Single cell genomics:
A deeper look at life's building blocks

Ramunas Stepanauskas


Art by Glynn Gorick

Bigelow | Laboratory for Ocean Sciences

Acknowledgements

SCGC staff:
Joseph Brown
Julia Brown
Wendy Bellows
Elizabeth Fergusson
Christian Harris
Monica Lluesma
Corianna Mascena
Dashiell Masland
Ben Tupper
Brian Thompson

FACS facility:
Nicole Poulton

SCGC Advisory Board
Rita Colwell
George Garrity
Kostas Konstantinidis
Craig Muir
Ger van den Engh
Tanja Woyke
Robert Young

RS postdocs (current and past):
Eric Becraft
Jane Heywood
Jessica Labonté
Manuel Garcia Martinez
Maria Pachiadaki
Brandon Swan

Key collaborators
Sallie Chisholm
Gerhad Herndl
Arvydas Lubys
Duane Moser
Stefan Sievert
Tullis Onstott
Kathuri Venkateswaran
Tanja Woyke

NSF NASA Office of Science U.S. DEPARTMENT OF ENERGY mti Maine Technology Institute *Investing in Innovation*

SIMONS FOUNDATION illumina® ThermoFisher SCIENTIFIC

Bigelow Laboratory for Ocean Sciences

Open postdoc positions

Scope: Single cell genomics and horizontal gene transfer in the marine microbiome

Application deadline: January 15, 2017

Further information:
https://www.bigelow.org/about/careers.html
rstepanauskas@bigelow.org

Bigelow | Laboratory for Ocean Sciences

Single cell genomics:
A deeper look at life's building blocks

Ramunas Stepanauskas

SCG opportunities in microbiology

Coffee break

Do we know how much we don't know?

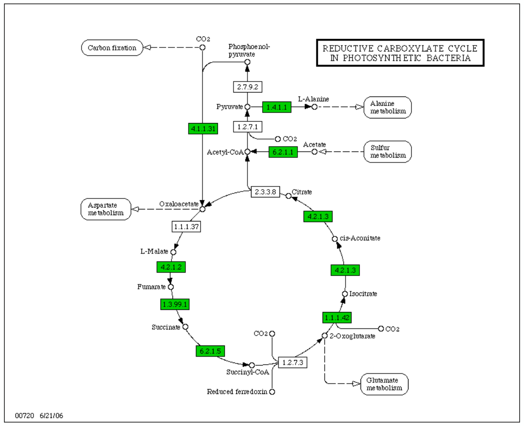
Leg stretcher

Infrastructure and method advances

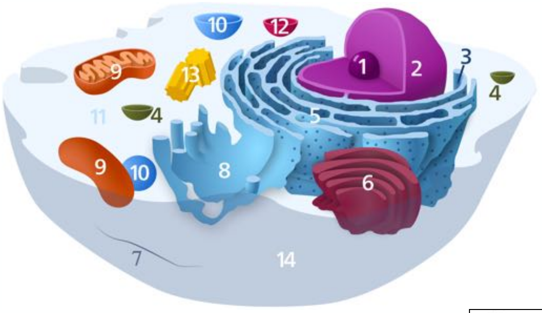
Art by Glynn Gorick

Bigelow | Laboratory for Ocean Sciences

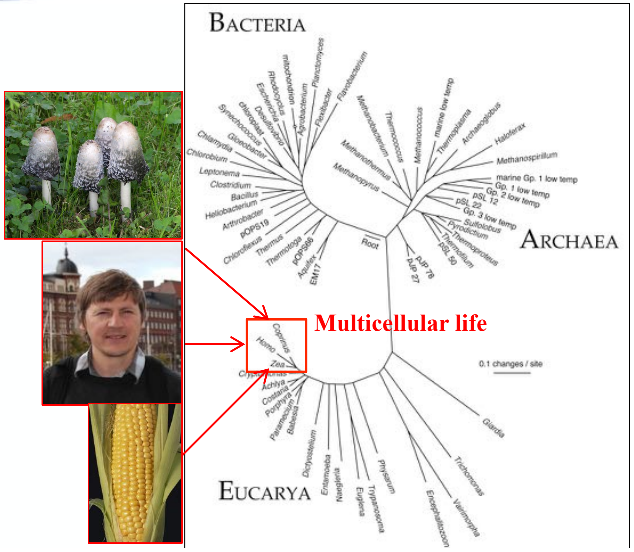
Why single cell genomics?



Genomes are blueprints of life



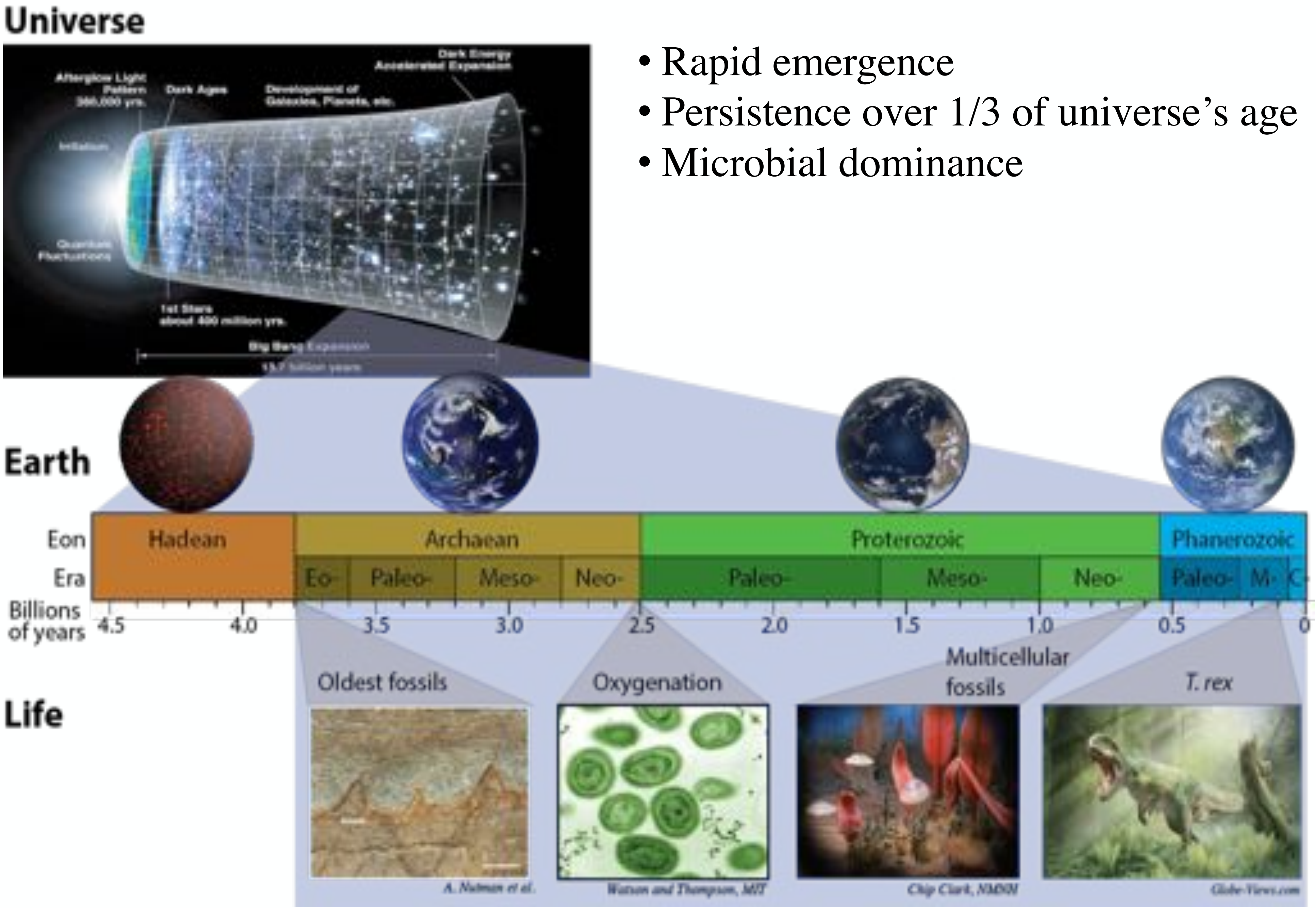
Cells are fundamental units of biology



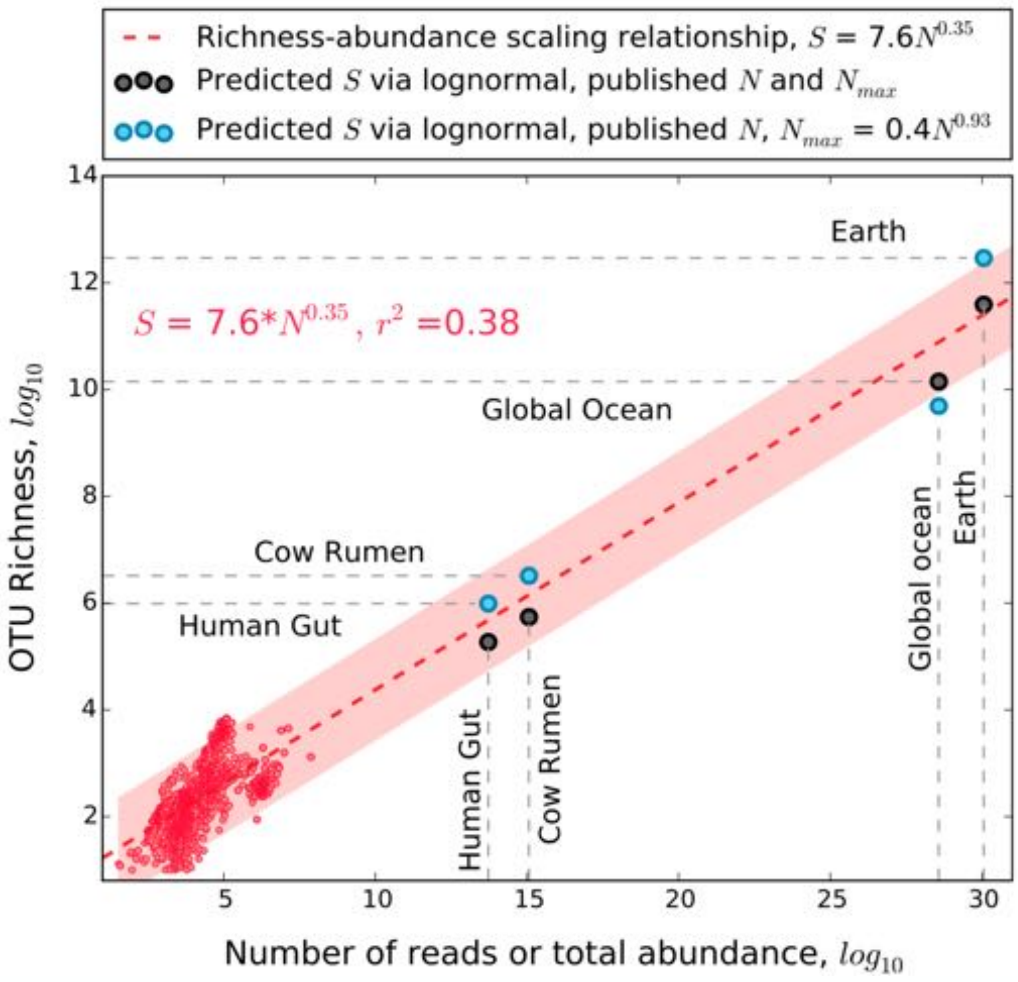
For most of life on Earth, 1 cell = 1 organism

