NIH Natio Rese
genome.gov/se
SEQUENCE
ALL THE THINGS
memegenerator.net

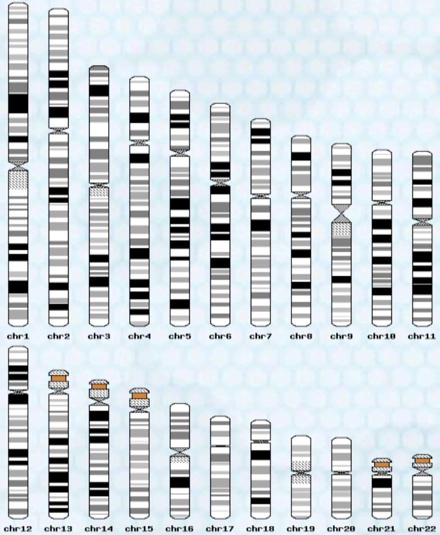
Reads < Contigs < Genome



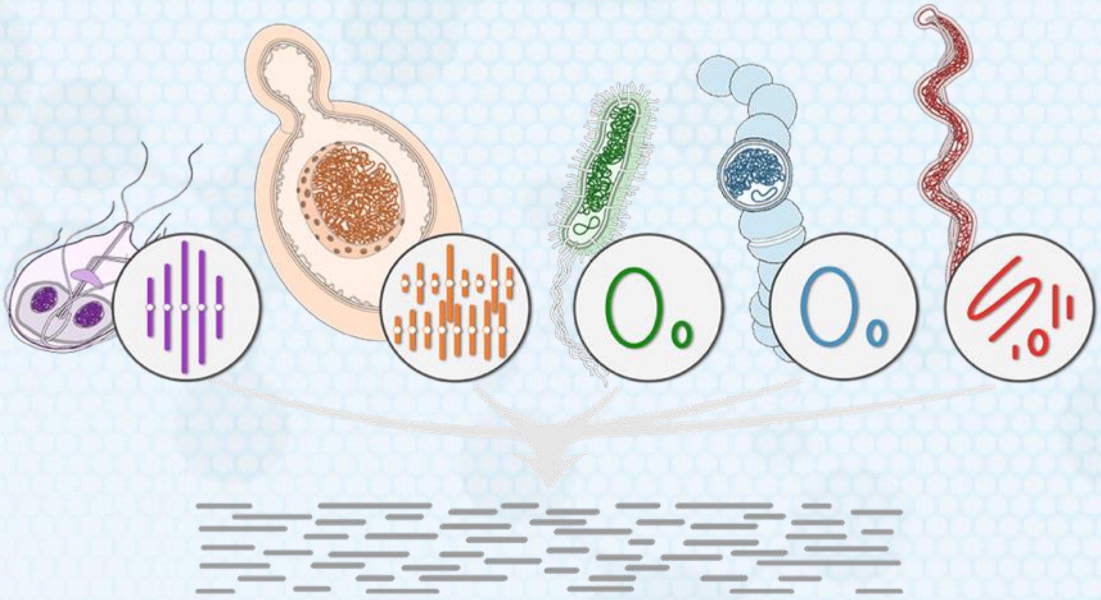
<



<



The problem with shotgun metagenomics

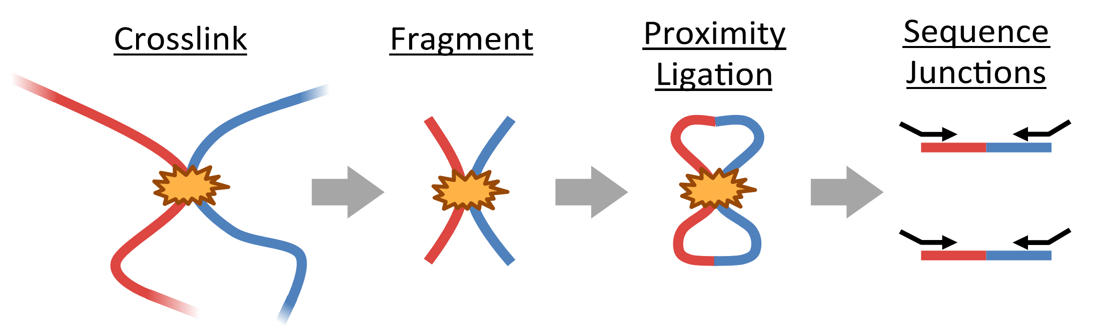


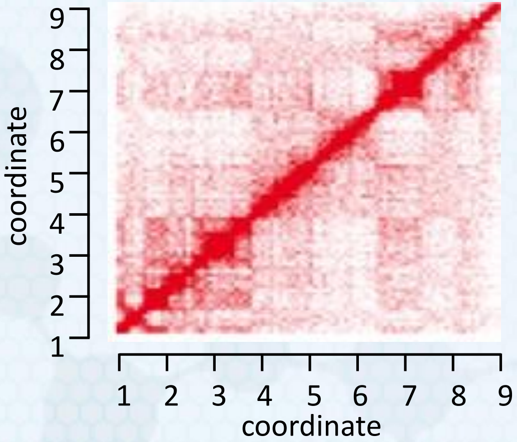
- Cannot tell which sequences belong to which organism
- Binning methods are inaccurate, hard to reproduce
- No way to track plasmids / viruses / antibiotic resistance
- Missing lots of organisms that are not in databases

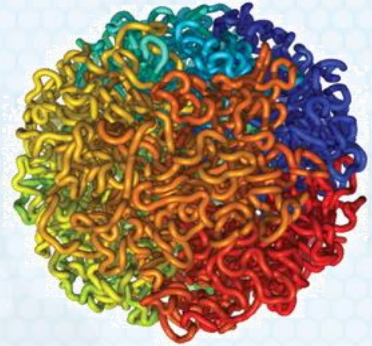
Genomes are packaged into 3D structures

Lieberman-Aiden, *et. al*. Science, 2009

Proximity Ligation (Hi-C) captures the 3D structure of chromosomes

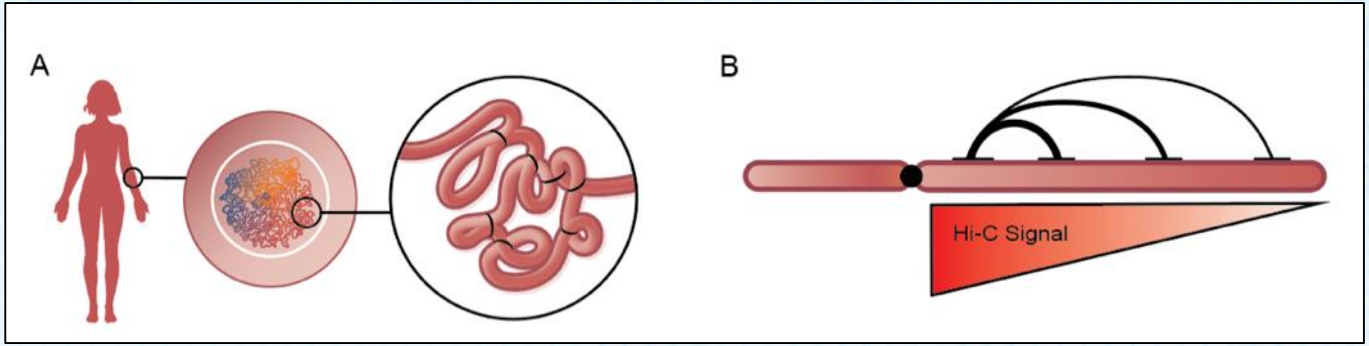






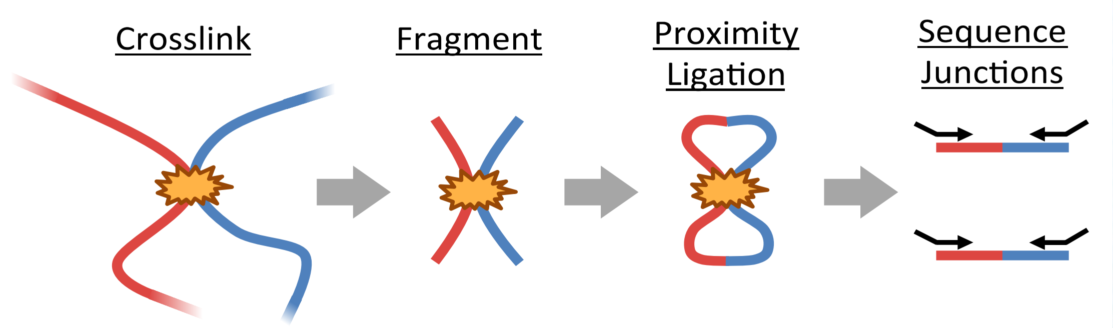
Lieberman-Aiden, *et. al*. Science, 2009

Proximity Ligation captures ultra-long genomic contiguity



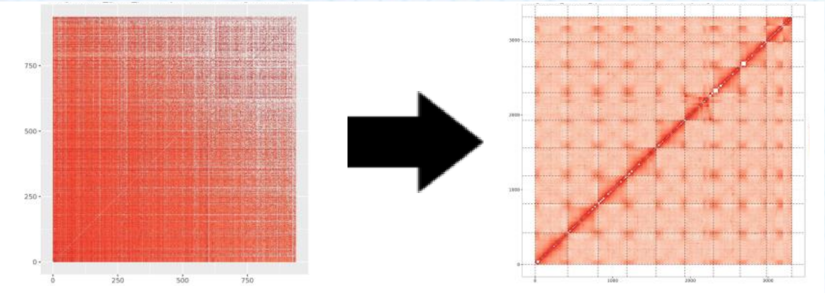
- Proximity in 3D is correlated with genomic distance
- Can be used to:
  - Scaffold and phase a genome of any size
  - Find rearrangements

Using Hi-C to assemble chromosome-scale genome scaffolds



Proximity-Guided Assembly™ :

- Clustering contigs into chromosome groups
- Ordering and orienting the contigs in scaffolds

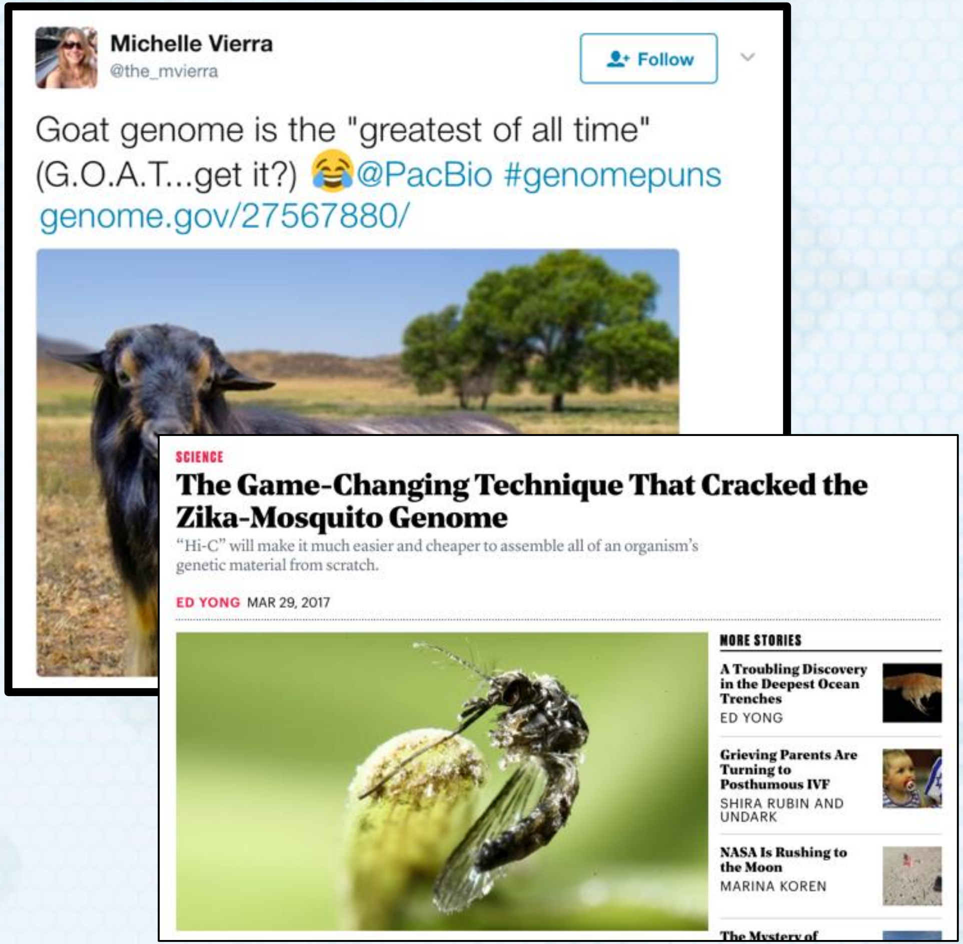


Un-scaffolded contigs

After Hi-C/PGA

Hi-C becomes a routine tool in eukaryotic genome assembly

- Human (Burton *et al*. 2013 *Nature Biotech*)
- Goat (Bickhart *et al.* 2017 *Nature Genetics*)
- Stickleback (Peichel *et al.* 2017 *Heredity*)
- Amaranth (Lightfoot *et al.* 2017 *BMC Biology*)
- Firefly (Fallon *et al.* 2017 *BioRxiv*)
- Black raspberry (Jibran *et al.* 2018 *Hort. Res.*)
- Clownfish (Lehmann *et al.* 2018 *BioRxiv*)
- Sugar beet (Funk *et al.* 2018 *Plant J*)
- Malaria Mosquito (Ghurye et al. 2018 *BioRxiv*)
- *Cannabis* (McKernan et al. 2018 *OSF*)
- *Cannabis* (Grassa et al. 2018 BioRxiv)
- *E. festucae* (Winter et al. 2018 *PLoS Genetics*)
- Honeybee (Wallberg *et al.* 2018 *BioRxiv*)
- Aphid (Chen et al. 2018 *BioRxiv*)
- *T. inflatum* (Olarte, *et al*. 2019 *BMC Genomics*)
- Bee mites (Techer et al. 2019 *BioRxiv*)
- …more from other labs…



Michelle Vierra
@the_mvierra

Goat genome is the "greatest of all time" (G.O.A.T...get it?) 😂@PacBio #genomepuns genome.gov/27567880/

SCIENCE

The Game-Changing Technique That Cracked the Zika-Mosquito Genome

"Hi-C" will make it much easier and cheaper to assemble all of an organism's genetic material from scratch.