Amazing features of life on Earth

- Rapid emergence
- Persistence over 1/3 of universe’s age
- Microbial dominance



Our planet may hold trillions of microbial species…



From: Locey and Lennon 2016

Implications:

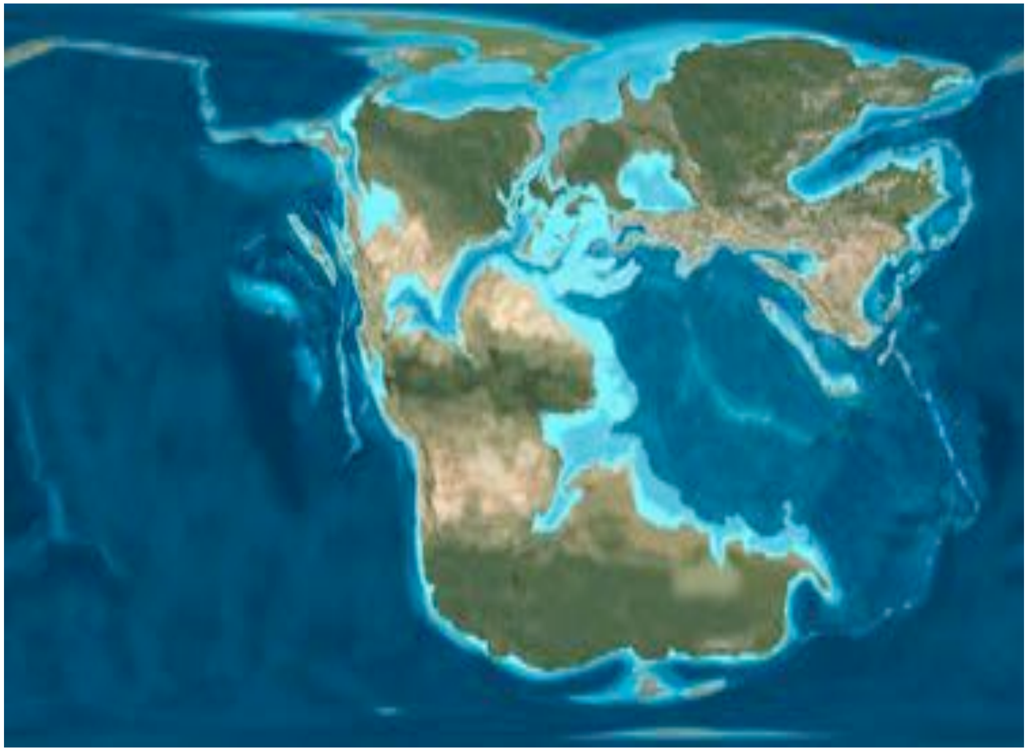
- Uncultured microbes run the planet
- Only a tiny fraction will ever be cultivated
- Molecular technologies have to be pushed further

OTUs are not species!

3% divergence in the 16S rRNA gene takes ~150 million years, so...

...last common ancestor of an average OTU lived in Jurassic!

Don't bacteria evolve faster than mammals?



Images from Wikipedia

Bigelow | Laboratory for Ocean Sciences

Oceans of genetic information

1 mL surface ocean water encodes ~1 TB genetic information

1 g of agricultural soil encodes ~1 PB

The entire planet encodes ~10^{21} PB

<1% of this genetic resource can be accessed by cultivation

Assumptions of metagenomic assemblies

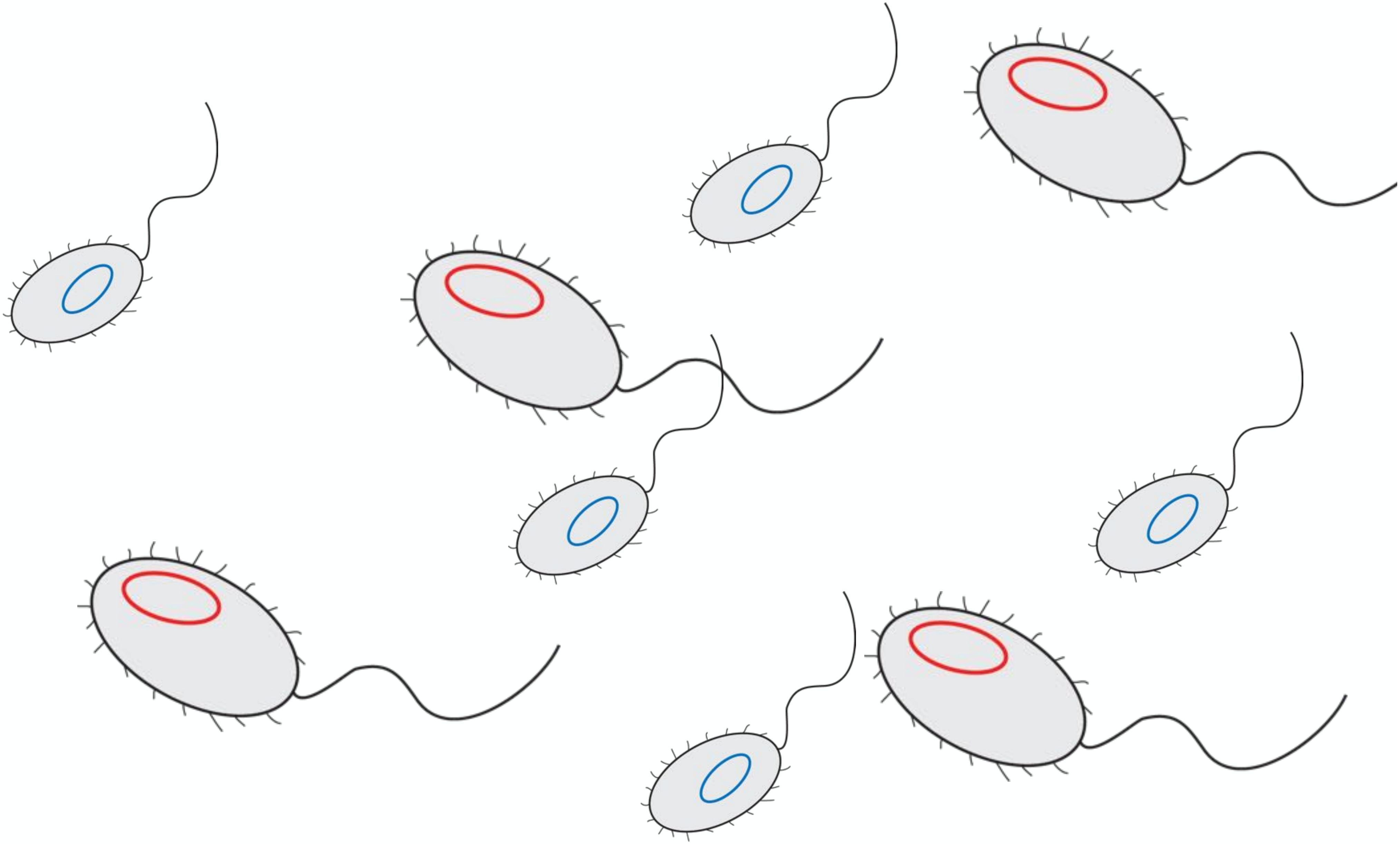
Image from Wikipedia

Microbial communities are composed of clonal populations

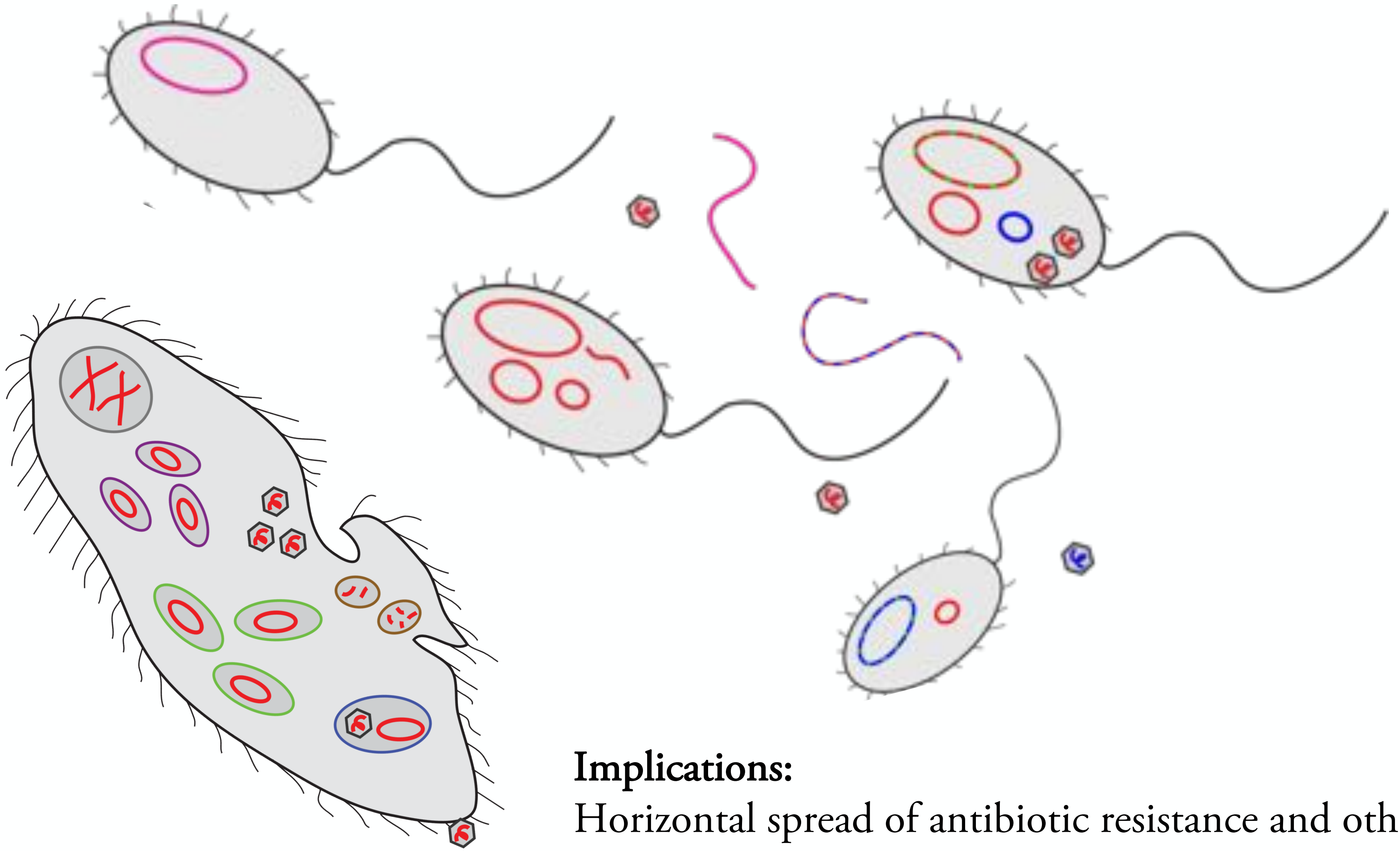
Entire genome evolves synchronously

De novo assembly software does not cross-assemble different populations

DNA in a idealized microbial assemblage



Complex DNA distribution in a microbial assemblage



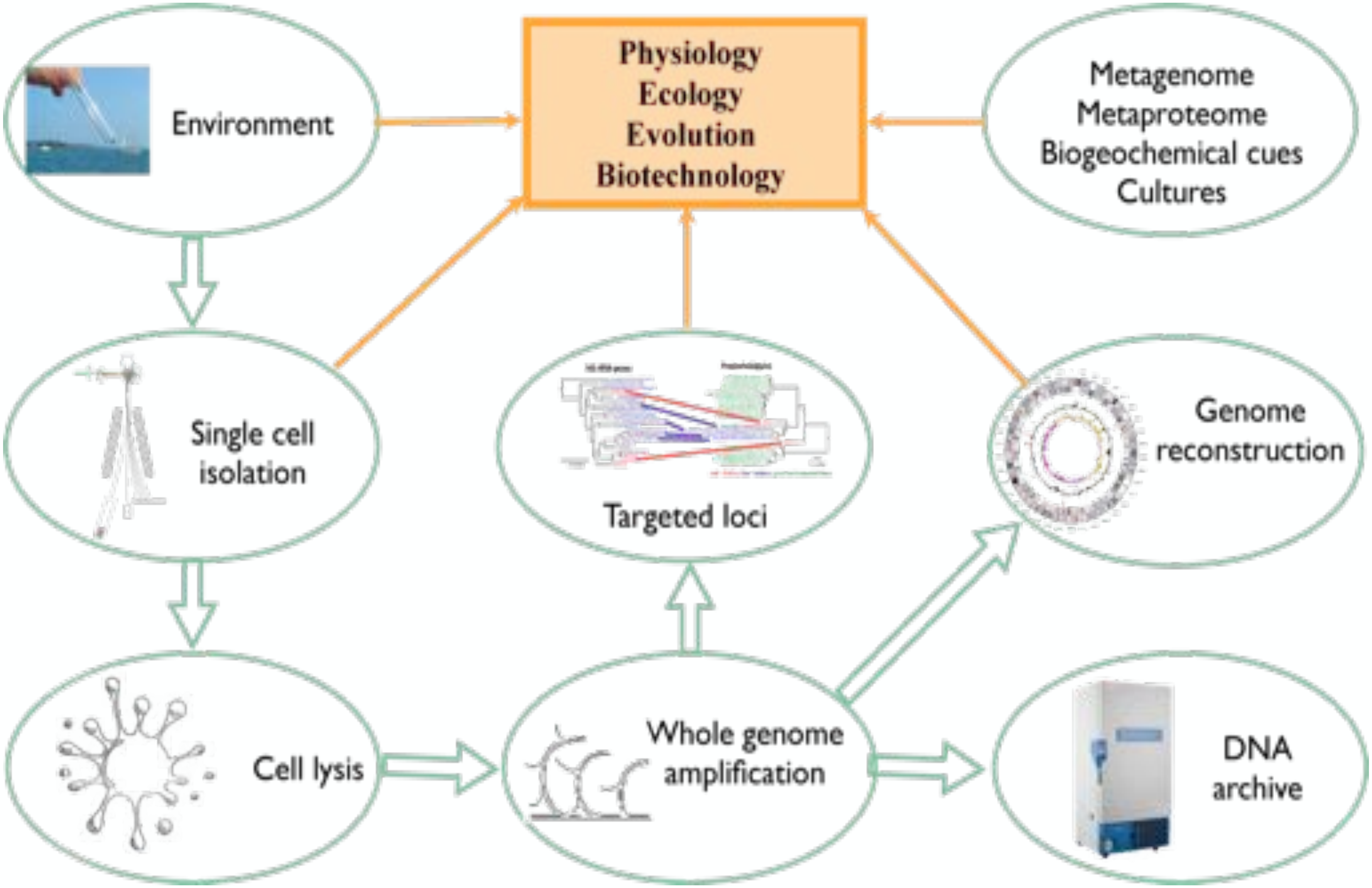
Implications:

Horizontal spread of antibiotic resistance and other traits

Large differences among close relatives

16S and metagenomics provide incomplete information

Microbial Single Cell Genomics Workflow



Environment

Physiology
Ecology
Evolution
Biotechnology

Metagenome
Metaproteome
Biogeochemical cues
Cultures

Single cell isolation

Targeted loci

Genome reconstruction

Cell lysis

Whole genome amplification

DNA archive

Bigelow | Laboratory for Ocean Sciences

Milestones of microbial single cell genomics

Pre-2000: Experimenting with PCR-based single cell DNA amplification

- *e.g. Zhang et al. (1992) PNAS 89:5847-5851*

2001-2002: Development of multiple displacement amplification (MDA)

- *Dean et al. (2001) Genome Research 11:1095-1099*
- *Dean et al. (2002) PNAS 99:5261-5266*

2005-2006: Proof-of-concept single cell MDA on cultured microorganisms

- *Raghunathan et al. (2005) AEM 71: 3342-3347*
- *Zhang et al. (2006) Nature Biotech. 24:680-686*

2007-2008: First genomic data from uncultured, single cells

- *Stepanauskas and Sieracki (2007) PNAS 104: 9052-9057*
- *Marcy et al. (2007) PNAS 104:11889-11894*

2009-now: High-throughput facilities; major research discoveries

- *Over 100 publications in microbial ecology, evolution, bioprospecting and human health*

Bigelow Laboratory Single Cell Genomics Center

scgc.*bigelow.org*

SCGC
SINGLE CELL GENOMICS CENTER
Reading nature's genetic tales, one cell at a time

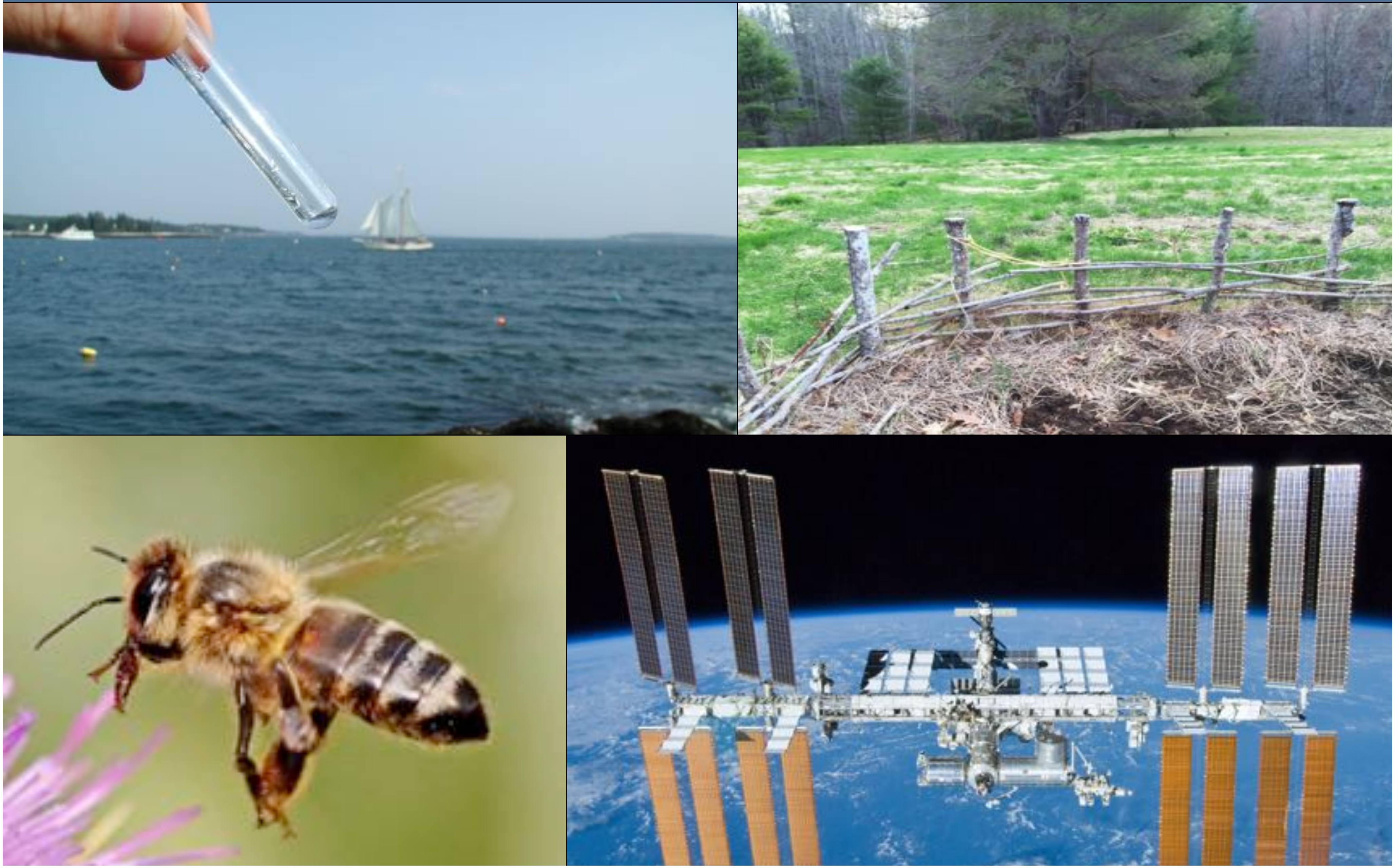
Mission: **make single cell genomics accessible to the broad research community; serve as an engine for discoveries in microbial ecology, evolution, bioprospecting and human health.**

- **First center of its kind, established 2009**
- **Diverse samples: aquatic, soil, organismal microbiomes, etc.**
- **>1,000,000 cells analyzed, representing >70 phyla**
- **60+ publications since 2011, 8 in *Science*, *Nature* & *PNAS***

Locations of SCGC customers

Microbiomes analyzed by SCGC

SCGC
SINGLE CELL GENOMICS CENTER
Reading nature's genetic tales, one cell at a time

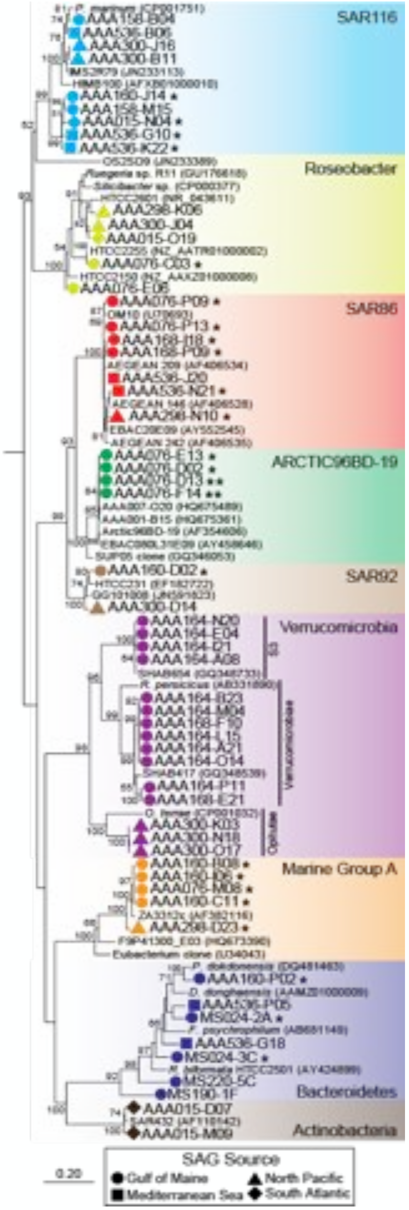


Research example: Surface ocean bacterioplankton

Garcia-Martinez et al. 2012 (PLoS ONE), Swan et al. 2013 (PNAS); Labonté et al. 2015 (ISME J)