ED YONG MAR 29, 2017

MORE STORIES

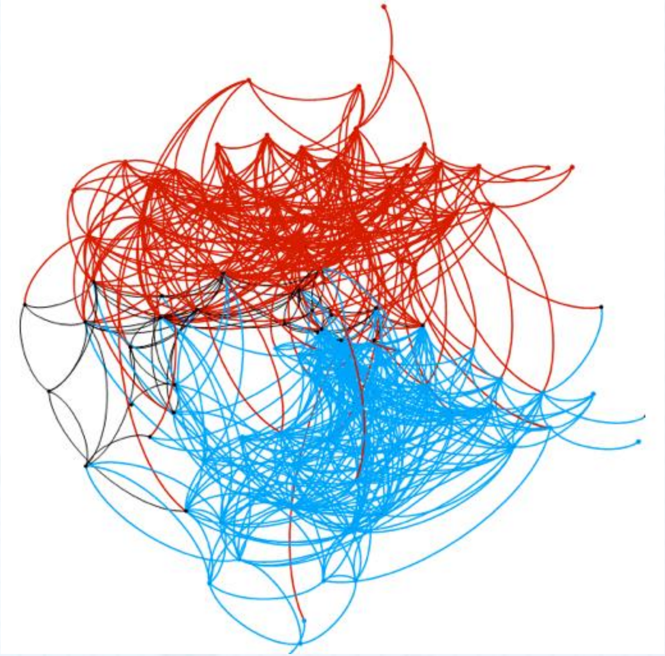
A Troubling Discovery in the Deepest Ocean Trenches
ED YONG

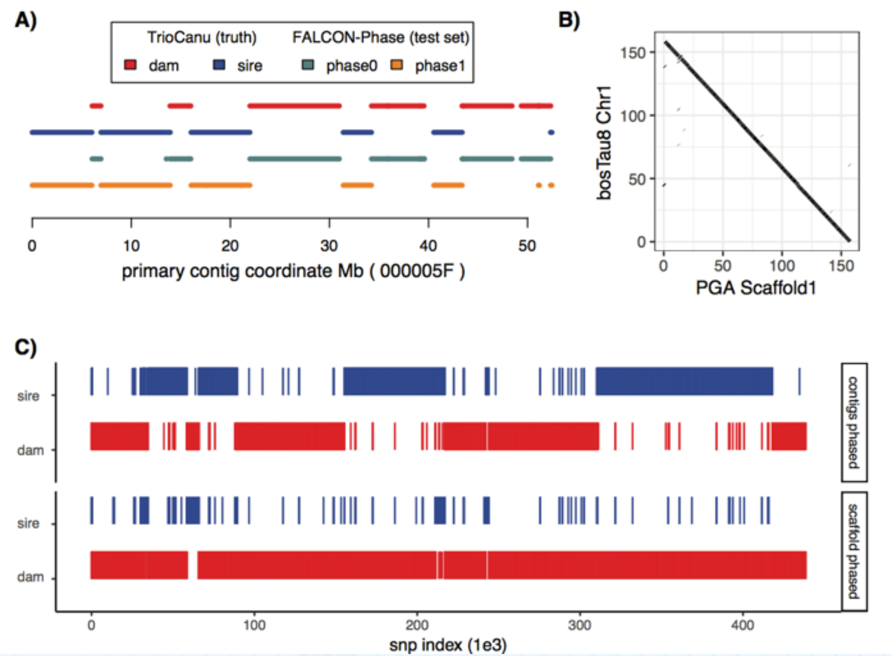
Grieving Parents Are Turning to Posthumous IVF
SHIRA RUBIN AND UNDARK

NASA Is Rushing to the Moon
MARINA KOREN

The Mystery of

FALCON-Phase: Combining SMRT and Hi-C data to generate fully phased genome assemblies.



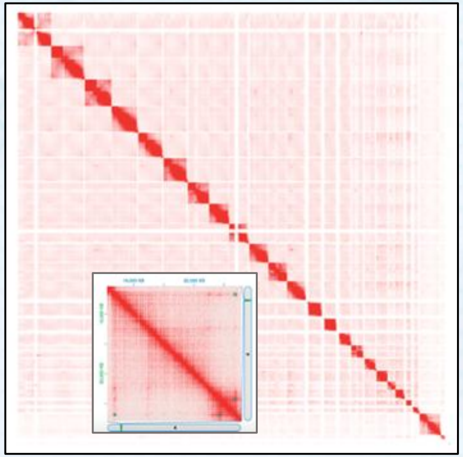


Kronenberg, *et. al.*, *BioRxiv* 2018; In collaboration with Sarah Kingan, Pacific Biosciences

Proximo™ yields high-res karyotype/SV data

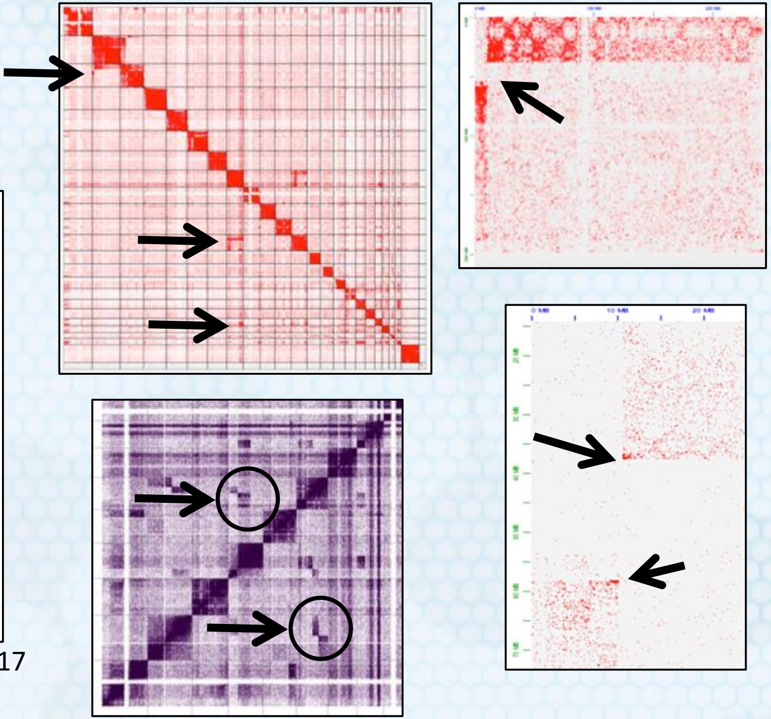
- Rapid
- Scalable
- No special machine
- Works on any sample type

Normal

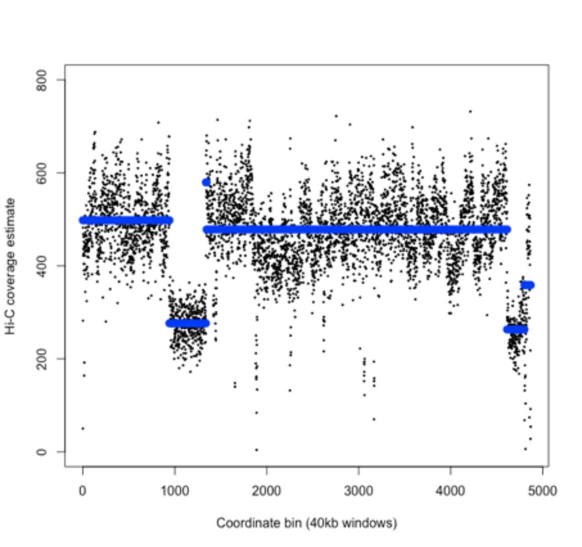


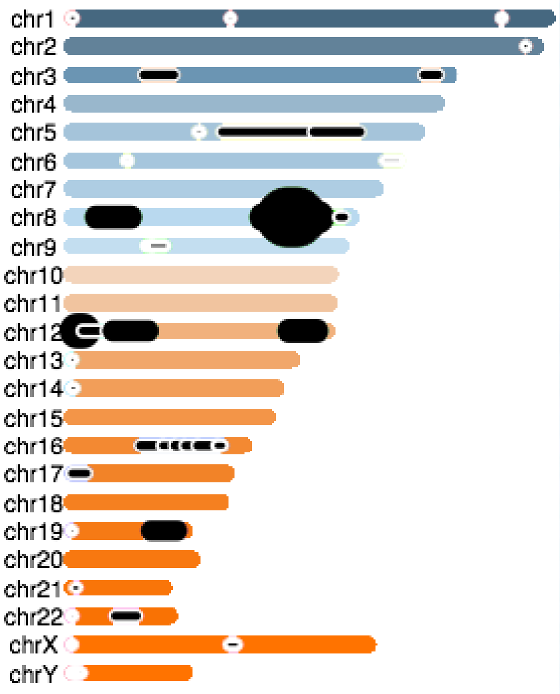
Yardimci and Noble, Genome Biol, 2017

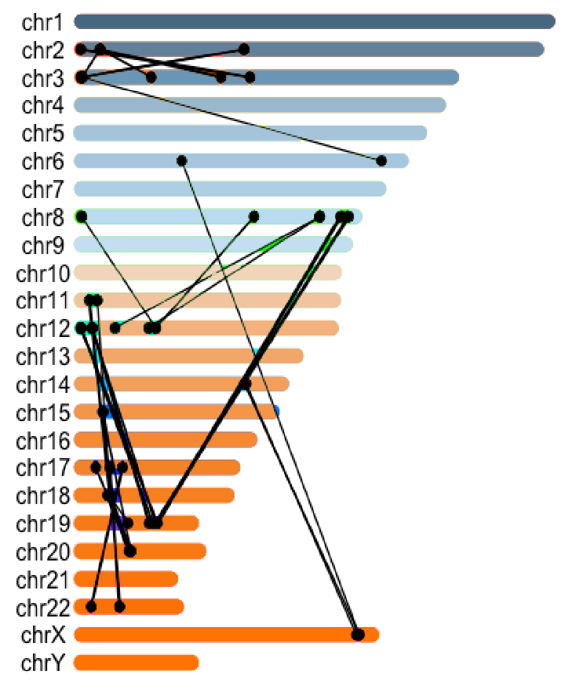
Cancer



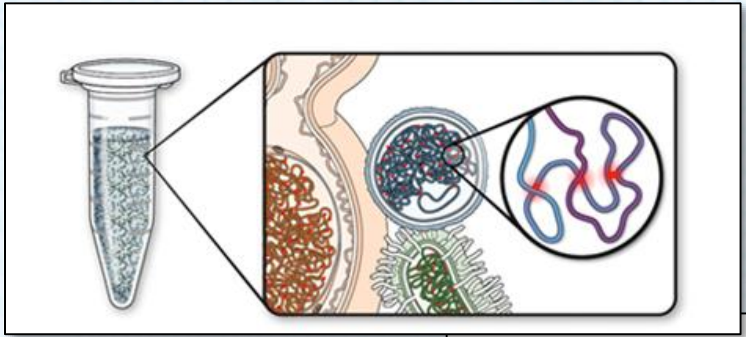
Simultaneous CNV and SV delineation in AML cells

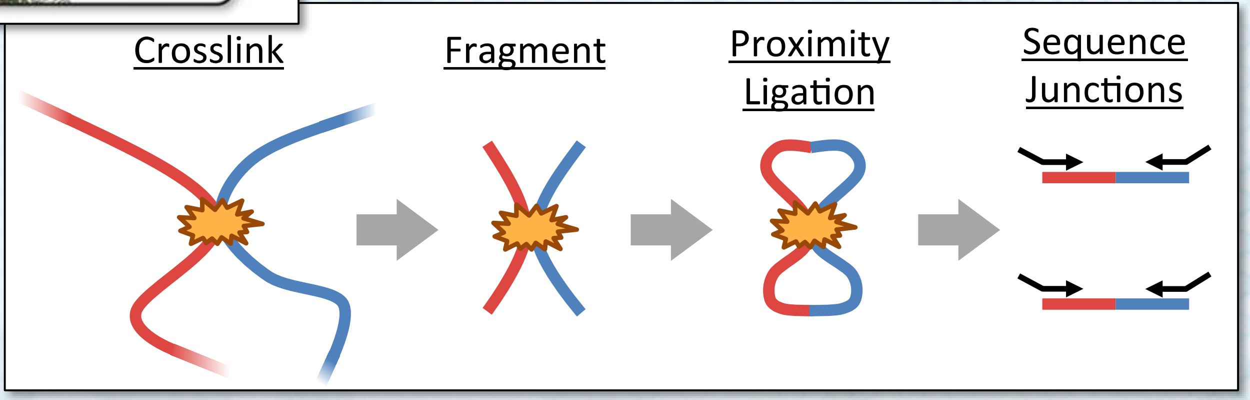


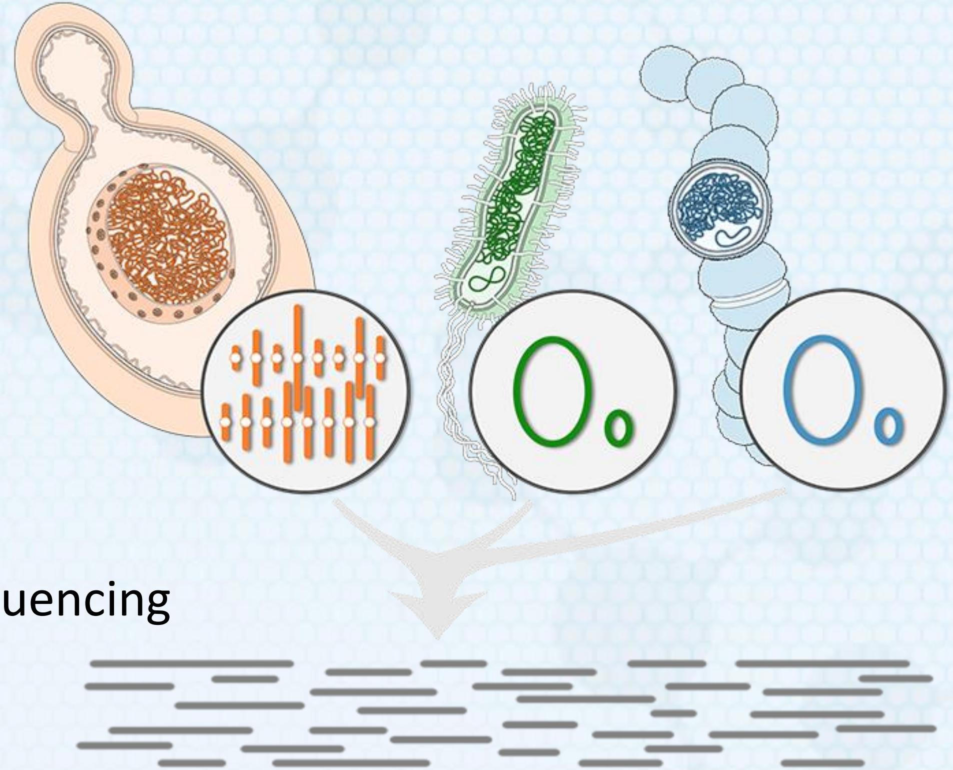




Any sequences that interact by Proximity must have originated from the same cell

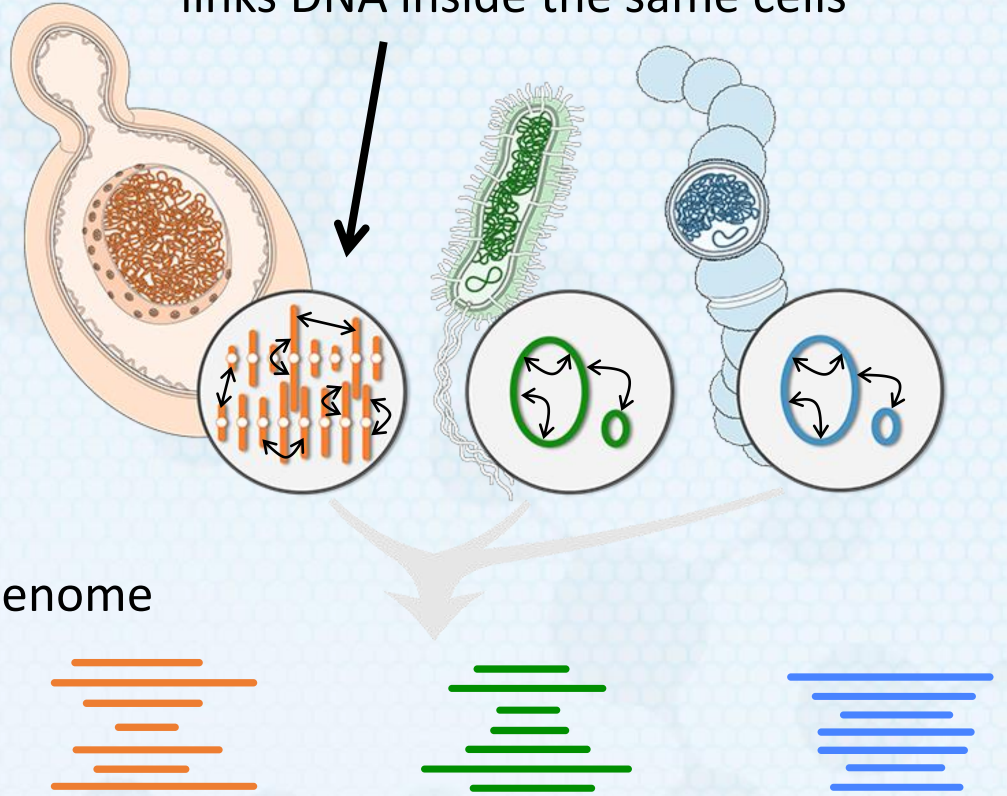






Shotgun sequencing

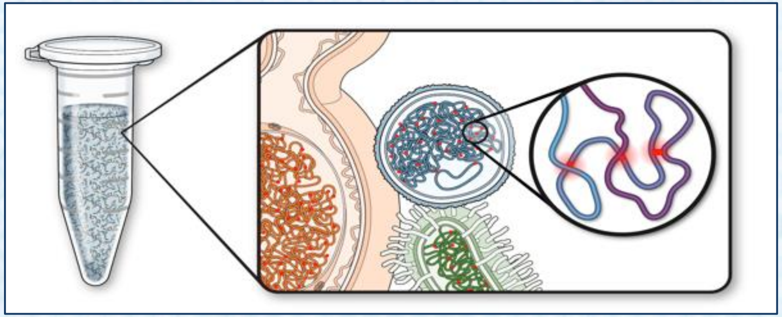
Proximity ligation chemically links DNA inside the same cells



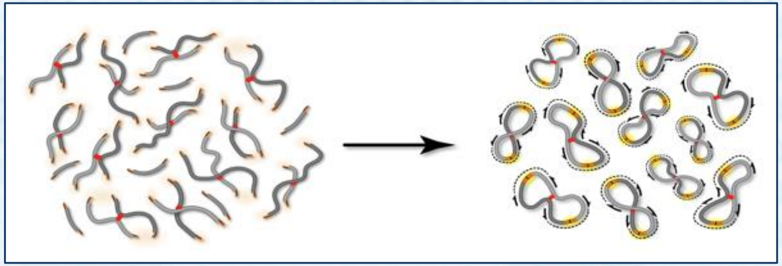
Connects metagenome sequences

Proximity-Guided Metagenome Assembly (ProxiMeta™)

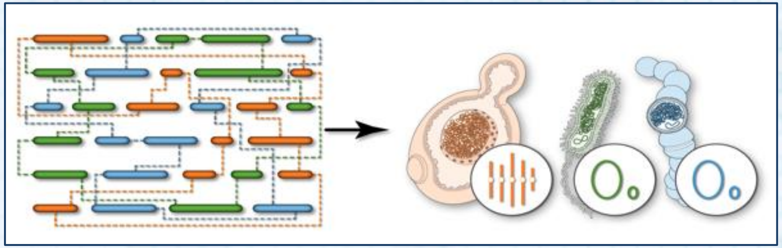
Crosslink intact cells to capture intra-cellular interactions



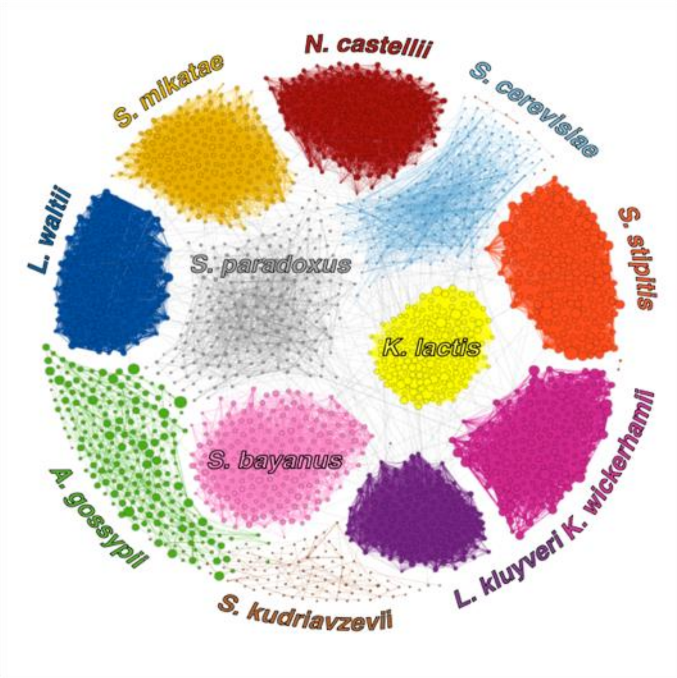
Isolate and sequence crosslinked junctions



Use proximity connections to deconvolute metagenome



Reference-quality pro- and eu- karyotic genomes from mixed populations

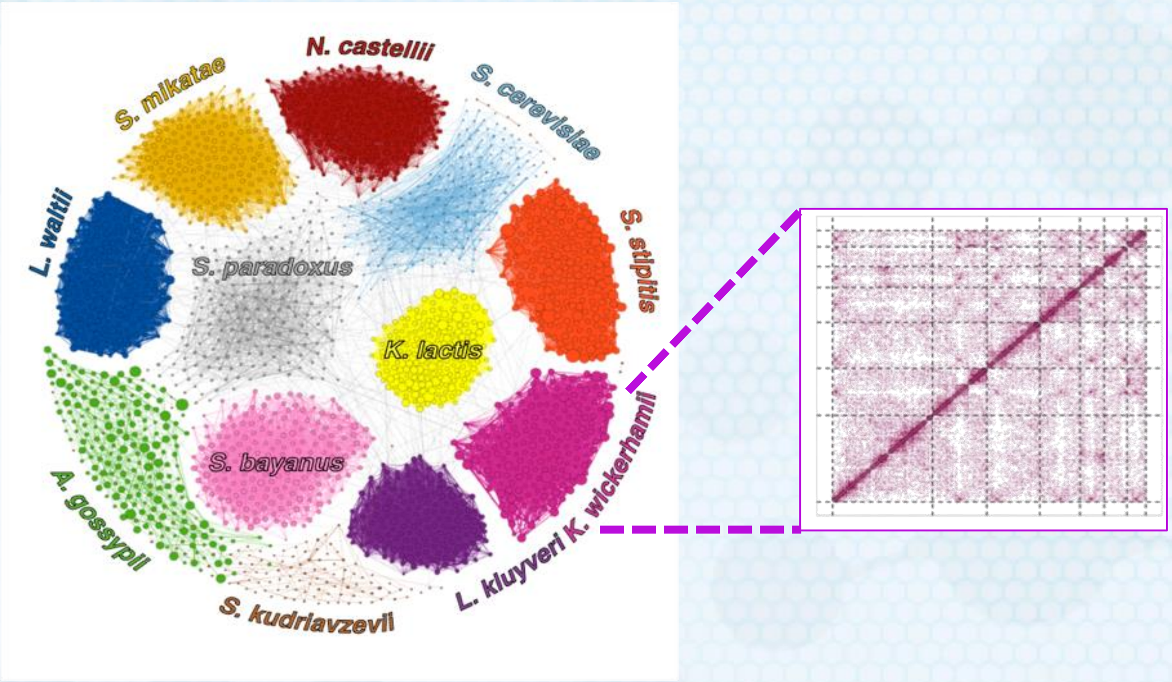


Draft assembly:
Size = 135.2 Mbp
Contig N50 = 17.3 Kb

Error rate <1%

Burton JN, Liachko I *et al*, G3 (2014)

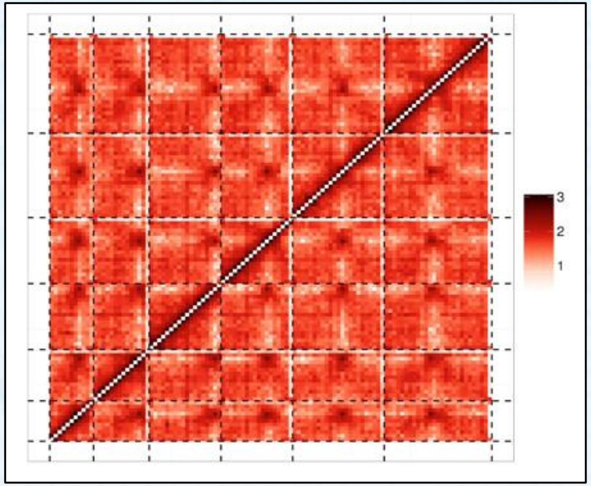
Reference-quality pro- and eu- karyotic genomes from mixed populations

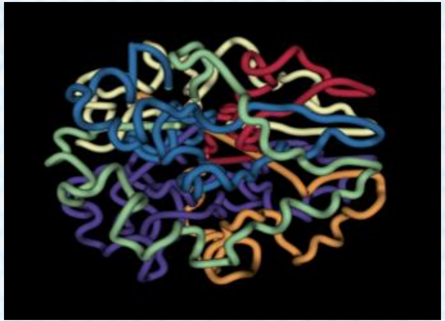


Burton JN, Liachko I *et al*, G3 (2014)

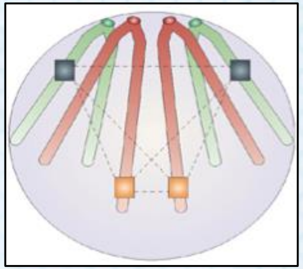
3D modeling of genomes directly from mixed populations

Kluyveromyces lactis





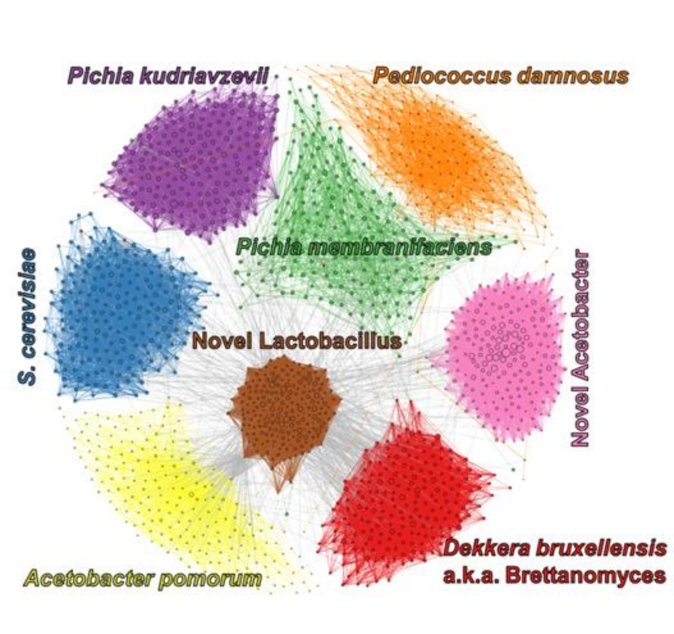
Nelle Varoquaux

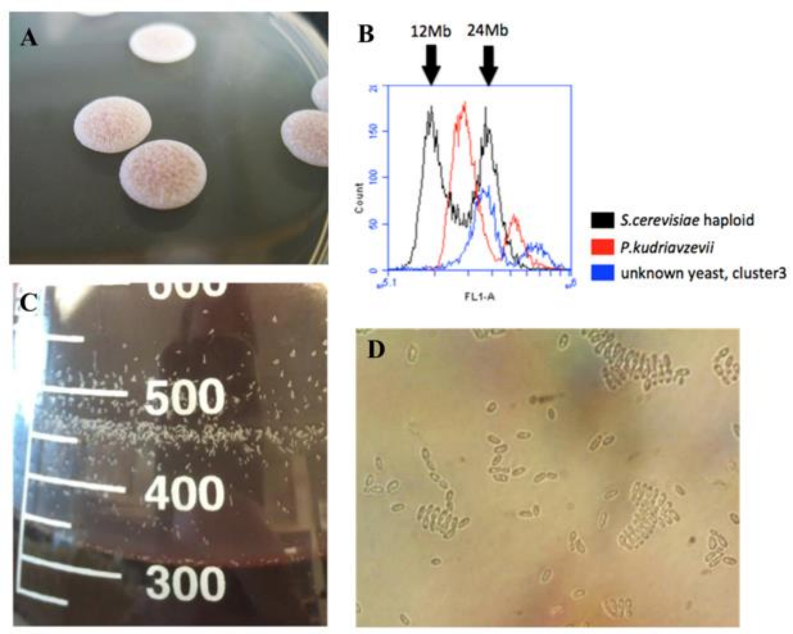


Barzel and Kupiec, 2008

*Dark spots in the middle of each chromosome are centromeric regions.

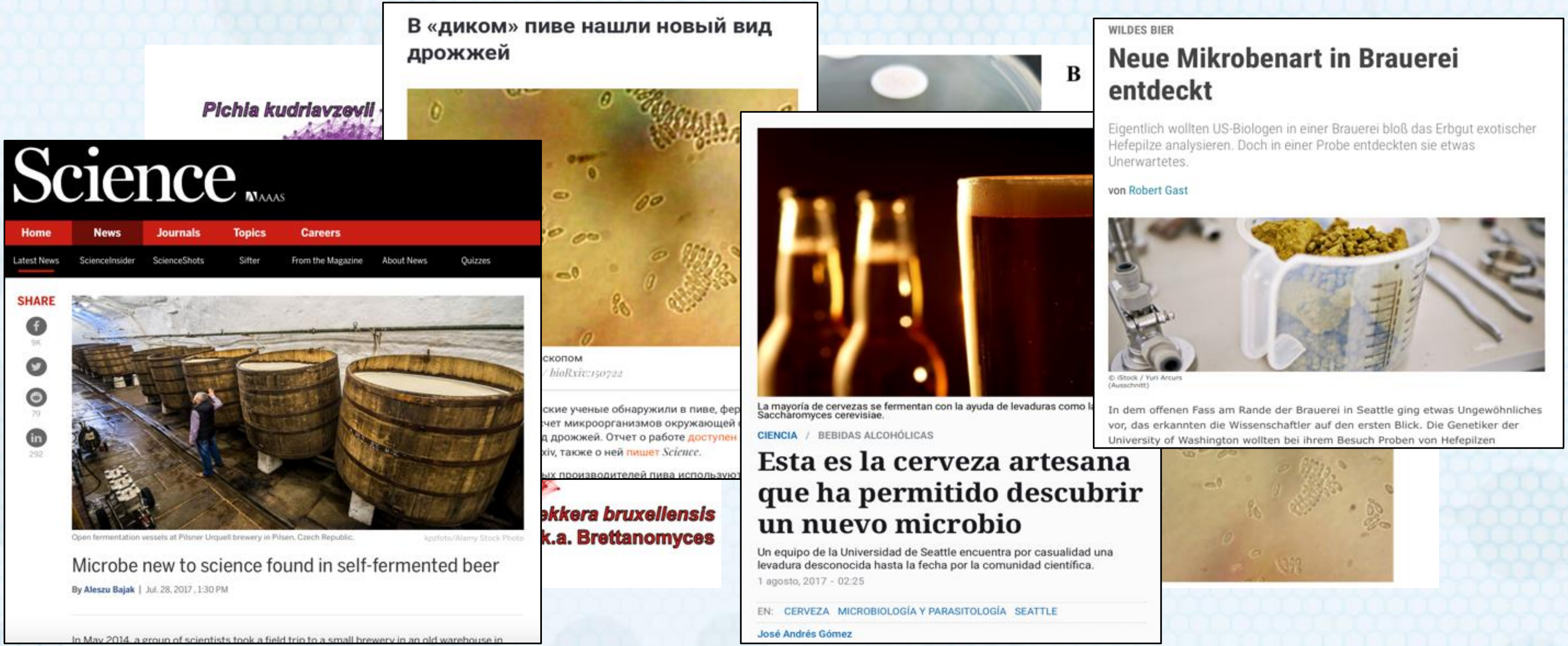
Assembly of a hybrid yeast from a beer metagenome





Heil, Burton, Liachko, *et al.*, Yeast, 2017

Assembly of a hybrid yeast from a beer metagenome



Pichia kudriavzevii

Science AAAS

Home News Journals Topics Careers

Latest News ScienceInsider ScienceShots Sifter From the Magazine About News Quizzes

SHARE 9K 79 292

Open fermentation vessels at Pilsner Urquell brewery in Pilsen, Czech Republic. kpzfoto/Alamy Stock Photo

Microbe new to science found in self-fermented beer

By Aleszu Bajak | Jul. 28, 2017 , 1:30 PM

In May 2014, a group of scientists took a field trip to a small brewery in an old warehouse in

В «диком» пиве нашли новый вид дрожжей

скопом

/ bioRxiv:150722

ские ученые обнаружили в пиве, фер

чет микроорганизмов окружающей

д дрожжей. Отчет о работе доступен

xiv, также о ней пишет Science.

ых производителей пива используют

ekkera bruxellensis

k.a. Brettanomyces

B

La mayoría de cervezas se fermentan con la ayuda de levaduras como la Saccharomyces cerevisiae.

CIENCIA / BEBIDAS ALCOHÓLICAS

Esta es la cerveza artesana que ha permitido descubrir un nuevo microbio

Un equipo de la Universidad de Seattle encuentra por casualidad una levadura desconocida hasta la fecha por la comunidad científica.