Sample collection sites

SAG phylogeny



- **Collected surface samples from the Gulf of Maine, the Mediterranean and the South Atlantic and North Pacific subtropical gyres in 2007-2009**
- **Generated >2,000 single amplified genomes (SAGs) of bacteria & archaea**
- **Genomically sequenced 57 SAGs representing various ubiquitous groups**
- **Used genomic data to analyze metabolism, biogeography and infections**

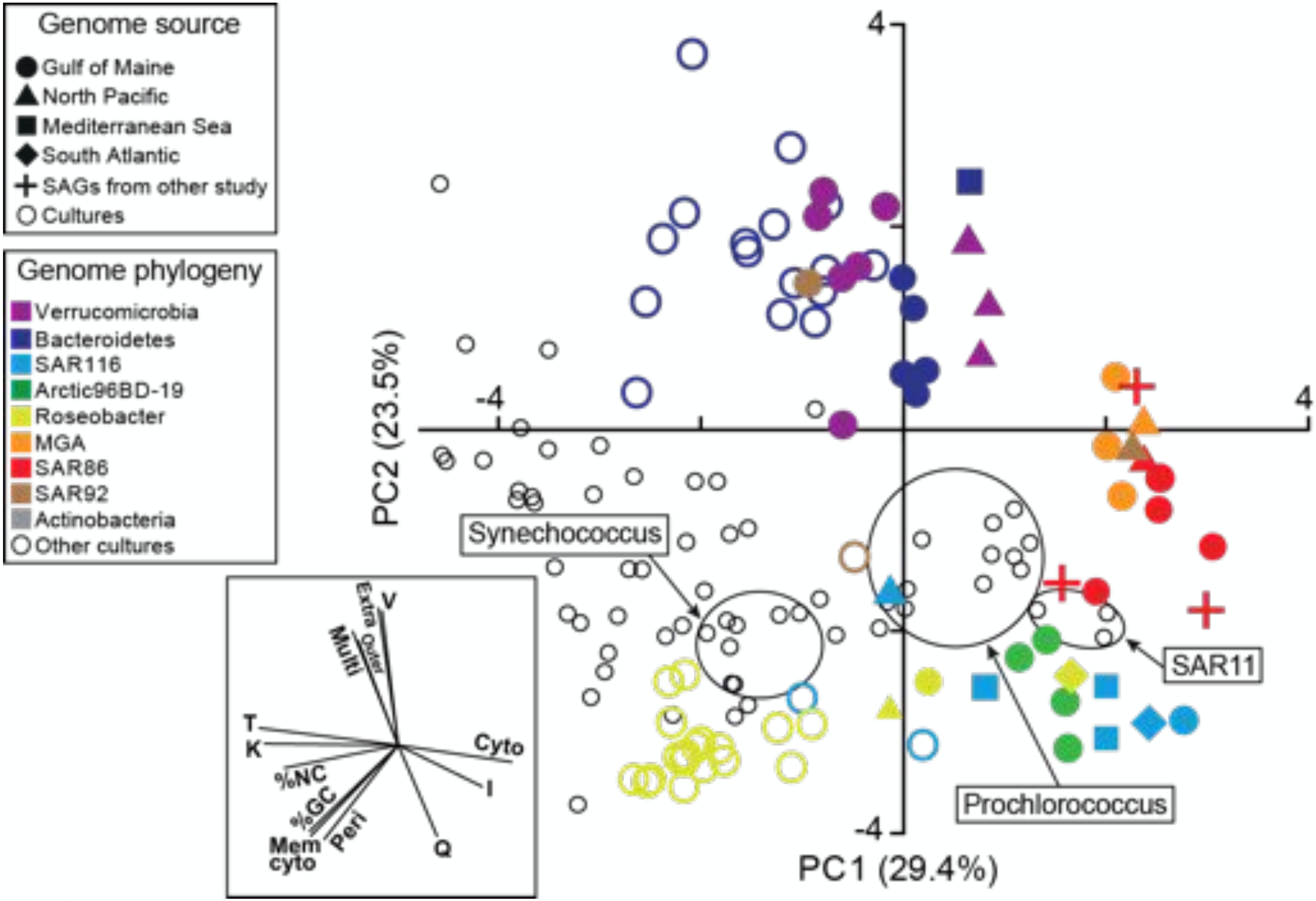
Four tales by 57 marine bacterioplankton cells:

- Metabolism
- Biogeography
- Microevolution
- Interactions

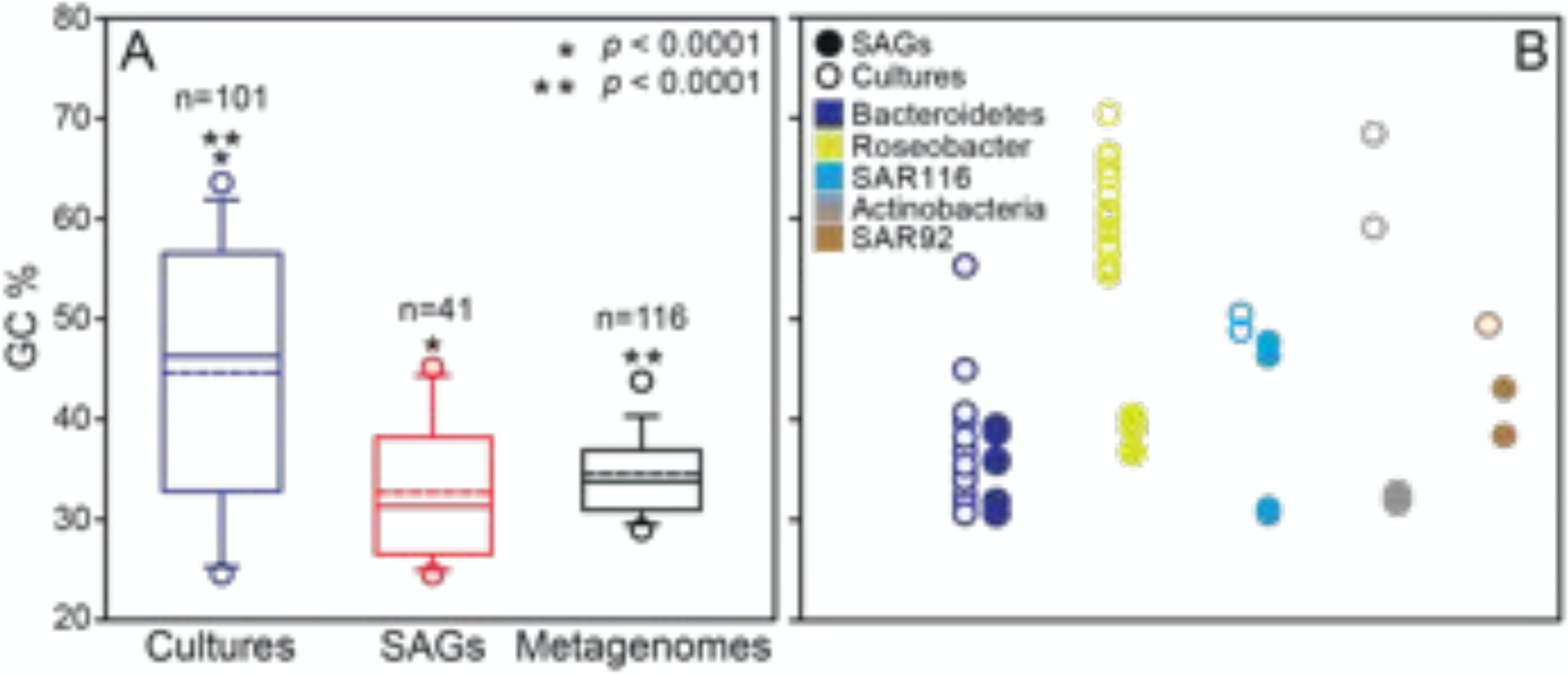
Art by Glynn Gorick

Bigelow | Laboratory for Ocean Sciences

Genomic divergence of cultured and uncultured bacterioplankton

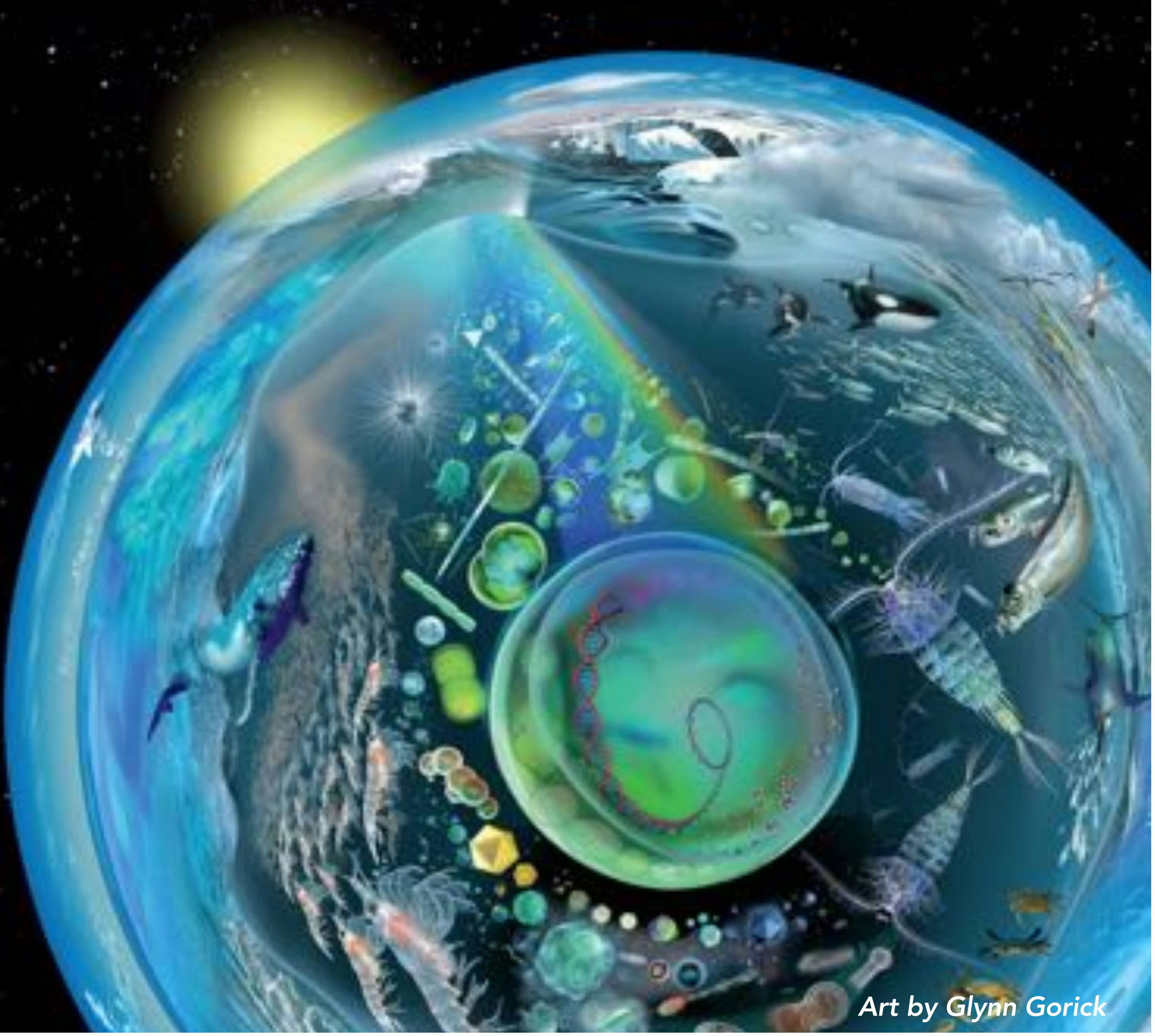


%GC divergence of cultured and uncultured bacterioplankton



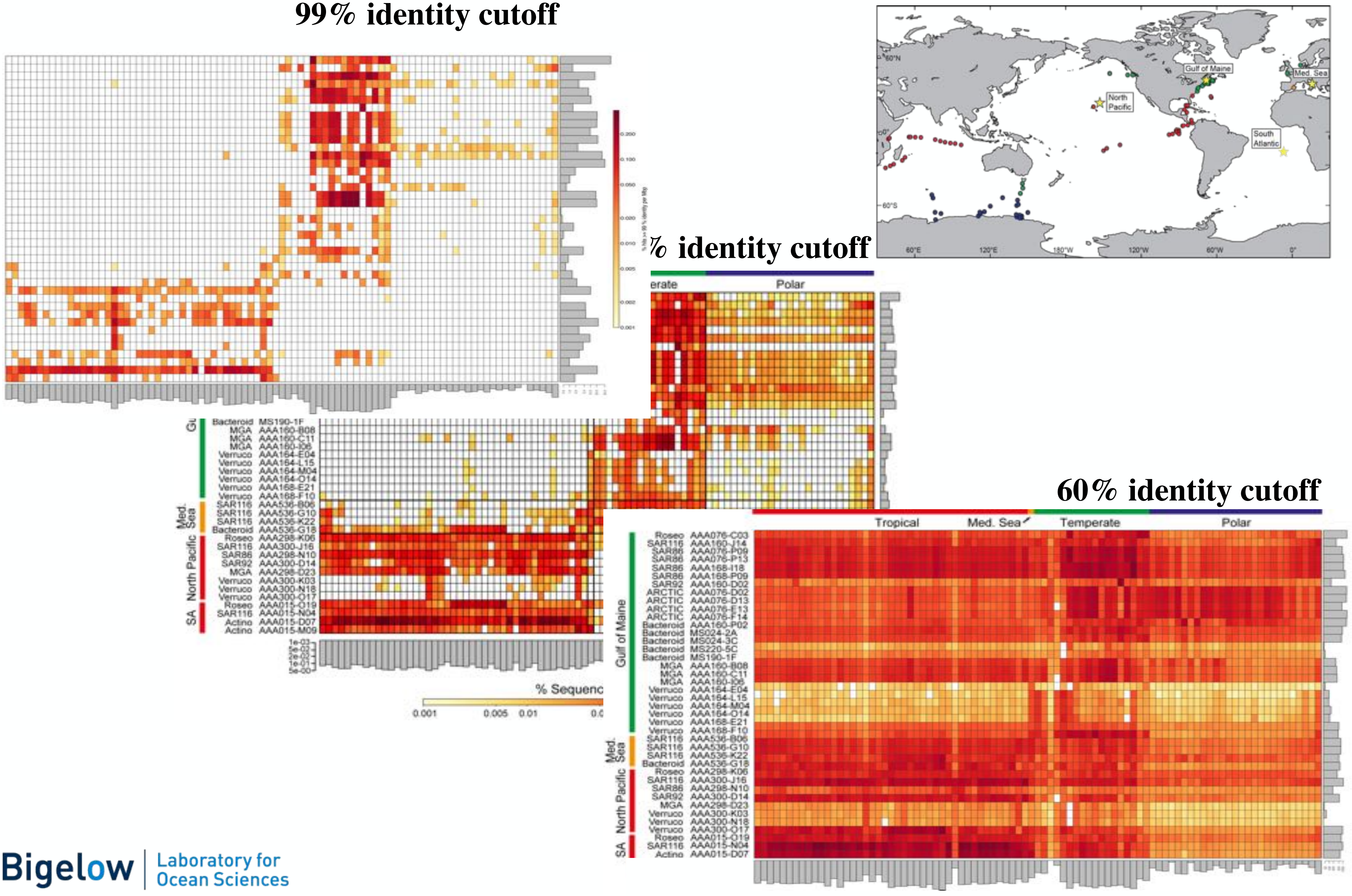
Four tales by 57 marine bacterioplankton cells:

- Metabolism
- Biogeography
- Microevolution
- Interactions


Art by Glynn Gorick

Bigelow | Laboratory for Ocean Sciences

Global biogeography: metagenomic fragment recruitment



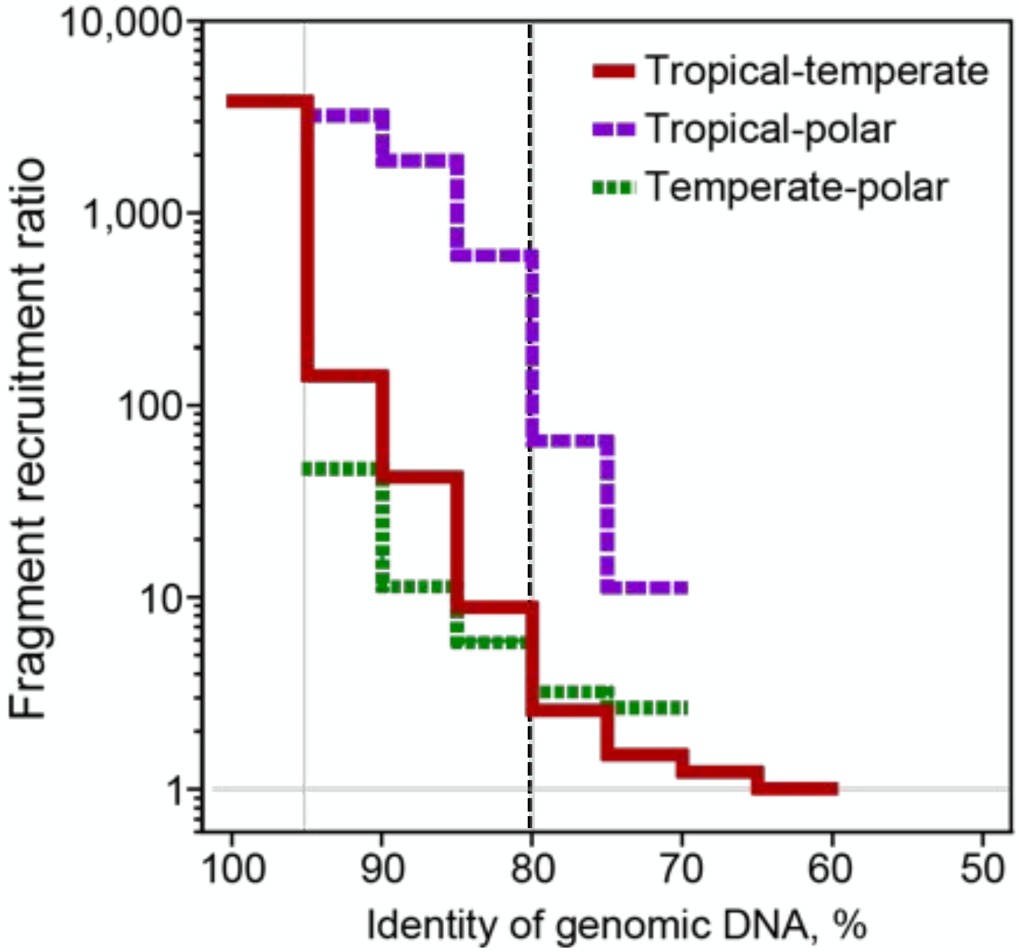
99% identity cutoff

% identity cutoff

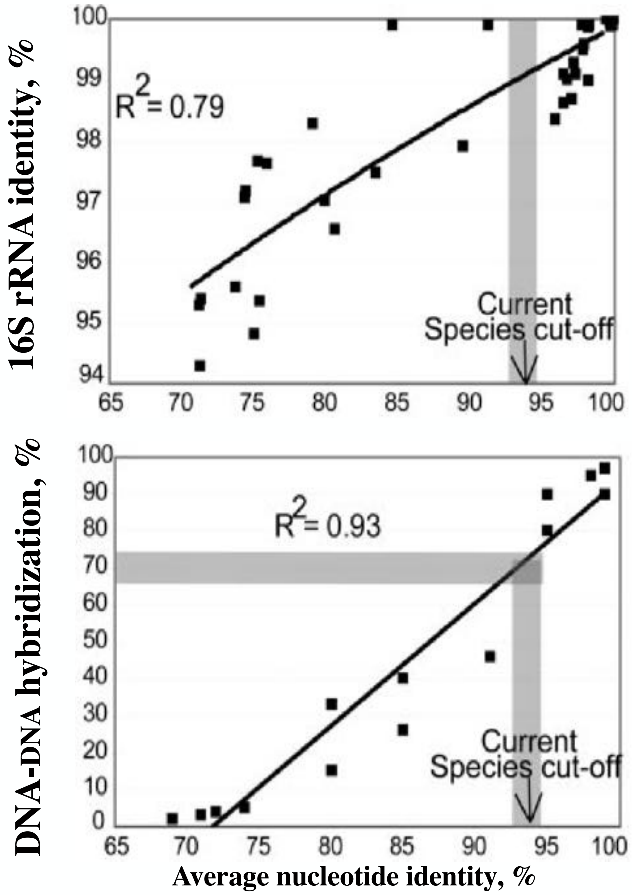
60% identity cutoff

Bigelow Laboratory for Ocean Sciences

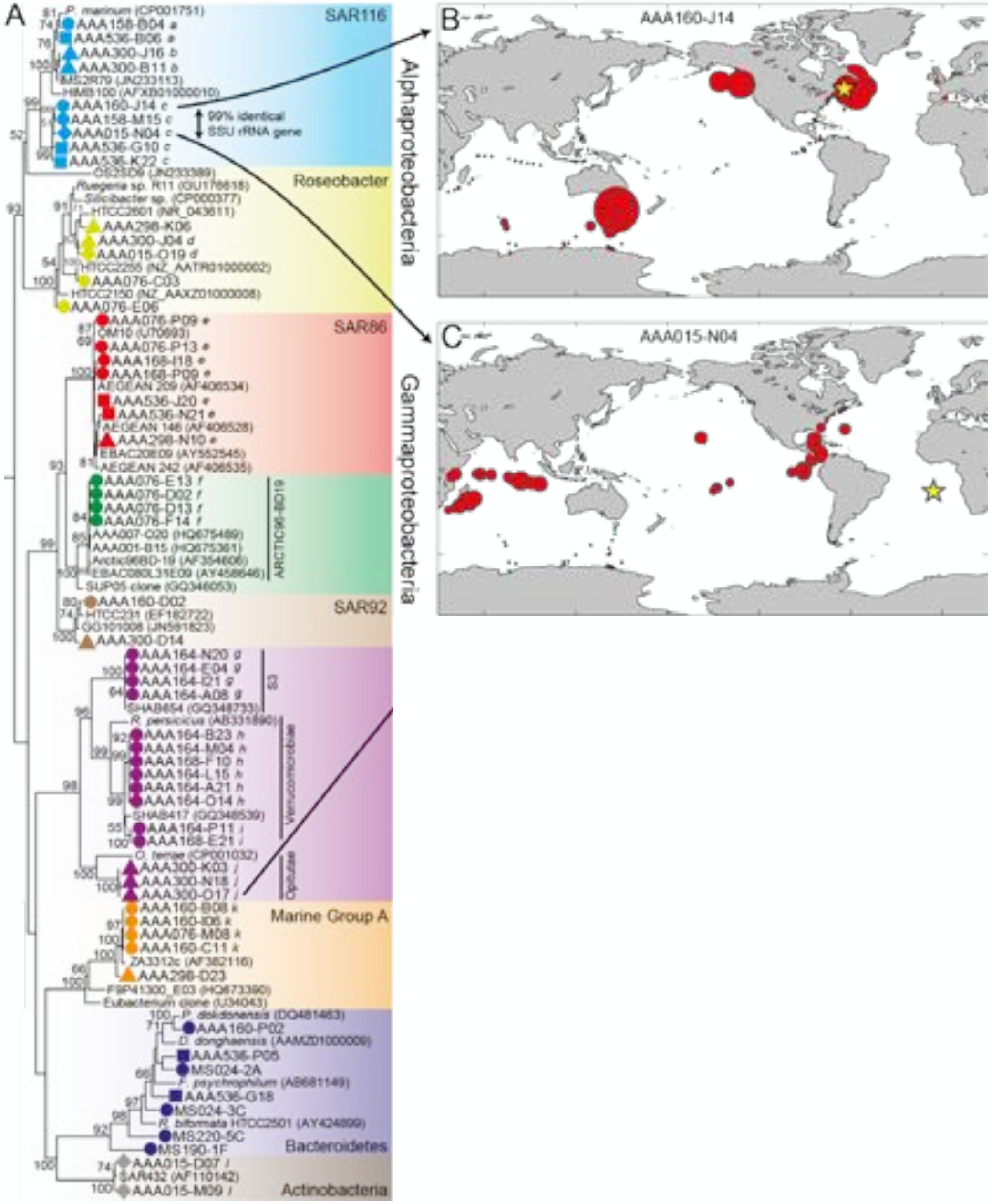
Recruitment ratio at various DNA identity intervals



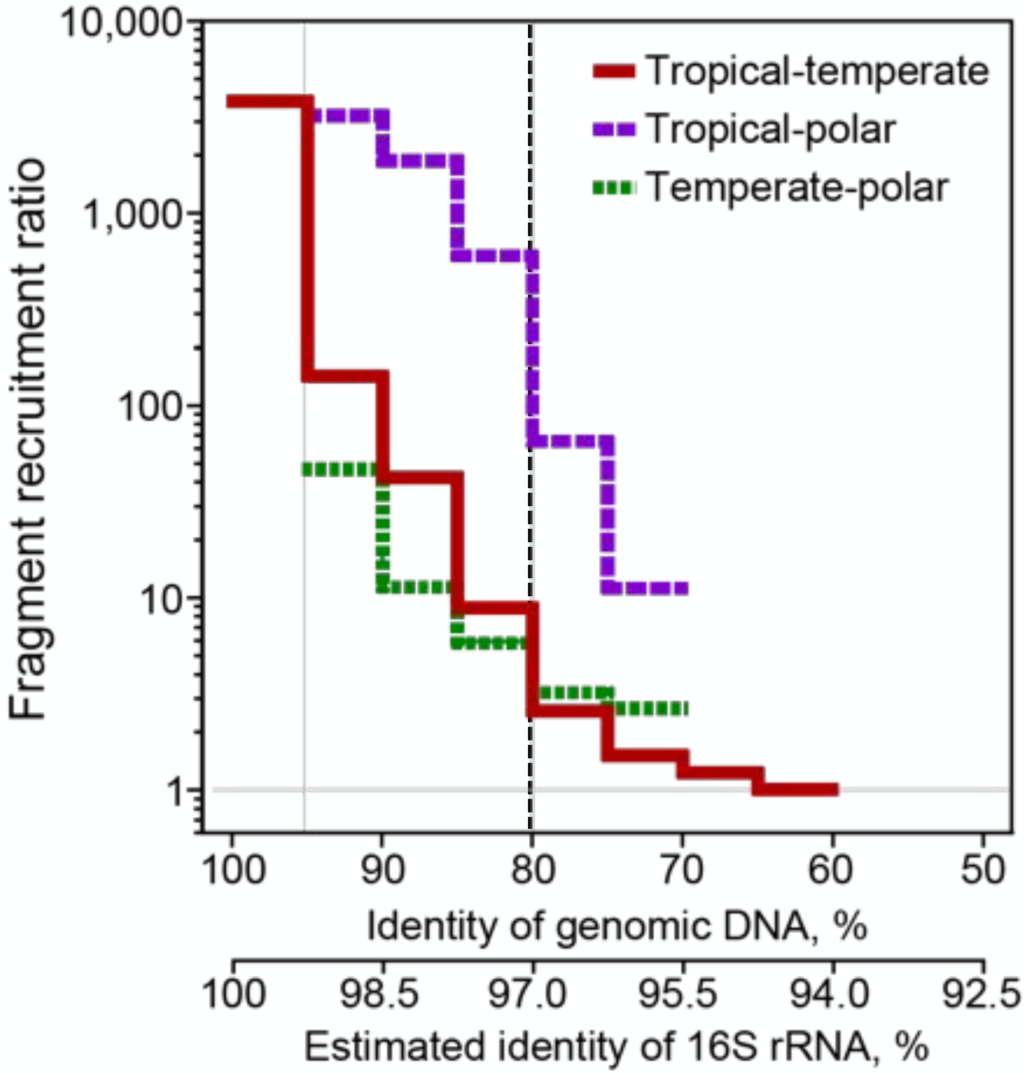
ANI-16S relationship source:
Konstantinidis and Tiedje 2005, PNAS 102:2567



Divergent biogeography of close relatives



Recruitment ratio at various DNA identity intervals



16S rRNA molecular clock:

Ochman and Wilson 1987, J. Mol.Evol. 26:74

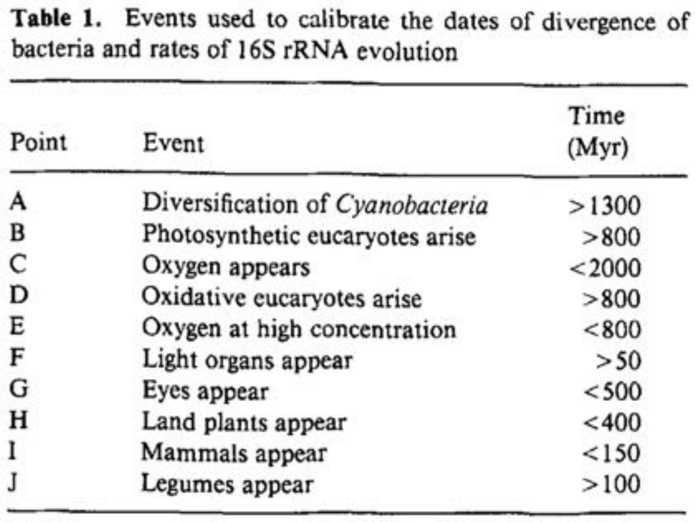


Table 1. Events used to calibrate the dates of divergence of bacteria and rates of 16S rRNA evolution

Point	Event	Time (Myr)
A	Diversification of *Cyanobacteria*	>1300
B	Photosynthetic eucaryotes arise	>800
C	Oxygen appears	<2000
D	Oxidative eucaryotes arise	>800
E	Oxygen at high concentration	<800
F	Light organs appear	>50
G	Eyes appear	<500
H	Land plants appear	<400
I	Mammals appear	<150
J	Legumes appear	>100