1 agosto, 2017 - 02:25

EN: CERVEZA MICROBIOLOGÍA Y PARASITOLOGÍA SEATTLE

José Andrés Gómez

WILDES BIER

Neue Mikrobenart in Brauerei entdeckt

Eigentlich wollten US-Biologen in einer Brauerei bloß das Erbgut exotischer Hefepilze analysieren. Doch in einer Probe entdeckten sie etwas Unerwartetes.

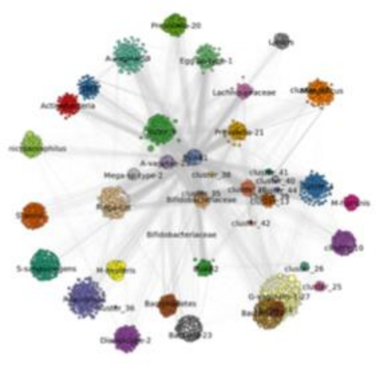
von Robert Gast

© iStock / Yuri Arcurs (Ausschnitt)

In dem offenen Fass am Rande der Brauerei in Seattle ging etwas Ungewöhnliches vor, das erkannten die Wissenschaftler auf den ersten Blick. Die Genetiker der University of Washington wollten bei ihrem Besuch Proben von Hefepilzen

Heil, Burton, Liachko, *et al.*, Yeast, 2017

New genomes and strains from a bacterial vaginosis sample

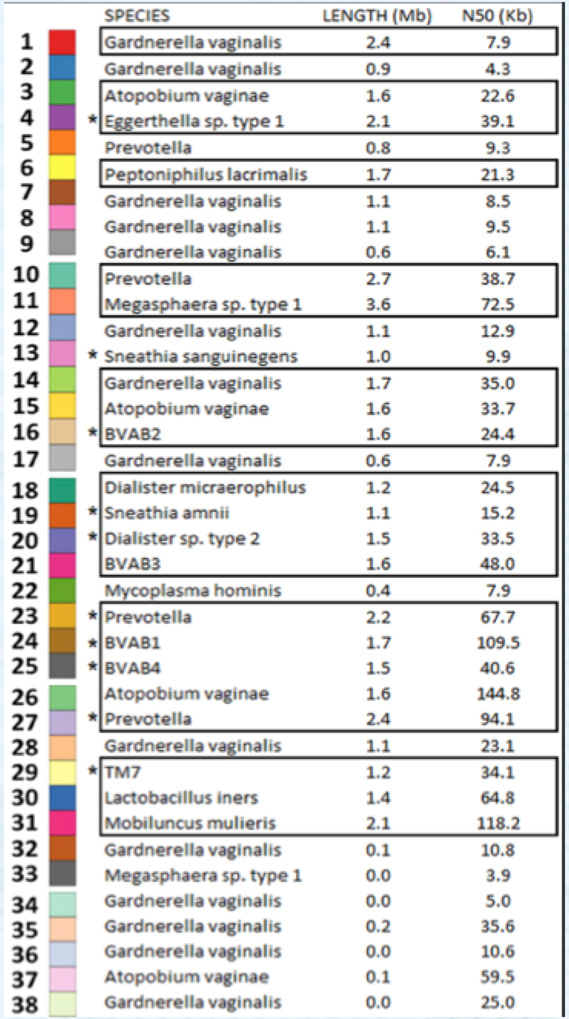


- ProxiMeta clustering of assembly containing all three read sets (N50 ~= 17 kb) yielded >20 high quality draft genomes with >95% core gene groups and N50 >20 kb.

- Least abundant species yielding a high quality genome was represented at 0.2% RA in the combined read set.

- Eight of the high quality genomes came from novel and/or previously unsequenced species (starred).

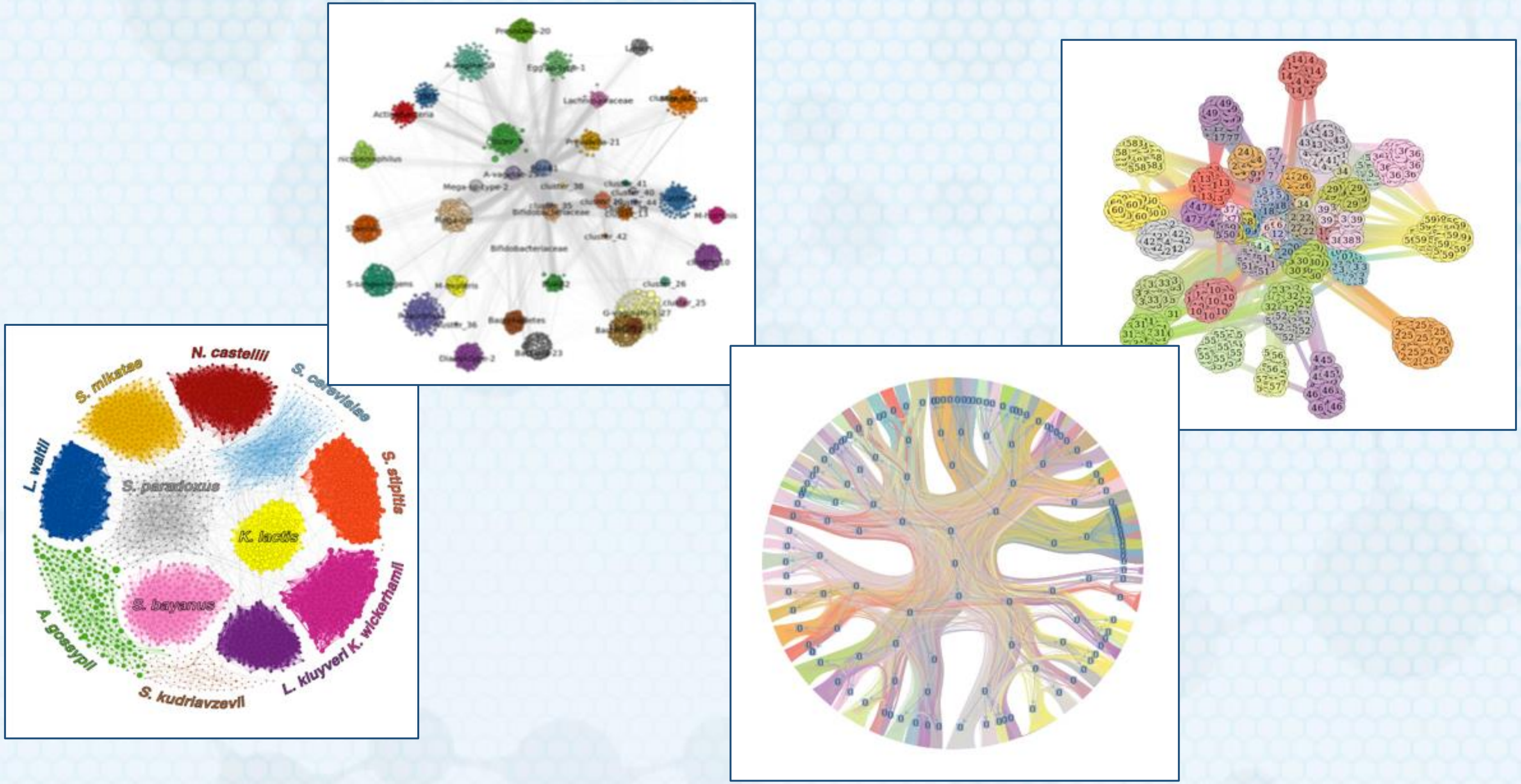
- PGA successfully segregated some strains that differed vastly in protein sequence identity.



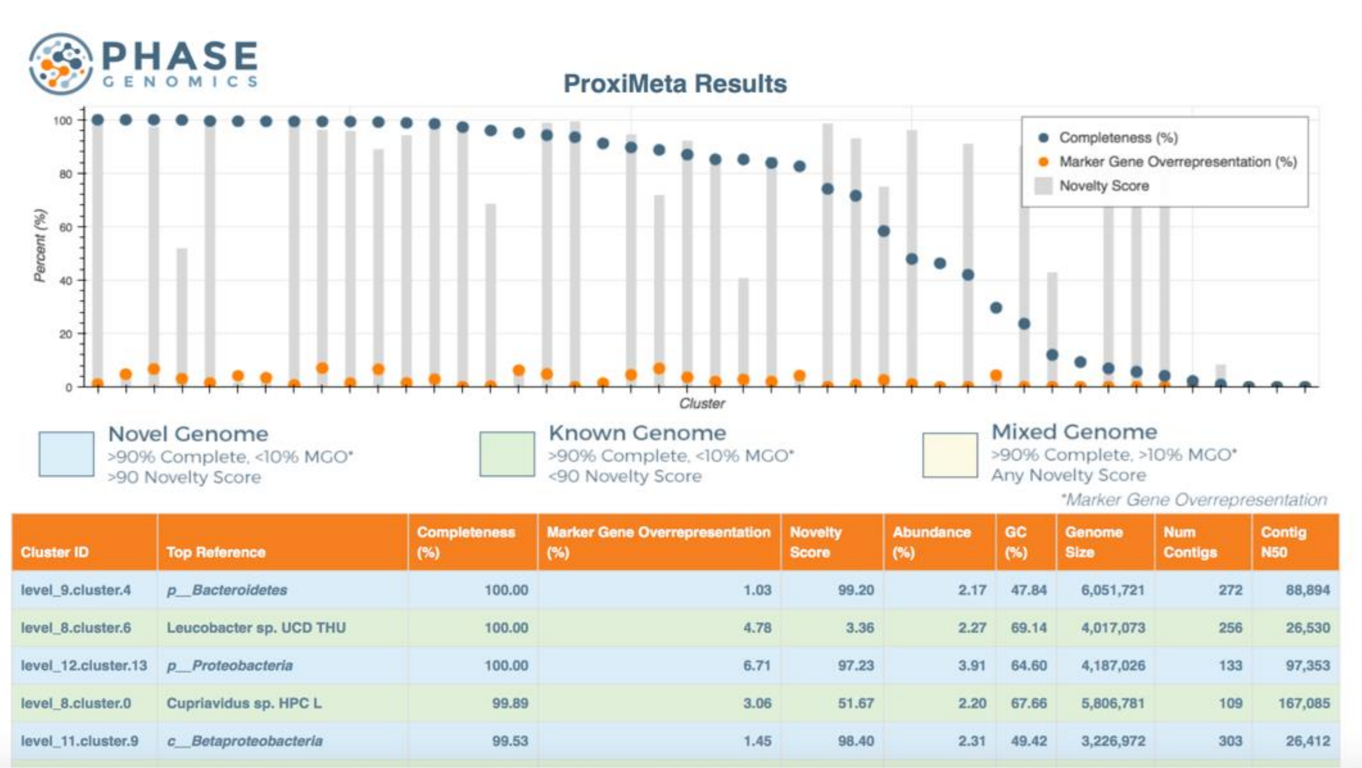
			SPECIES	LENGTH (Mb)	N50 (Kb)
1	🟥		Gardnerella vaginalis	2.4	7.9
2	🟦		Gardnerella vaginalis	0.9	4.3
3	🟩		Atopobium vaginae	1.6	22.6
4	🟪	*	Eggerthella sp. type 1	2.1	39.1
5	🟧		Prevotella	0.8	9.3
6	🟨		Peptoniphilus lacrimalis	1.7	21.3
7	🟫		Gardnerella vaginalis	1.1	8.5
8	🟪		Gardnerella vaginalis	1.1	9.5
9	⬜		Gardnerella vaginalis	0.6	6.1
10	🟩		Prevotella	2.7	38.7
11	🟧		Megasphaera sp. type 1	3.6	72.5
12	🟦		Gardnerella vaginalis	1.1	12.9
13	🟪	*	Sneathia sanguinegens	1.0	9.9
14	🟩		Gardnerella vaginalis	1.7	35.0
15	🟨		Atopobium vaginae	1.6	33.7
16	🟫	*	BVAB2	1.6	24.4
17	⬜		Gardnerella vaginalis	0.6	7.9
18	🟩		Dialister micraerophilus	1.2	24.5
19	🟧	*	Sneathia amnii	1.1	15.2
20	🟪	*	Dialister sp. type 2	1.5	33.5
21	🟥		BVAB3	1.6	48.0
22	🟩		Mycoplasma hominis	0.4	7.9
23	🟨	*	Prevotella	2.2	67.7
24	🟫	*	BVAB1	1.7	109.5
25	⬛	*	BVAB4	1.5	40.6
26	🟩		Atopobium vaginae	1.6	144.8
27	🟪	*	Prevotella	2.4	94.1
28	🟧		Gardnerella vaginalis	1.1	23.1
29	🟨	*	TM7	1.2	34.1
30	🟦		Lactobacillus iners	1.4	64.8
31	🟥		Mobiluncus mulieris	2.1	118.2
32	🟫		Gardnerella vaginalis	0.1	10.8
33	⬛		Megasphaera sp. type 1	0.0	3.9
34	🟩		Gardnerella vaginalis	0.0	5.0
35	🟧		Gardnerella vaginalis	0.2	35.6
36	🟦		Gardnerella vaginalis	0.0	10.6
37	🟪		Atopobium vaginae	0.1	59.5
38	🟩		Gardnerella vaginalis	0.0	25.0

Laura Sycuro, Andrew Wiser, Fredricks lab, Fred Hutch

Reference-quality genomes from mixed populations



Novel, high-completeness genomes from diverse samples



ProxiMeta Results

Novel Genome
>90% Complete, <10% MGO*
>90 Novelty Score

Known Genome
>90% Complete, <10% MGO*
<90 Novelty Score

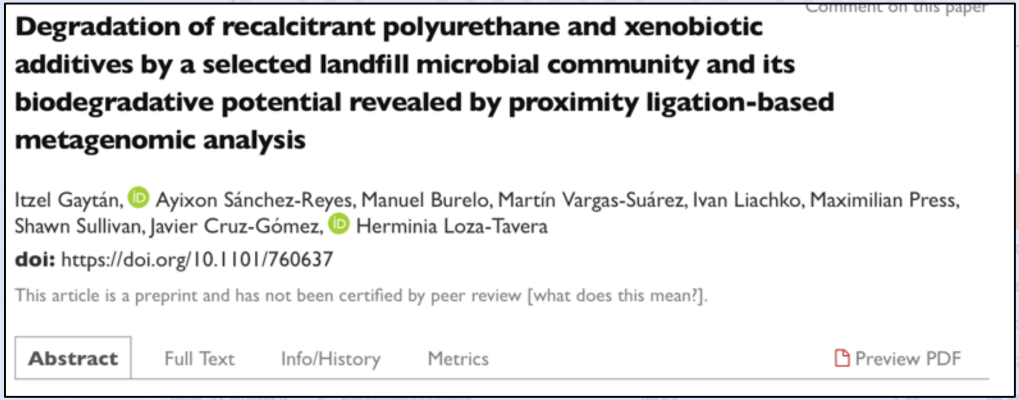
Mixed Genome
>90% Complete, >10% MGO*
Any Novelty Score

*Marker Gene Overrepresentation

Cluster ID	Top Reference	Completeness (%)	Marker Gene Overrepresentation (%)	Novelty Score	Abundance (%)	GC (%)	Genome Size	Num Contigs	Contig N50
level_9.cluster.4	*p__Bacteroidetes*	100.00	1.03	99.20	2.17	47.84	6,051,721	272	88,894
level_8.cluster.6	Leucobacter sp. UCD THU	100.00	4.78	3.36	2.27	69.14	4,017,073	256	26,530
level_12.cluster.13	*p__Proteobacteria*	100.00	6.71	97.23	3.91	64.60	4,187,026	133	97,353
level_8.cluster.0	Cupriavidus sp. HPC L	99.89	3.06	51.67	2.20	67.66	5,806,781	109	167,085
level_11.cluster.9	*c__Betaproteobacteria*	99.53	1.45	98.40	2.31	49.42	3,226,972	303	26,412

In collaboration with Herminia Loza-Tavera, Ayixon Sanchez-Reyez

Novel, high-completeness genomes from diverse samples



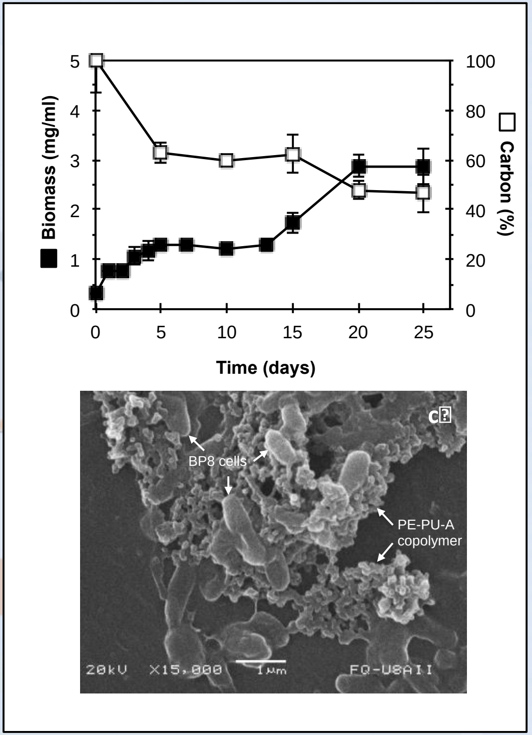
Comment on this paper

Degradation of recalcitrant polyurethane and xenobiotic additives by a selected landfill microbial community and its biodegradative potential revealed by proximity ligation-based metagenomic analysis

Itzel Gaytán, Ayixon Sánchez-Reyes, Manuel Burelo, Martín Vargas-Suárez, Ivan Liachko, Maximilian Press, Shawn Sullivan, Javier Cruz-Gómez, Herminia Loza-Tavera

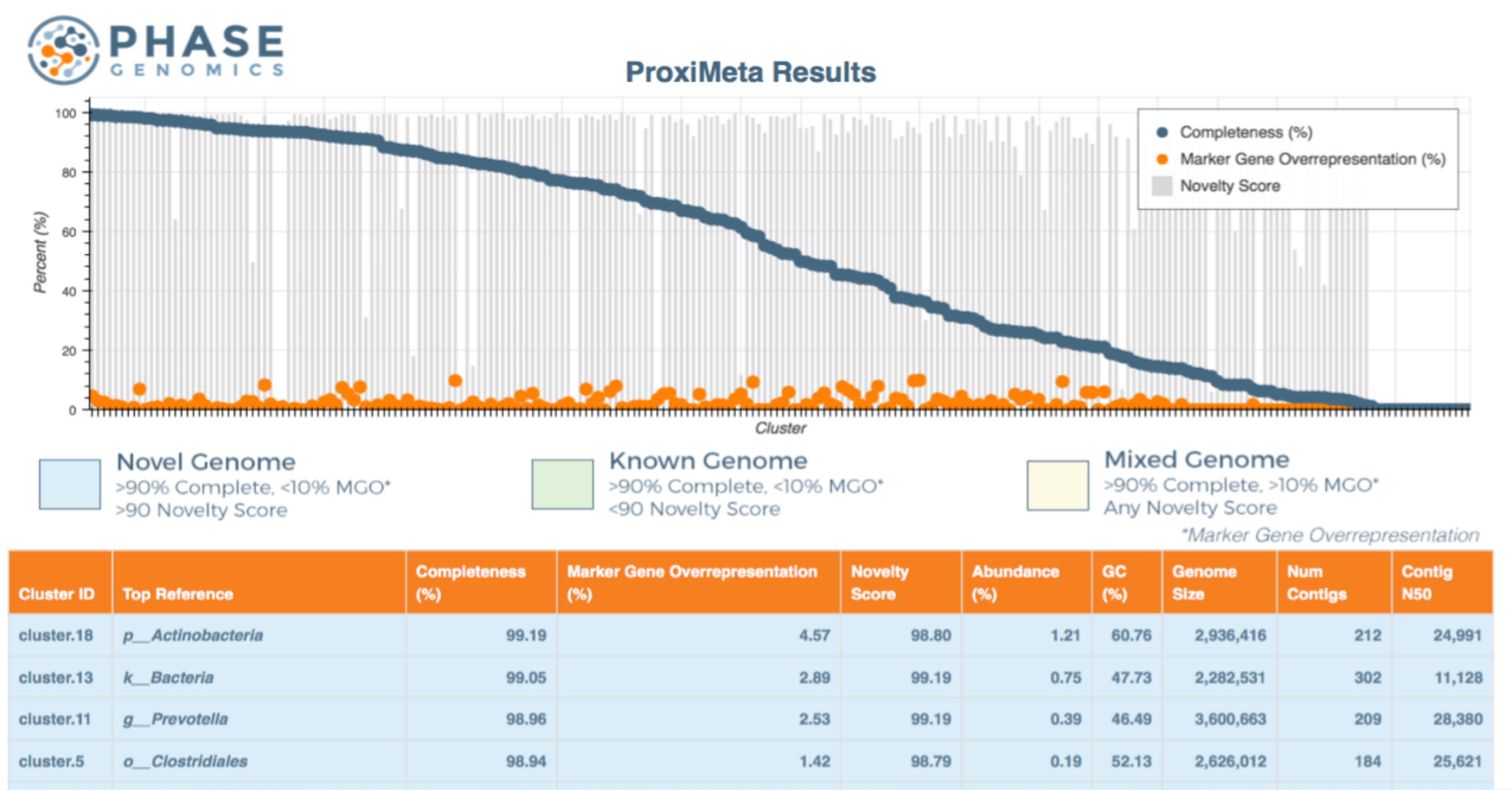
doi: https://doi.org/10.1101/760637

This article is a preprint and has not been certified by peer review [what does this mean?].



In collaboration with Herminia Loza-Tavera, Ayixon Sanchez-Reyez

High numbers of high-quality, novel genomes directly from rumen samples



Novel Genome
>90% Complete, <10% MGO*
>90 Novelty Score

Known Genome
>90% Complete, <10% MGO*
<90 Novelty Score

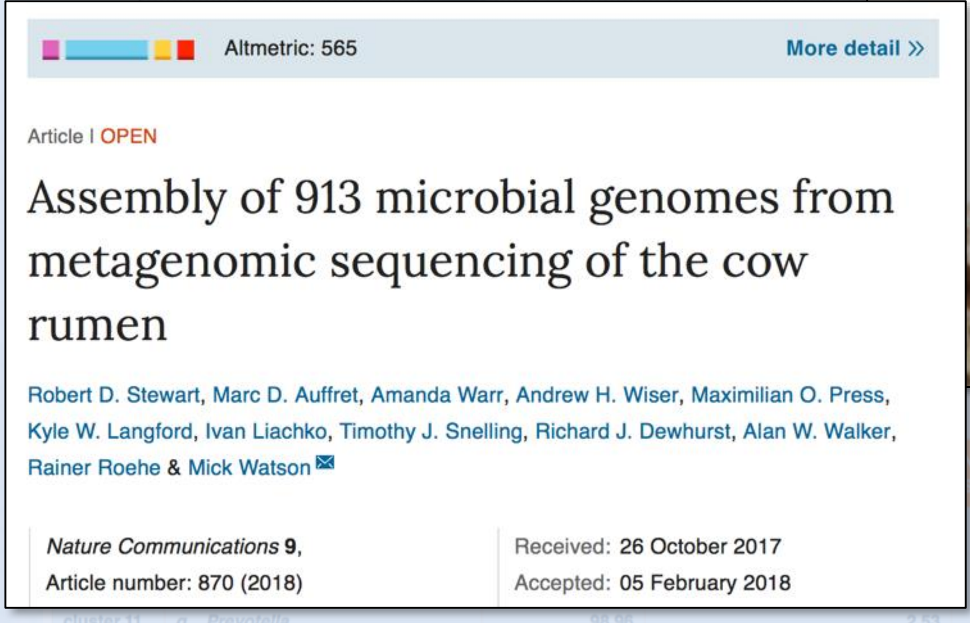
Mixed Genome
>90% Complete, >10% MGO*
Any Novelty Score

*Marker Gene Overrepresentation

Cluster ID	Top Reference	Completeness (%)	Marker Gene Overrepresentation (%)	Novelty Score	Abundance (%)	GC (%)	Genome Size	Num Contigs	Contig N50
cluster.18	*p__Actinobacteria*	99.19	4.57	98.80	1.21	60.76	2,936,416	212	24,991
cluster.13	*k__Bacteria*	99.05	2.89	99.19	0.75	47.73	2,282,531	302	11,128
cluster.11	*g__Prevotella*	98.96	2.53	99.19	0.39	46.49	3,600,663	209	28,380
cluster.5	*o__Clostridiales*	98.94	1.42	98.79	0.19	52.13	2,626,012	184	25,621

Stewart *et al.*, Nature Communications, 2018

High numbers of high-quality, novel genomes directly from rumen samples



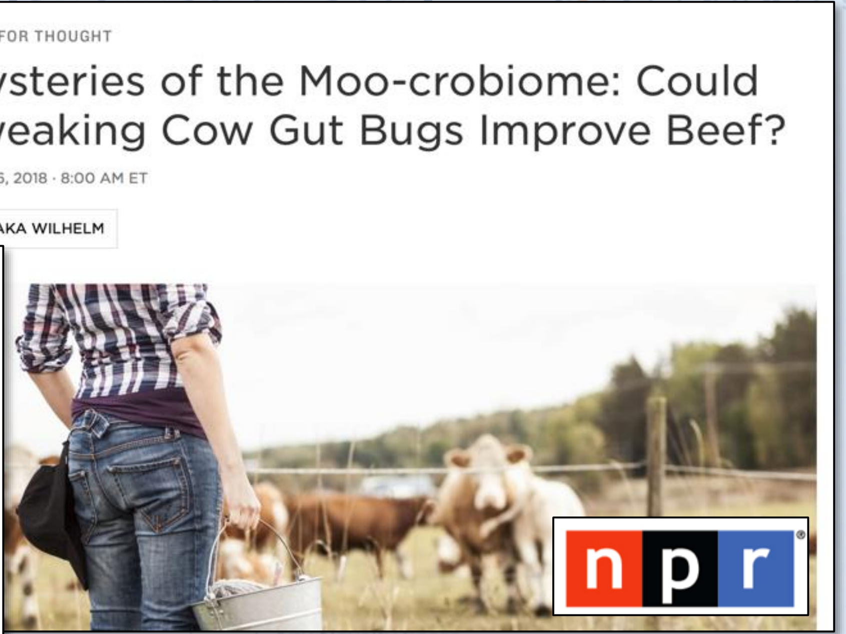
Article | OPEN

Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen

Robert D. Stewart, Marc D. Auffret, Amanda Warr, Andrew H. Wiser, Maximilian O. Press, Kyle W. Langford, Ivan Liachko, Timothy J. Snelling, Richard J. Dewhurst, Alan W. Walker, Rainer Roehe & Mick Watson

Nature Communications **9**, Article number: 870 (2018)

Received: 26 October 2017
Accepted: 05 February 2018



FOOD FOR THOUGHT

Mysteries of the Moo-crobiome: Could Tweaking Cow Gut Bugs Improve Beef?

March 6, 2018 · 8:00 AM ET

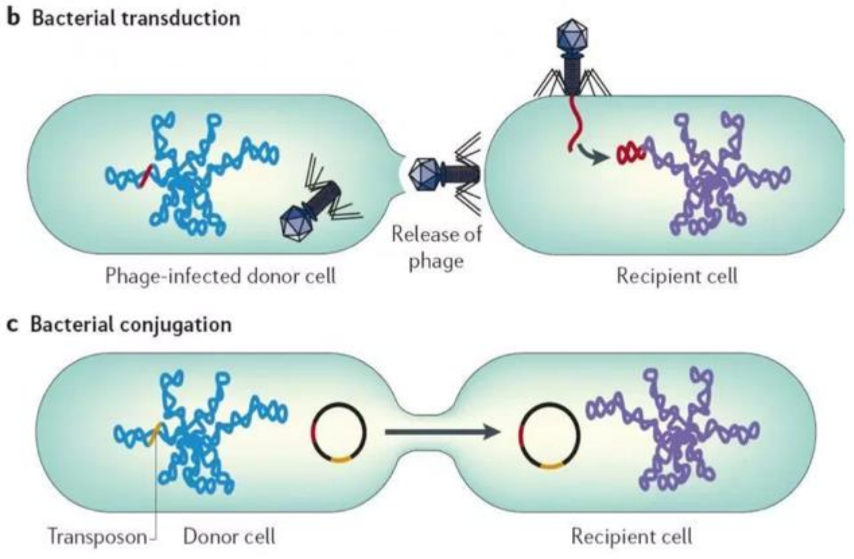
MENAKA WILHELM

Stewart *et al.*, Nature Communications, 2018

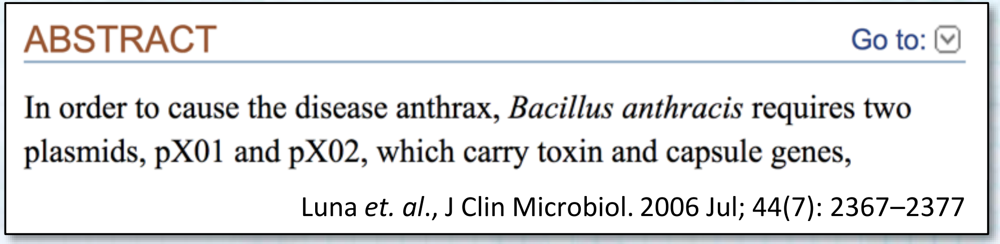
PHASE GENOMICS

Plasmids/viruses are key players in the microbiome

- Plasmids/Phage transmit AMR (Anti-microbial resistance)
- Plasmids often transmit pathogenic/toxic genes (ex. Anthrax)
- Virtually impossible to connect AMR and mobile elements with host strains using normal NGS
- Need a method that can directly link plasmids/viruses to hosts.



Furuya and Lowy, *Nat Rev Micro* 2006

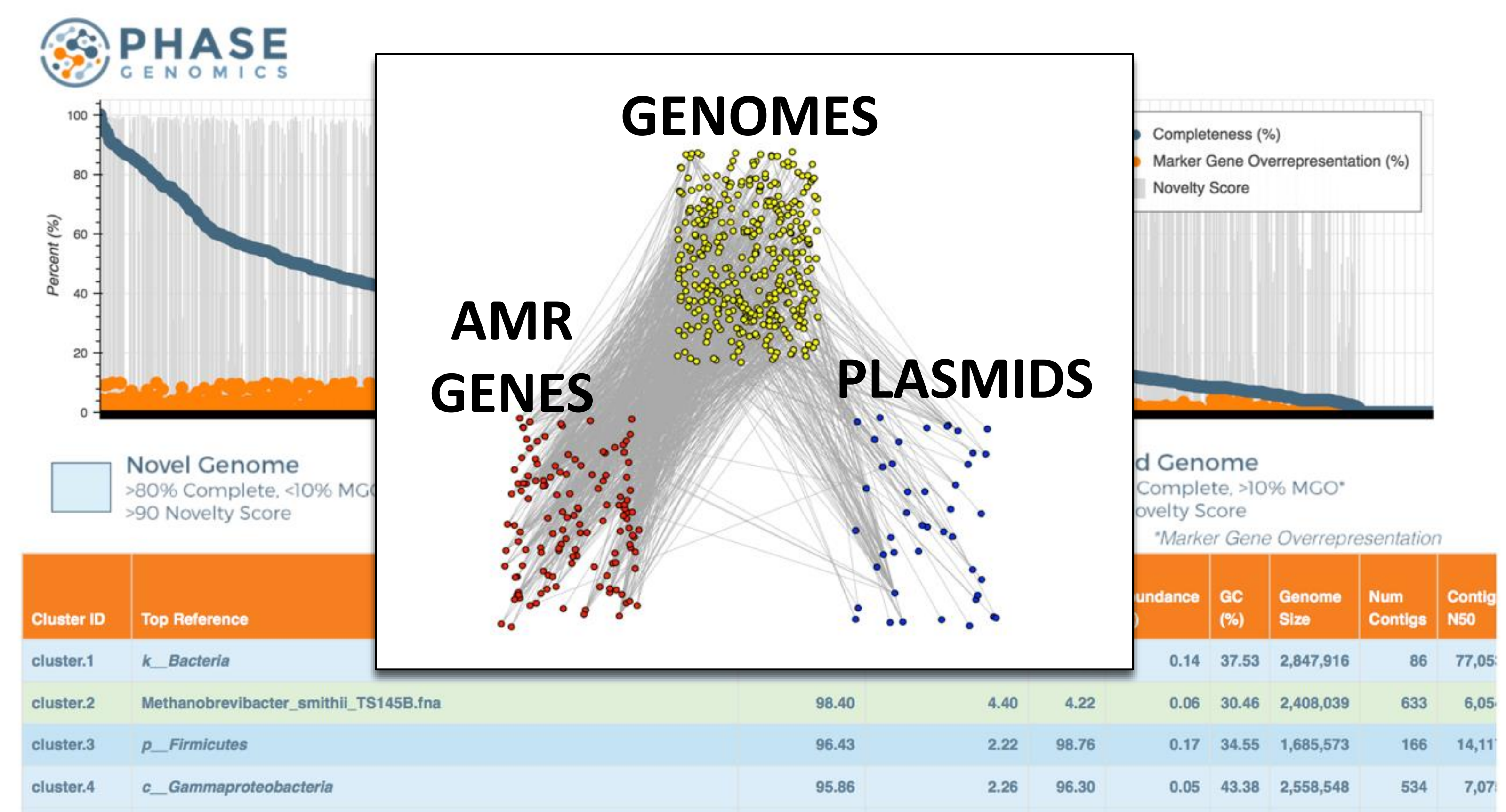


ABSTRACT

In order to cause the disease anthrax, *Bacillus anthracis* requires two plasmids, pX01 and pX02, which carry toxin and capsule genes,