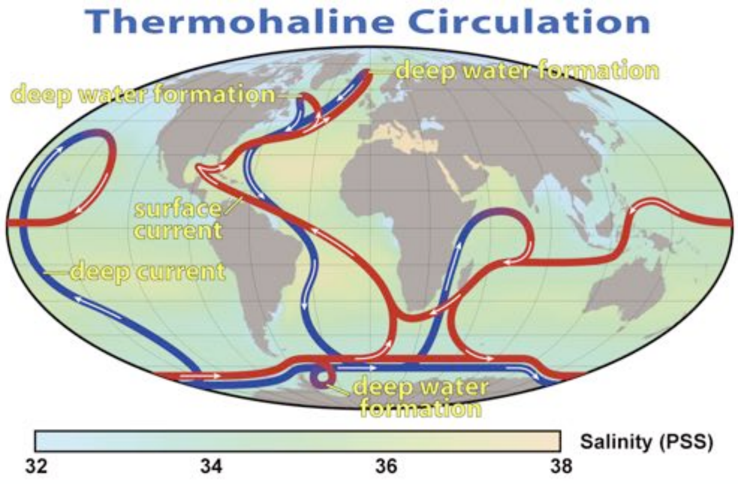
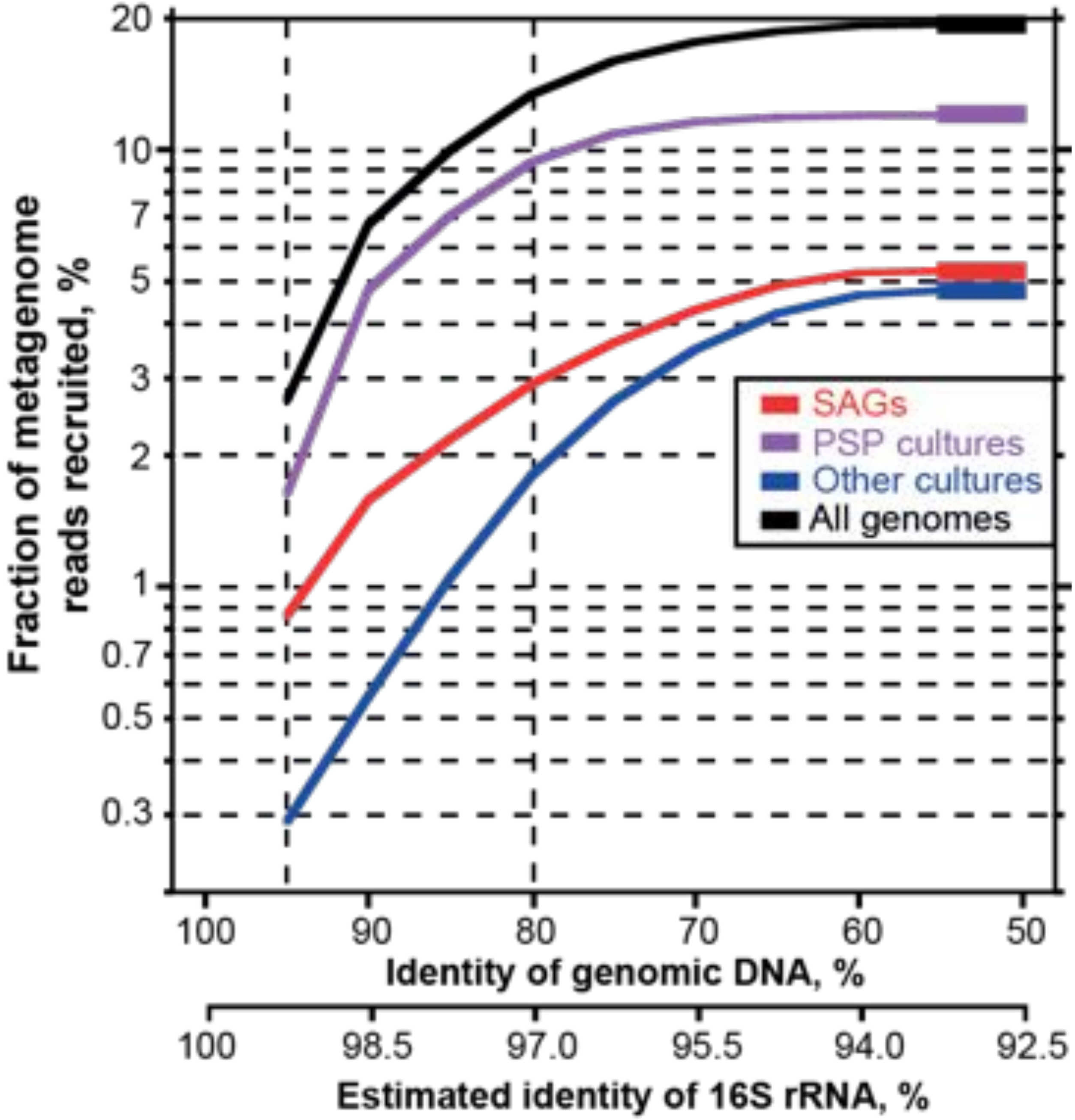
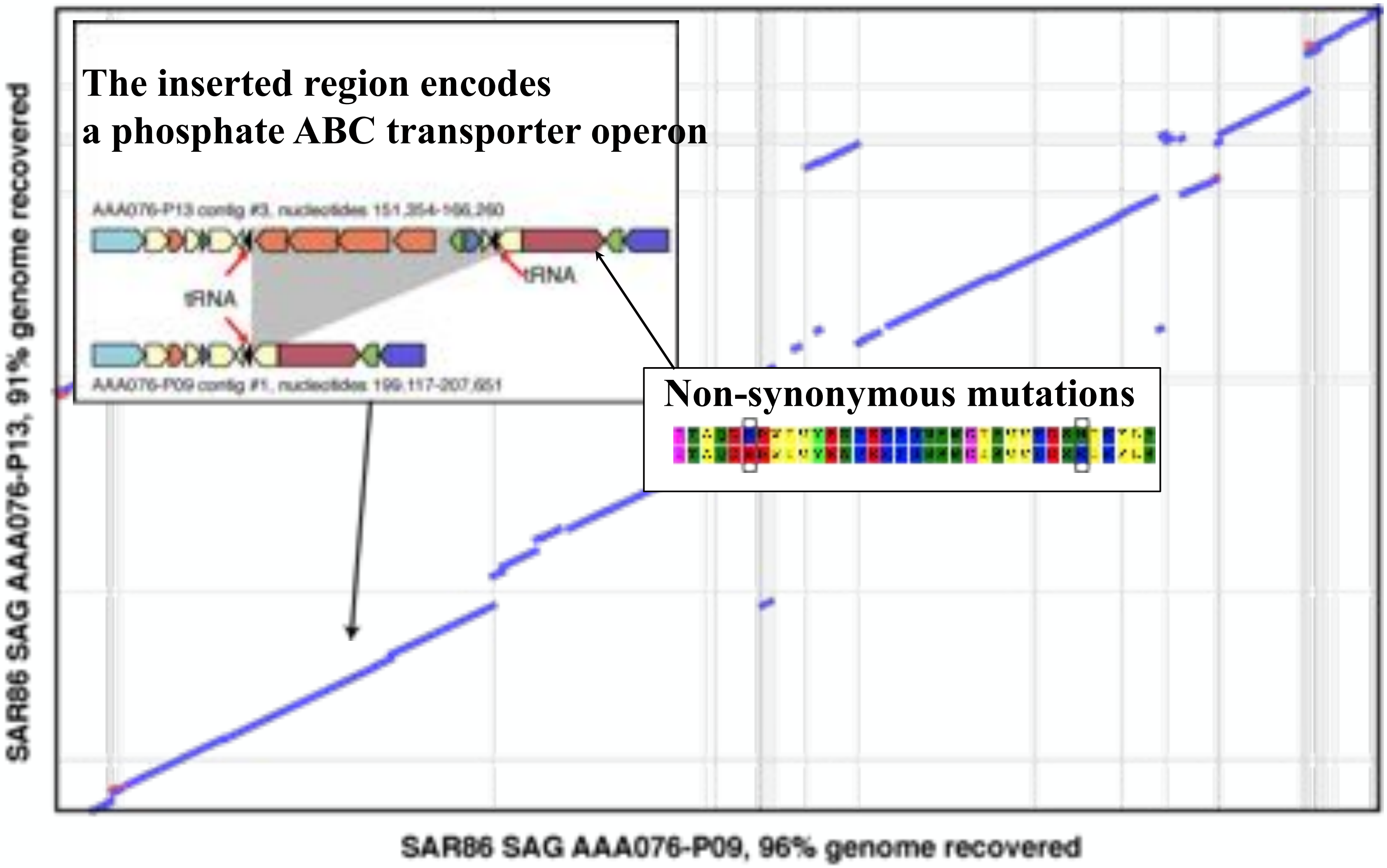


Figure from: Robert Simmon, NASA

We are still only scratching the surface



Genome content differences between two cells of SAR86 from the same drop with 100% identical 16S rRNA