Luna *et. al.*, J Clin Microbiol. 2006 Jul; 44(7): 2367–2377

Highly complex wastewater community



Novel Genome
>80% Complete, <10% MG
>90 Novelty Score

d Genome
Complete, >10% MGO*
ovelty Score
*Marker Gene Overrepresentation

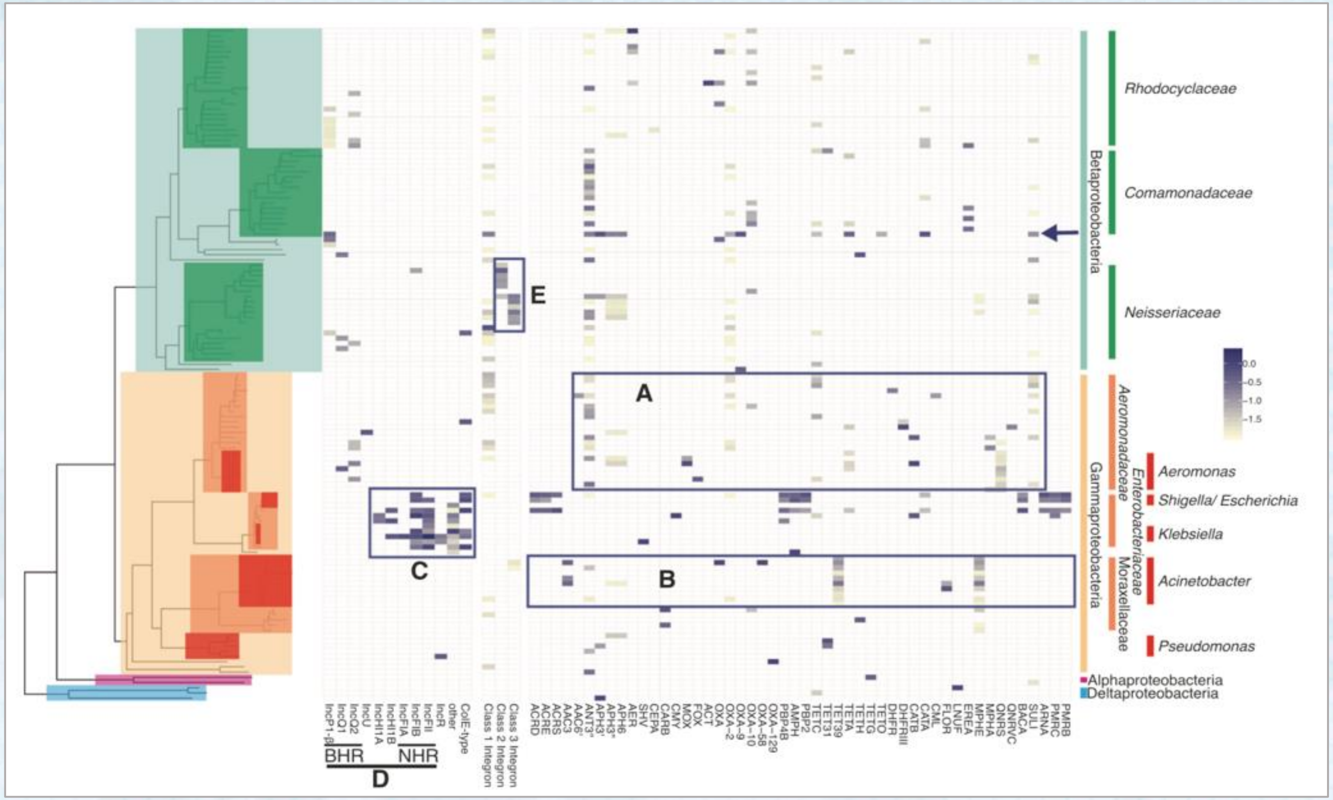
Cluster ID	Top Reference				undance	GC (%)	Genome Size	Num Contigs	Contig N50
cluster.1	k__Bacteria				0.14	37.53	2,847,916	86	77,05
cluster.2	Methanobrevibacter_smithii_TS145B.fna	98.40	4.40	4.22	0.06	30.46	2,408,039	633	6,05
cluster.3	p__Firmicutes	96.43	2.22	98.76	0.17	34.55	1,685,573	166	14,11
cluster.4	c__Gammaproteobacteria	95.86	2.26	96.30	0.05	43.38	2,558,548	534	7,07

Stalder *et al.*, 2019, *ISME J*

Linking the ‘Mobilome’ to the Microbiome

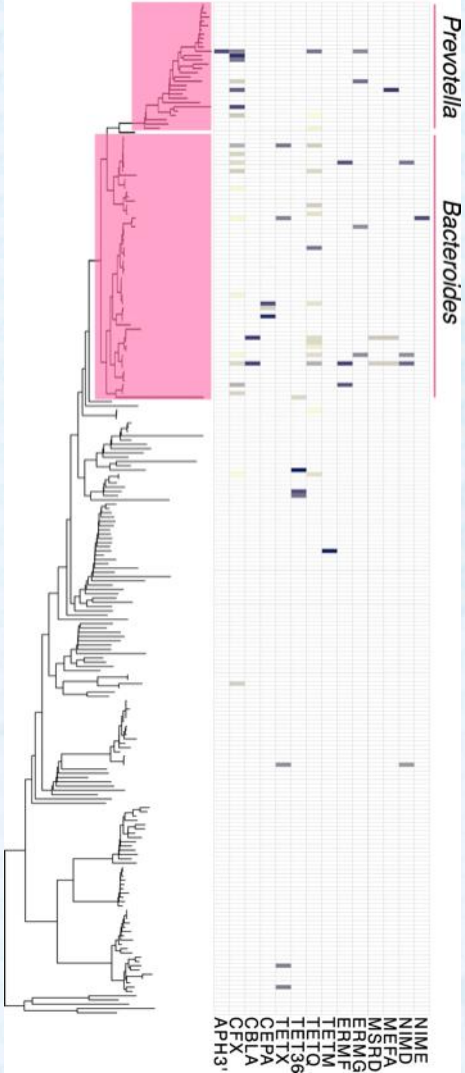
Genomes Plasmids Antibiotic resistance genes

Proteobacteria

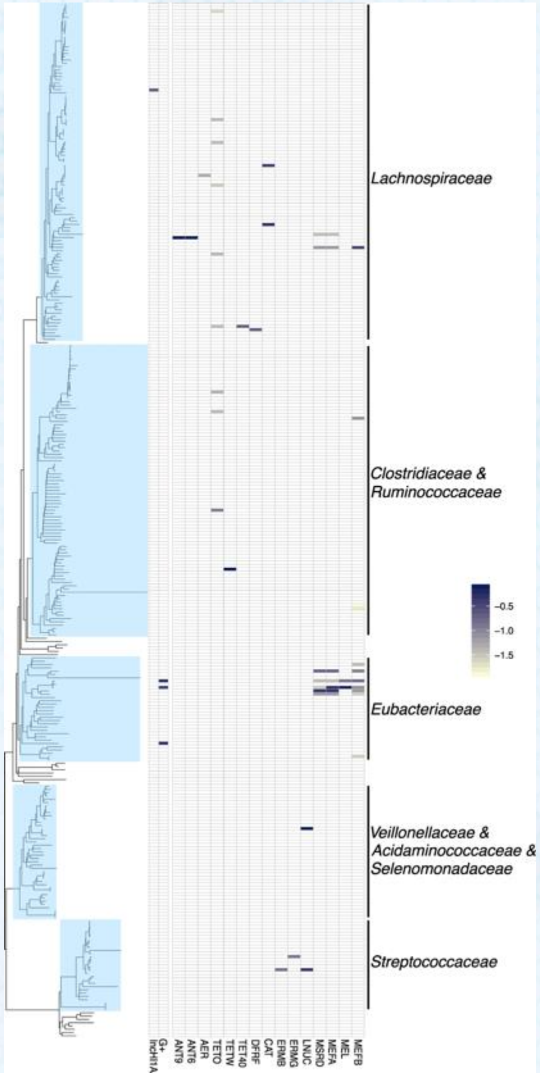


Stalder *et al.*, 2019, *ISME J*

Bacterioidetes



Firmicutes



~1200 PAGs

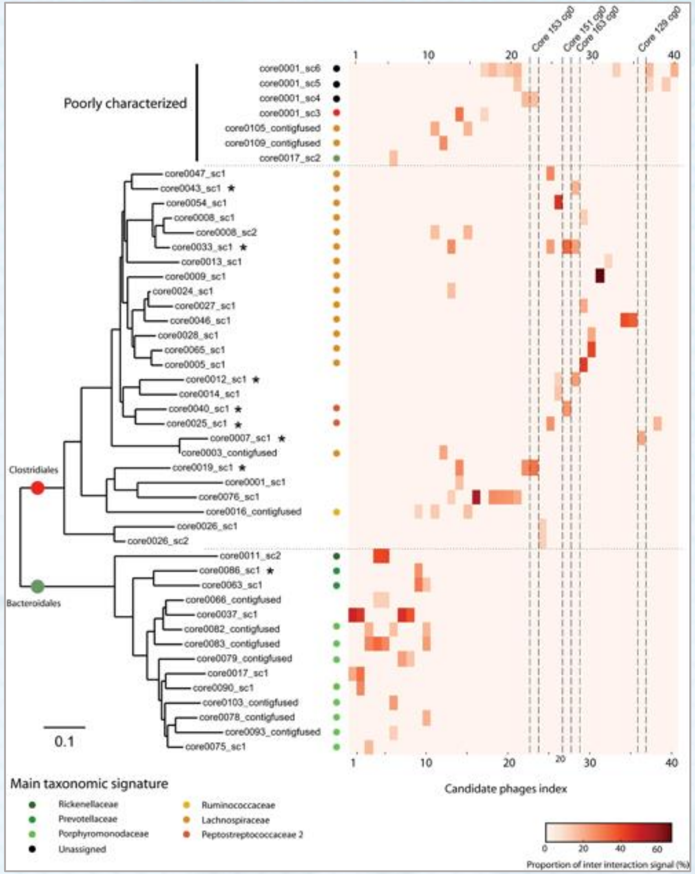
396 host-ARG assn

83 plasmid-host assn

58 integron-host assn

Stalder *et al*., 2019, *ISME J*

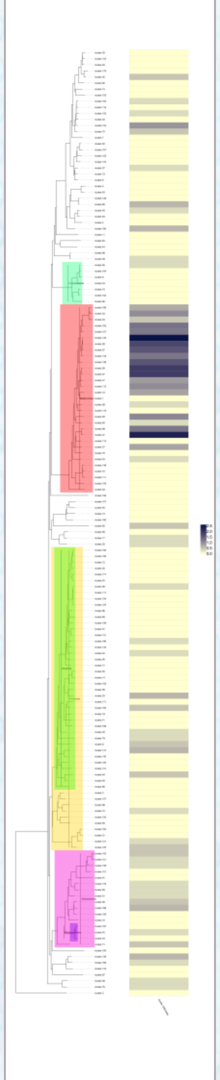
Tracking viral-host association in metagenomes



Marbouty *et al*., Science Advances 2017

Stalder et al., ISME J, 2019

ex: Where is crAssphage?



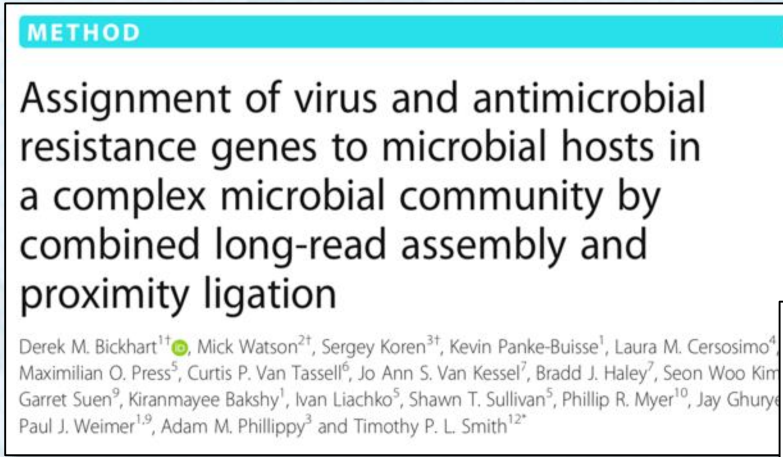
Bacteriodetes

Clostridia

Actinobacteria

Combining long reads and ProxiMeta in a complex microbiome context

PHASE GENOMICS

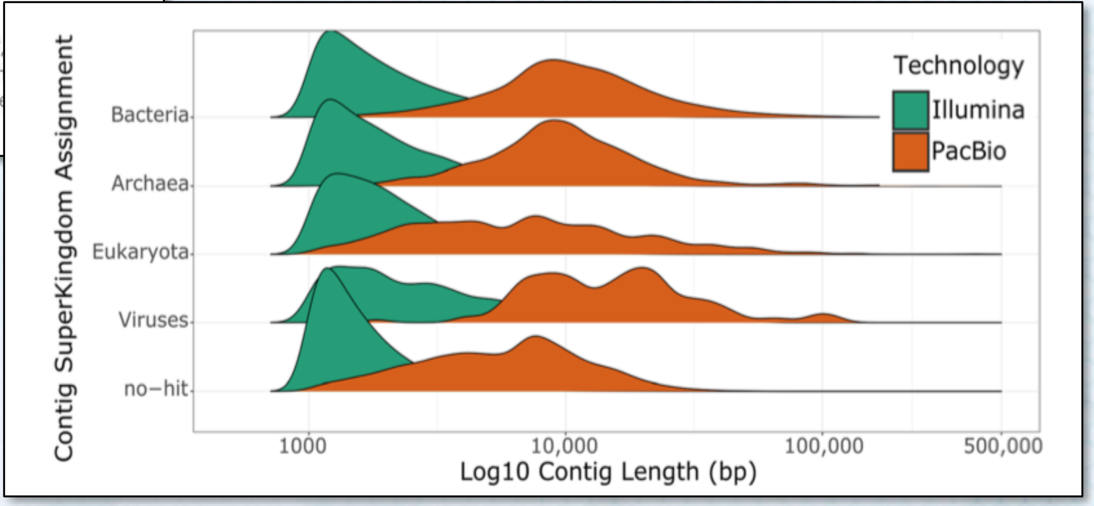


METHOD Open Access

Check for updates

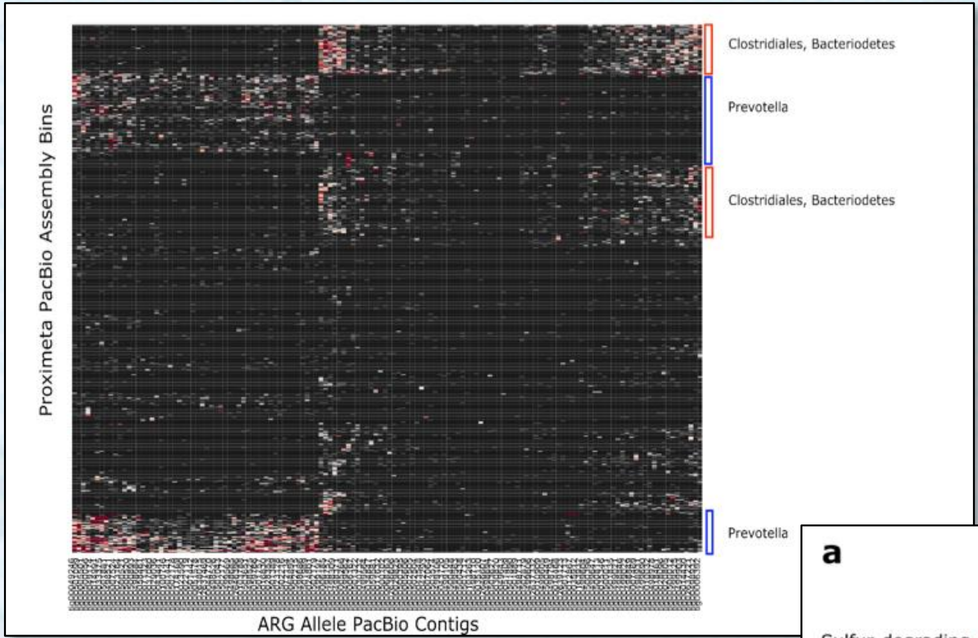
Assignment of virus and antimicrobial resistance genes to microbial hosts in a complex microbial community by combined long-read assembly and proximity ligation

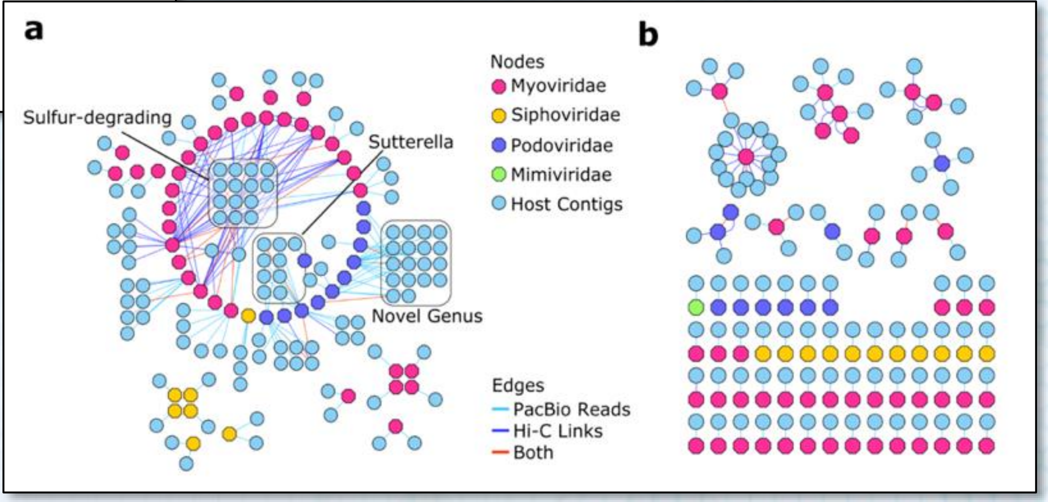
Derek M. Bickhart[1†], Mick Watson[2†], Sergey Koren[3†], Kevin Panke-Buisse[1], Laura M. Cersosimo[4], Maximilian O. Press[5], Curtis P. Van Tassell[6], Jo Ann S. Van Kessel[7], Bradd J. Haley[7], Seon Woo Kim Garret Suen[9], Kiranmayee Bakshy[1], Ivan Liachko[5], Shawn T. Sullivan[5], Phillip R. Myer[10], Jay Ghurye Paul J. Weimer[1,9], Adam M. Phillippy[3] and Timothy P. L. Smith[12*]



Bickhart *et al.*, *Genome Biology*, 2019

Connecting ARGs and viruses to their hosts





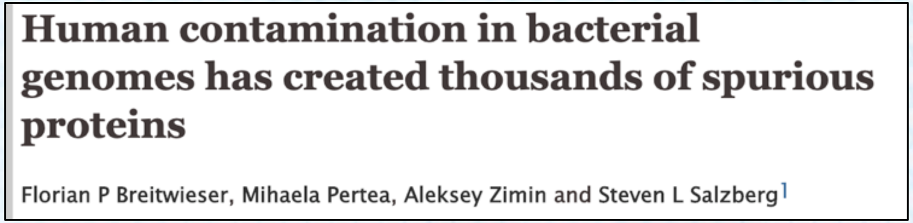
*188 Novel viruses and host interactions discovered from one rumen sample

Bickhart *et al.*, *Genome Biology*, 2019

Sneak Peek…

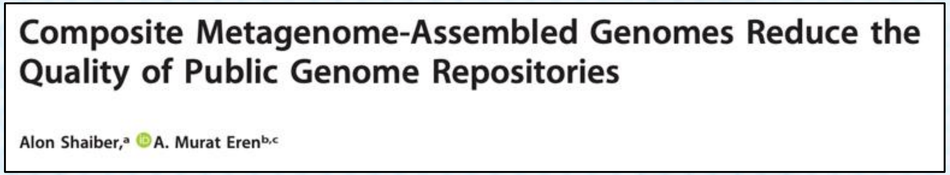
The limitations of metagenomic binning

PHASE GENOMICS



Human contamination in bacterial genomes has created thousands of spurious proteins

Florian P Breitwieser, Mihaela Pertea, Aleksey Zimin and Steven L Salzberg[1]

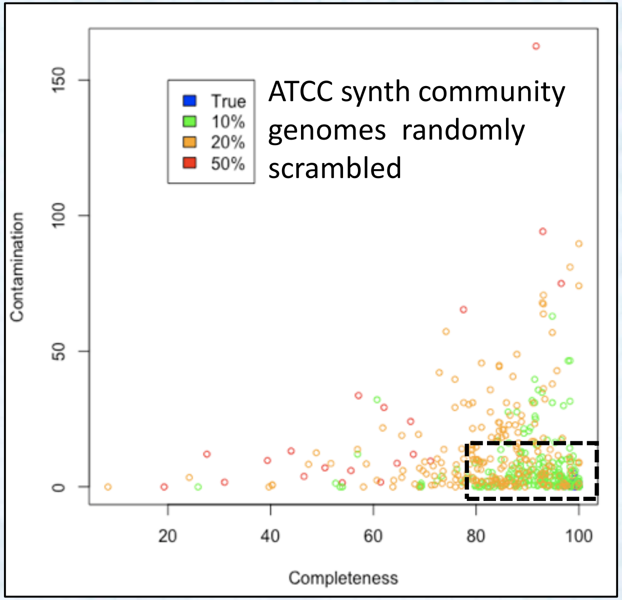


Composite Metagenome-Assembled Genomes Reduce the Quality of Public Genome Repositories

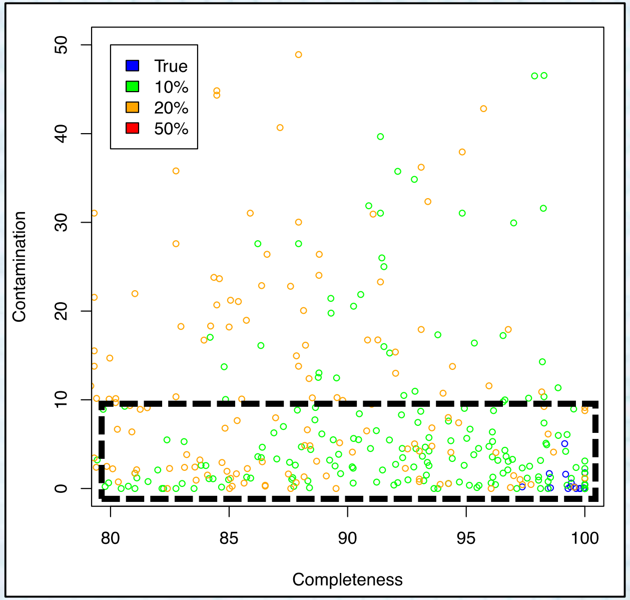
Alon Shaiber,[a] A. Murat Eren[b,c]

CheckM uses core gene content to QC MAGs

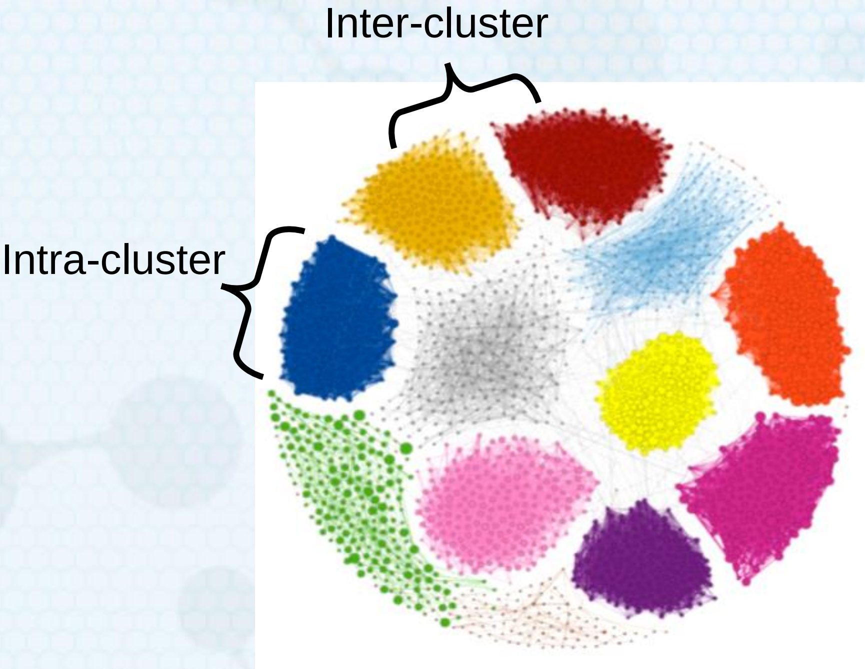
Scrambled genomes are often called as High-Quality MAGs



ATCC synth community genomes randomly scrambled



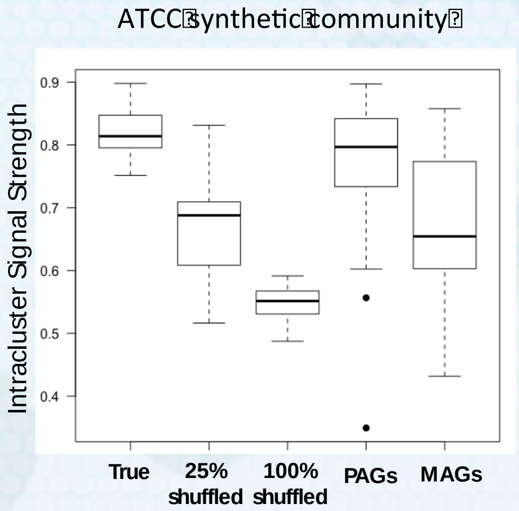
ProxiMeta data provides a direct orthogonal datatype to QC metagenome bins

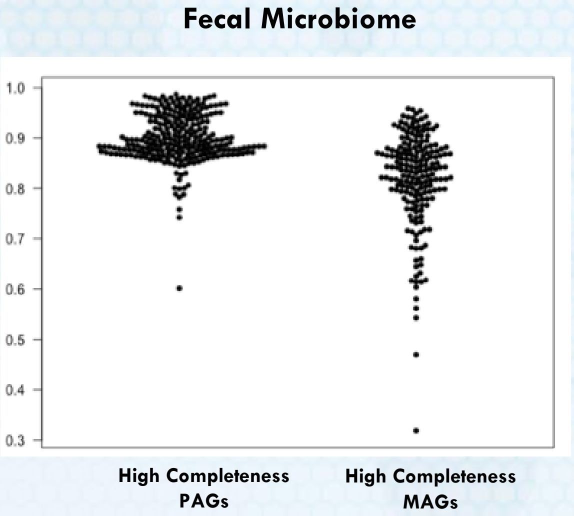


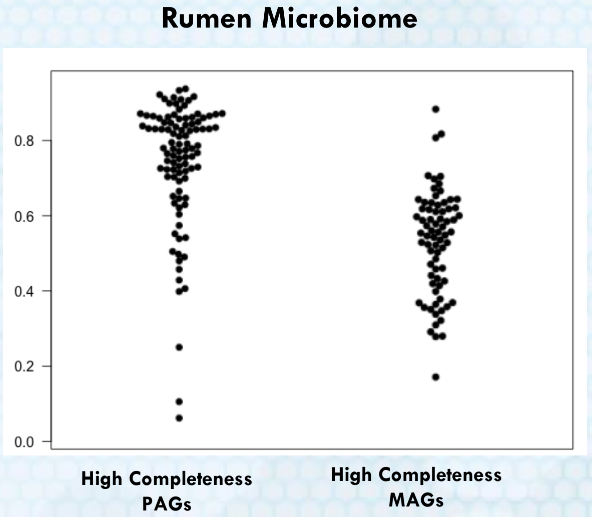
Correctly binned: $\frac{\text{intra}}{\text{inter}}$ = ↑

Scrambled: $\frac{\text{intra}}{\text{inter}}$ = ↓

Proximity-assembled genomes are more accurate than MAGs

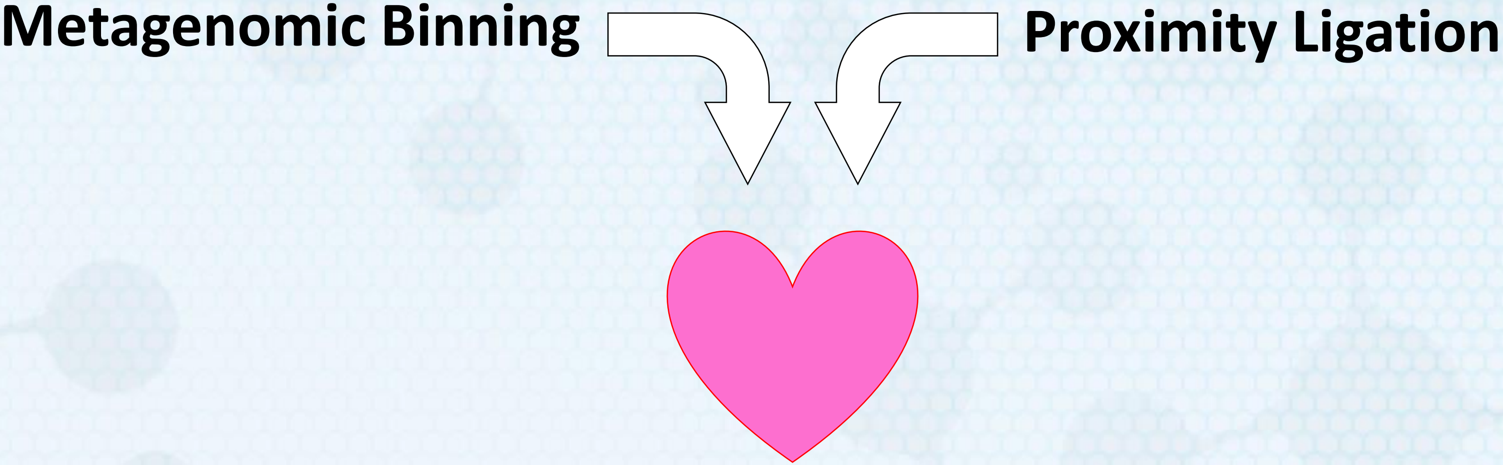

ATCC synthetic community


Fecal Microbiome


Rumen Microbiome

*High-completeness MAGs demonstrate low intra-cluster enrichment values, indicating high degree of error

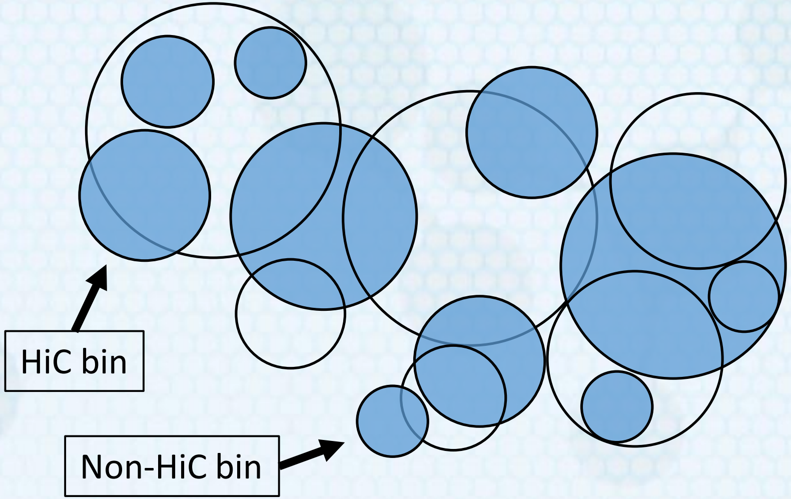
Can we get the best of both worlds?



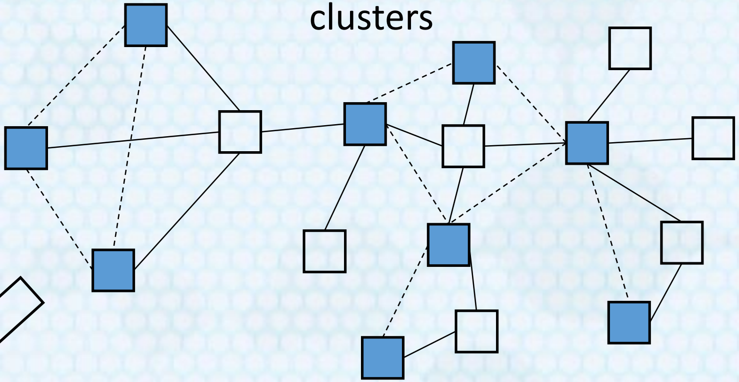
Metagenomic Binning

Proximity Ligation

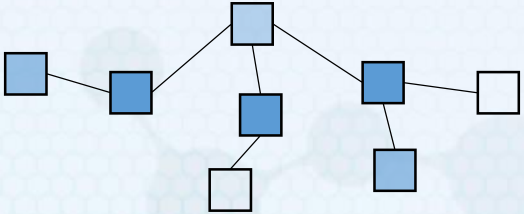
1. Compute overlaps between HiC and non-HiC bins



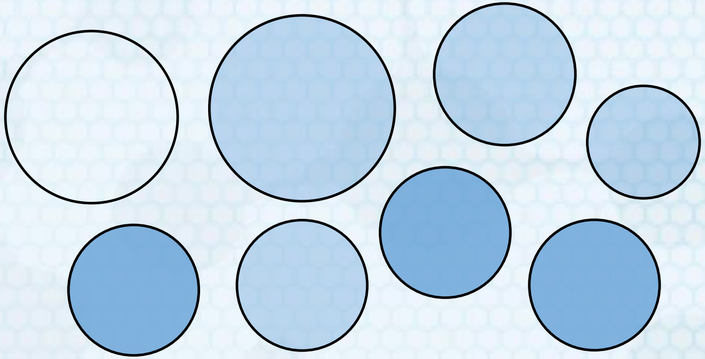
2. Generate connectivity network of contig clusters



3. Supervised algorithm to hierarchically de-convolve network

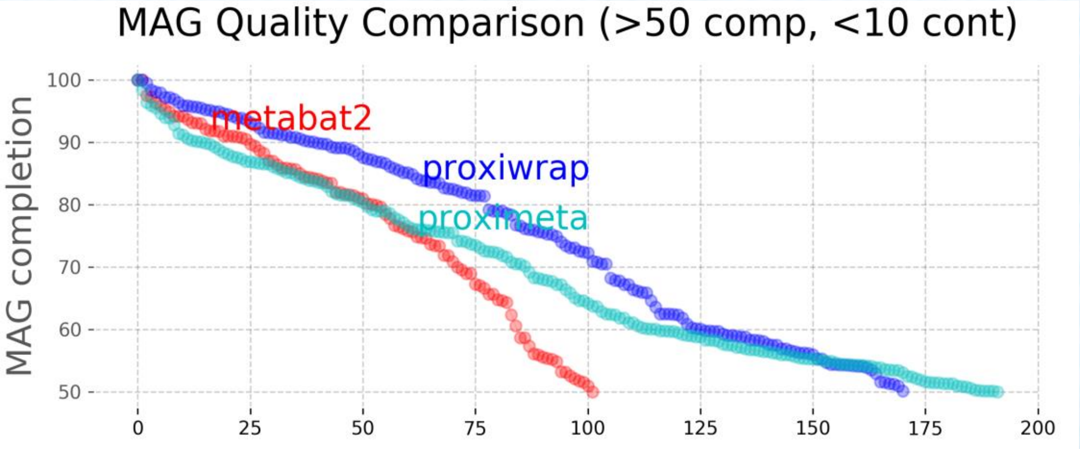


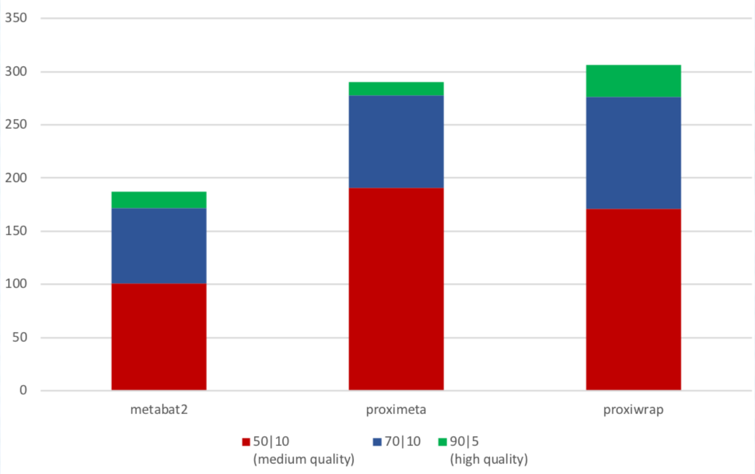
4. Export final set of bins



Gherman Uritskiy

Applying ProxiWrap™ to highly complex wastewater sample

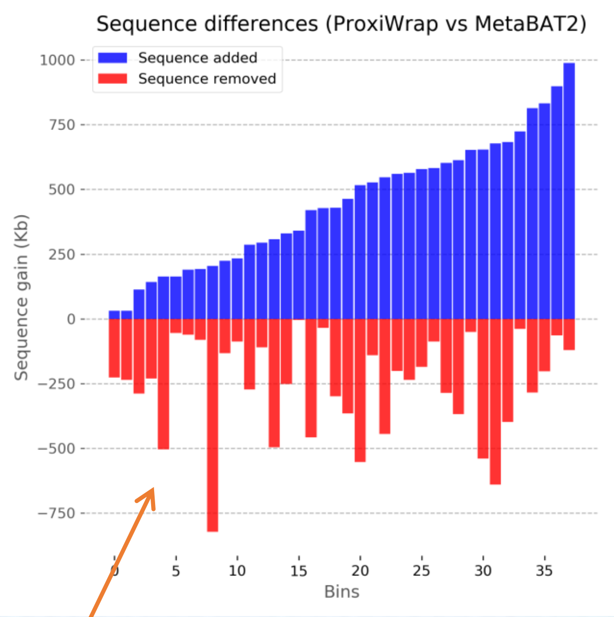


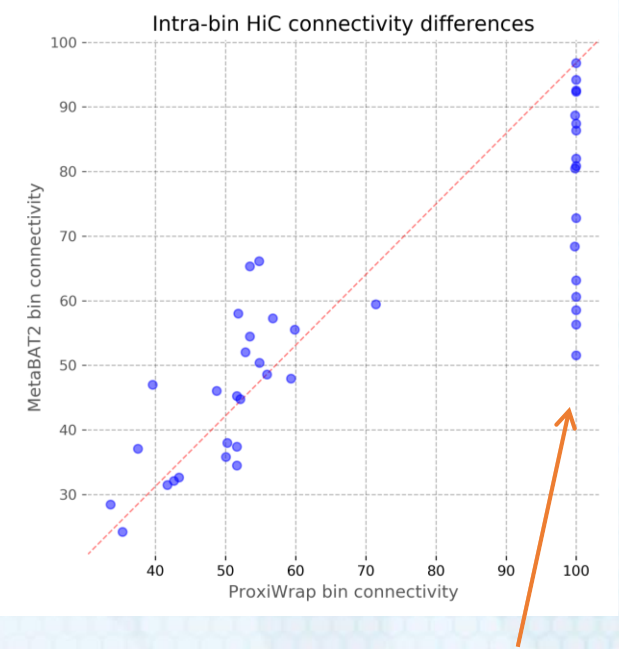


*ProxiWrap yields more high quality genomes than metaBAT2 (binning) or ProxiMeta (Hi-C)

Gherman Uritskiy

Applying ProxiWrap to highly complex wastewater sample

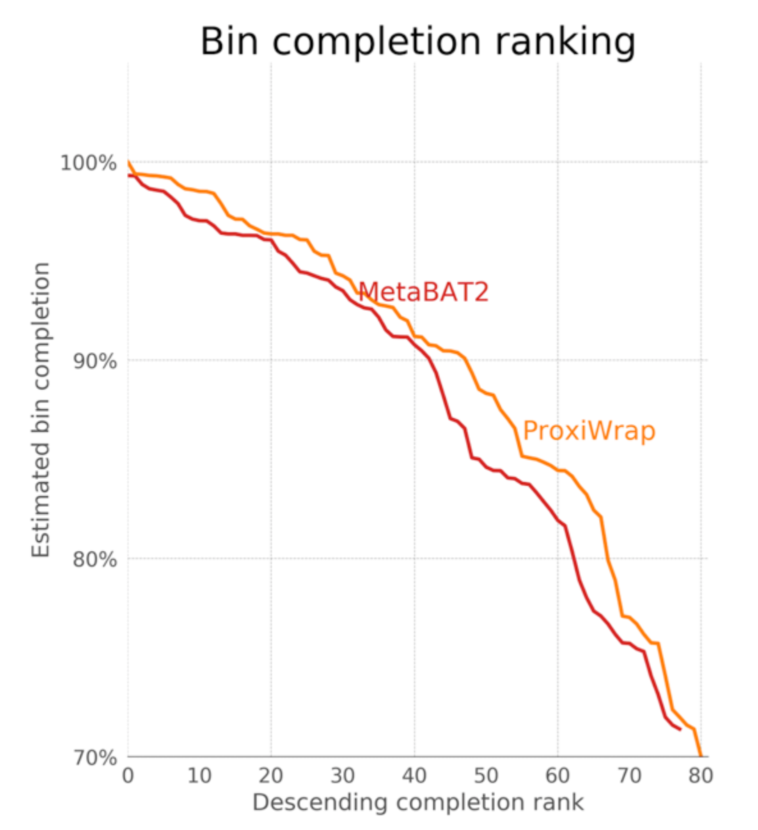




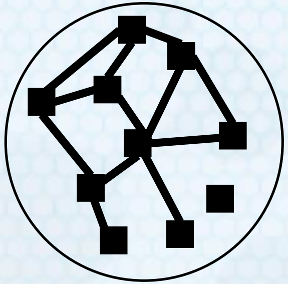
Sequences added and removed by ProxiWrap

Improving intra-cluster connectivity

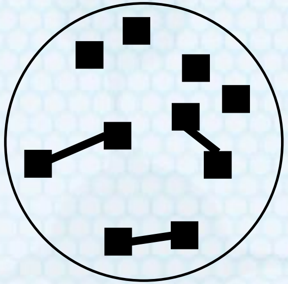
ProxiWrap™ significantly improves MAG completion and provides estimates of MAG validity



MAG	Completion	Contamination	Connectivity
bin.73	100%	4.2%	88.3%
bin.7	95.8%	2.7%	68.8%
bin.35	98.6%	0.3%	3.2%
bin.111	71.8%	5.7%	0.01%



High-confidence MAG (high inter-contig Hi-C connectivity)



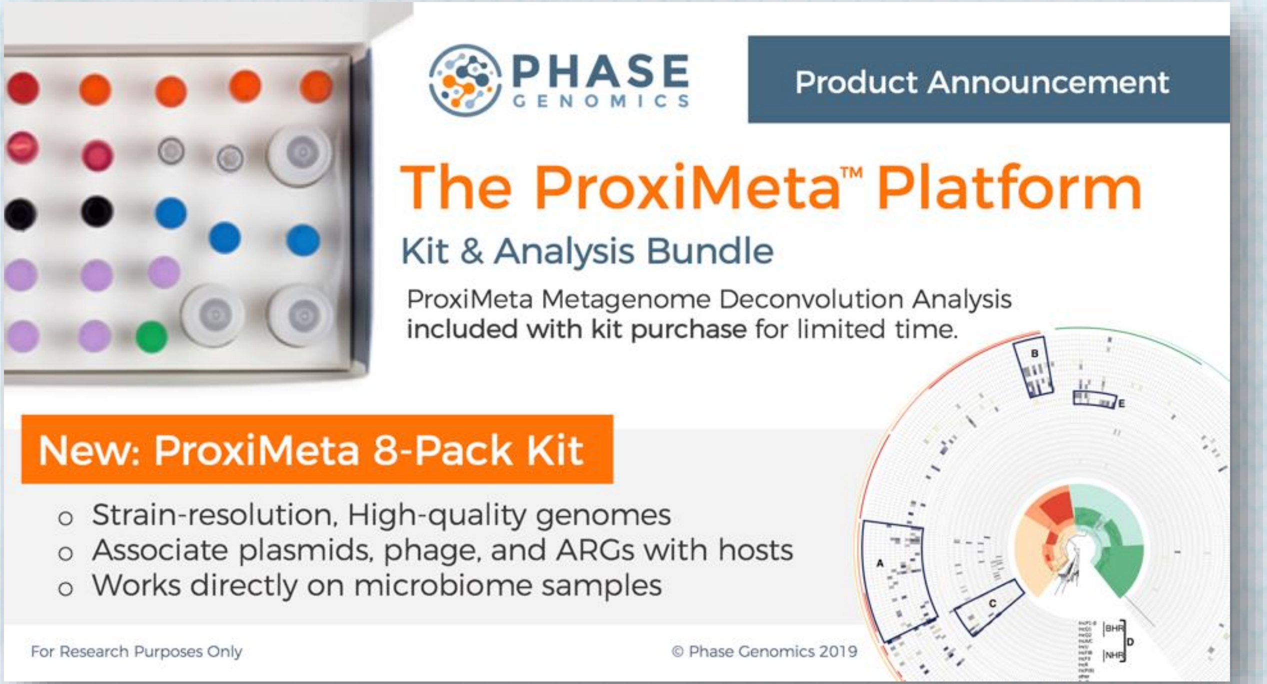
Low-confidence MAG (low inter-contig Hi-C connectivity)

Gherman Uritskiy

Proximity-Guided Metagenome Assembly™

Genomes, Strains, Mobile Elements

- No culturing
- No binning/de-replication
- No *a priori* information
- No HMW-DNA
- No special machinery



PHASE GENOMICS

Product Announcement

The ProxiMeta™ Platform

Kit & Analysis Bundle

ProxiMeta Metagenome Deconvolution Analysis **included with kit purchase** for limited time.

New: ProxiMeta 8-Pack Kit

- Strain-resolution, High-quality genomes
- Associate plasmids, phage, and ARGs with hosts
- Works directly on microbiome samples

For Research Purposes Only

© Phase Genomics 2019

New 8-pack kits, **ANALYSIS INCLUDED**

Acknowledgements

PHASE GENOMICS

UW Genome Sciences

Jay Shendure
Maitreya Dunham
Ehanan Borenstein
Josh Burton
Bill Noble
Nelle Varoquaux
Ferhat Ay

Collaborators

Eva Top
Mick Watson
Patrick Chain
Herminia Loza-Tavera
Tim Smith
Adam Phillippy
Justin Sonnenburg
Team PacBio

Phase Genomics

Shawn Sullivan
Andrew Wiser
Max Press
Zev Kronenberg
Kyle Langford
Kaylee Mueller
Kayla Young
Gherman Uritskiy

BILL & MELINDA GATES foundation

NIH National Institute of Allergy and Infectious Diseases

CONGRUENT VENTURES

WRF CAPITAL

Genome Sciences UNIVERSITY OF WASHINGTON | SCHOOL OF MEDICINE

NATIONAL INSTITUTES OF HEALTH

NSF

RITA ALLEN FOUNDATION