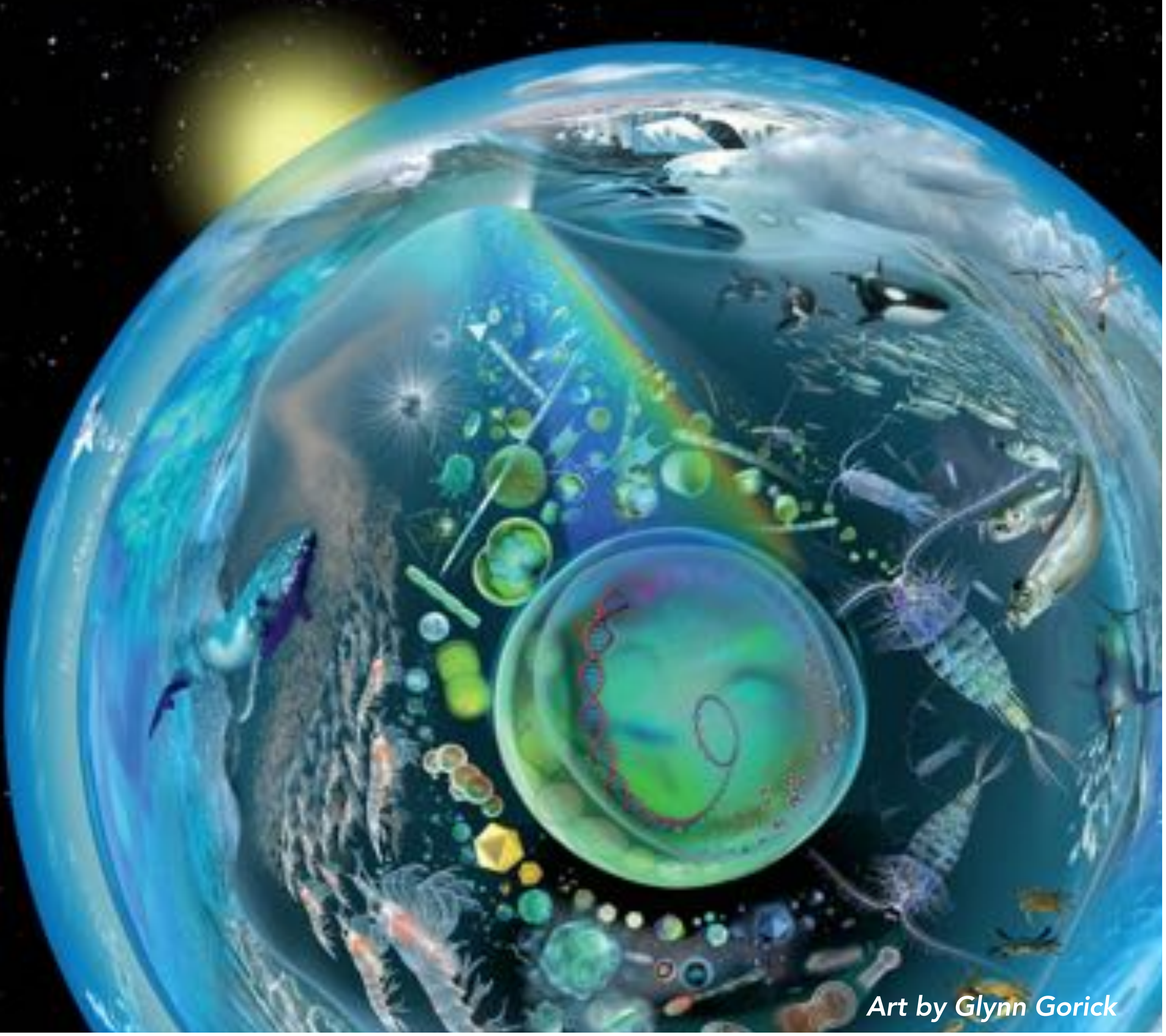


The inserted region encodes a phosphate ABC transporter operon

Non-synonymous mutations

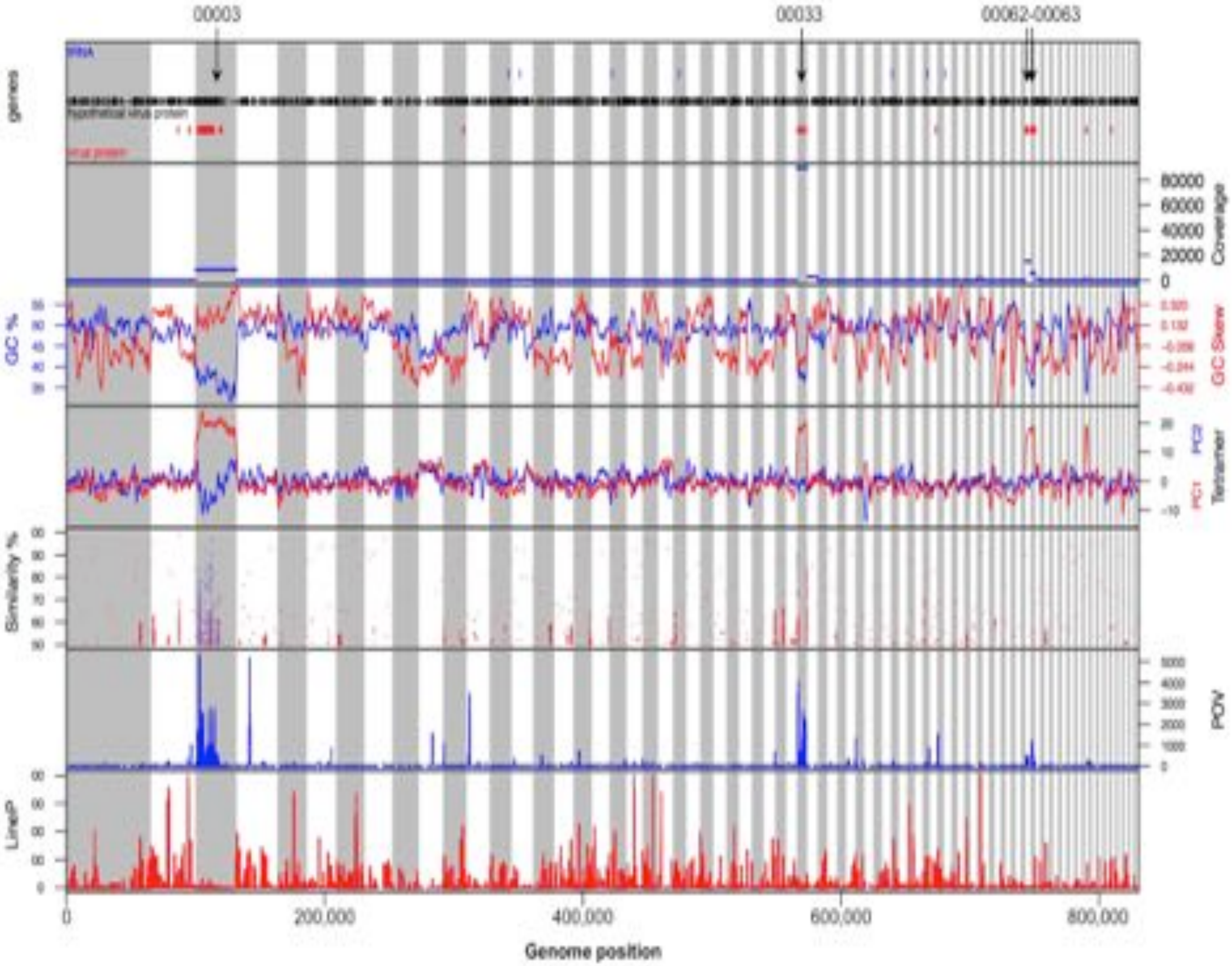
Four tales by 57 marine bacterioplankton cells:

- Metabolism
- Biogeography
- Microevolution
- Interactions


Art by Glynn Gorick

Bigelow | Laboratory for Ocean Sciences

Search for viral DNA in SAGs of bacteria and archaea



Viral marker genes

Sequence coverage depth

GC % and skew

Tetramer frequency

Recruitment of viral versus bacterial metagenomic reads

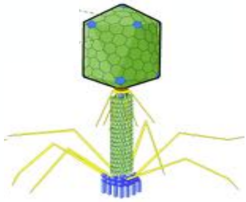
19 out of 57 SAGs (33%) contained viral sequences

10 *Podoviridae* phages

- Marinimicrobia SAR406
- Verrucomicrobia (5)
- Gammaproteobacteria SAR92
- Bacteroidetes
- Roseobacter

5 *Myoviridae* phages

- Verrucomicrobia
- Roseobacter
- Marine Group I crenarchaeon
- Marinimicrobia SAR406
- SAR86



3 *Siphoviridae* phages

- SAR116
- Verrucomicrobia
- Flavobacteria

1 *Phycodnaviridae* virus (likely contaminant)

- Verrucomicrobia

First known viruses of phyla Thaumarchaeota, Marinimicrobia, and Verrucomicrobia

Viruses were also found in SAGs of Alphaproteobacteria, Gammaproteobacteria and Bacteroidetes

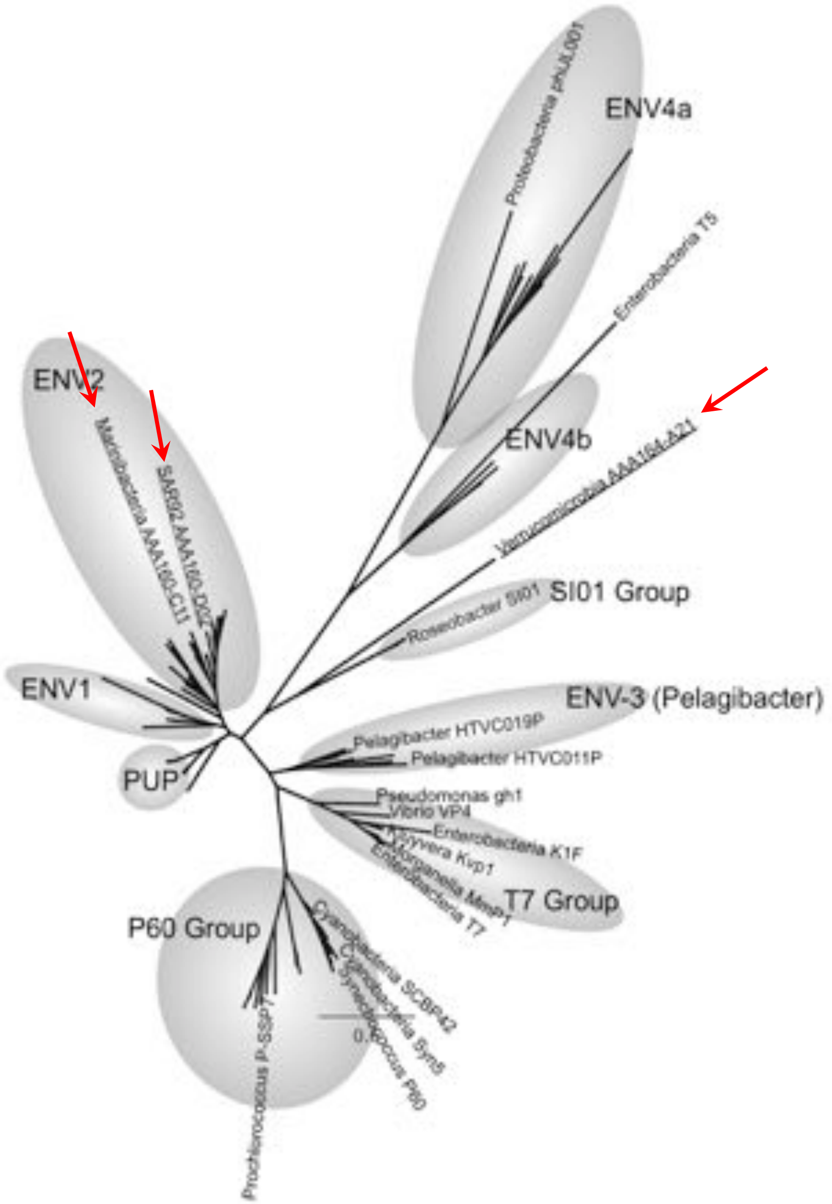
Complete genome recovery of 3 phages

High fragment recruitment of viral metagenomic reads confirmed that most SAG-associated viruses are abundant in the ocean

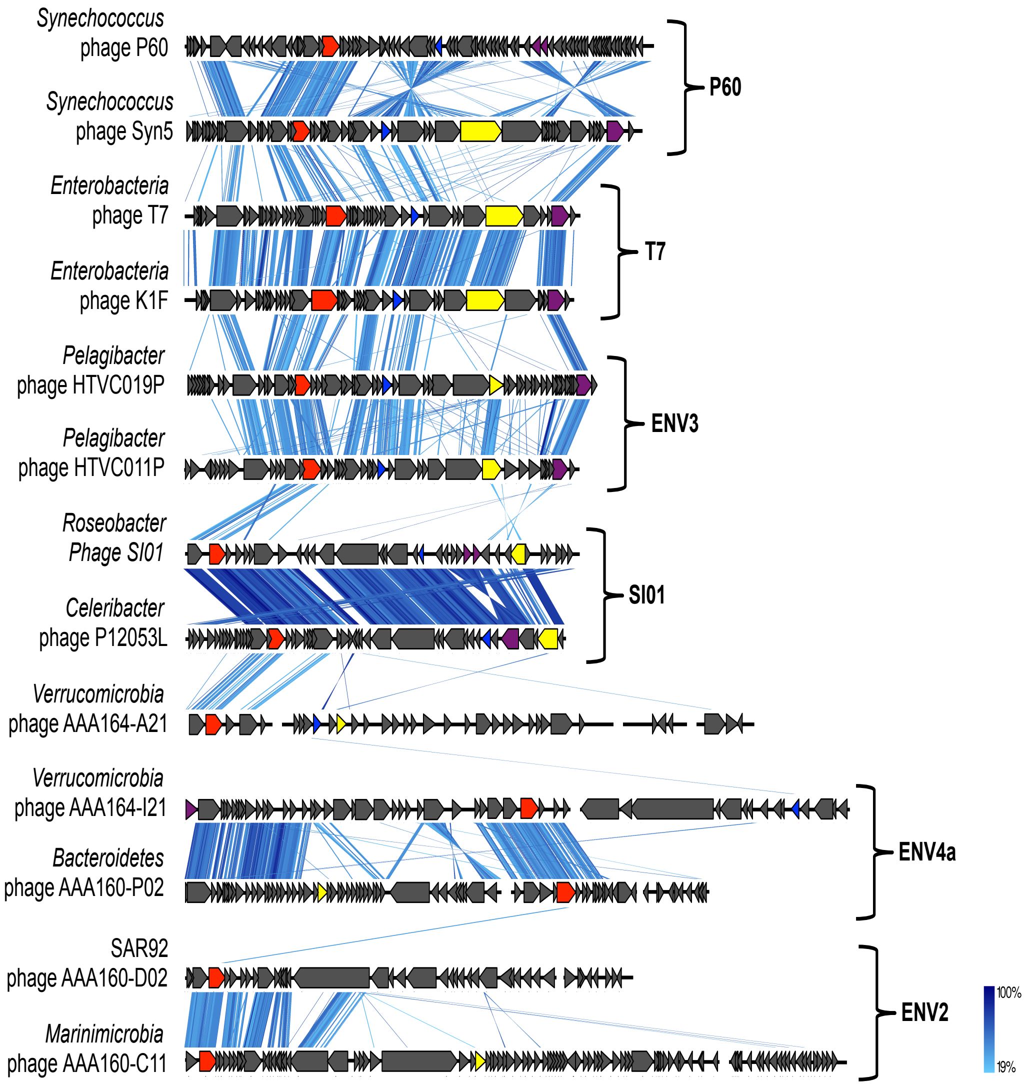
Bigelow Laboratory for Ocean Sciences

Podoviridae phages from SAGs

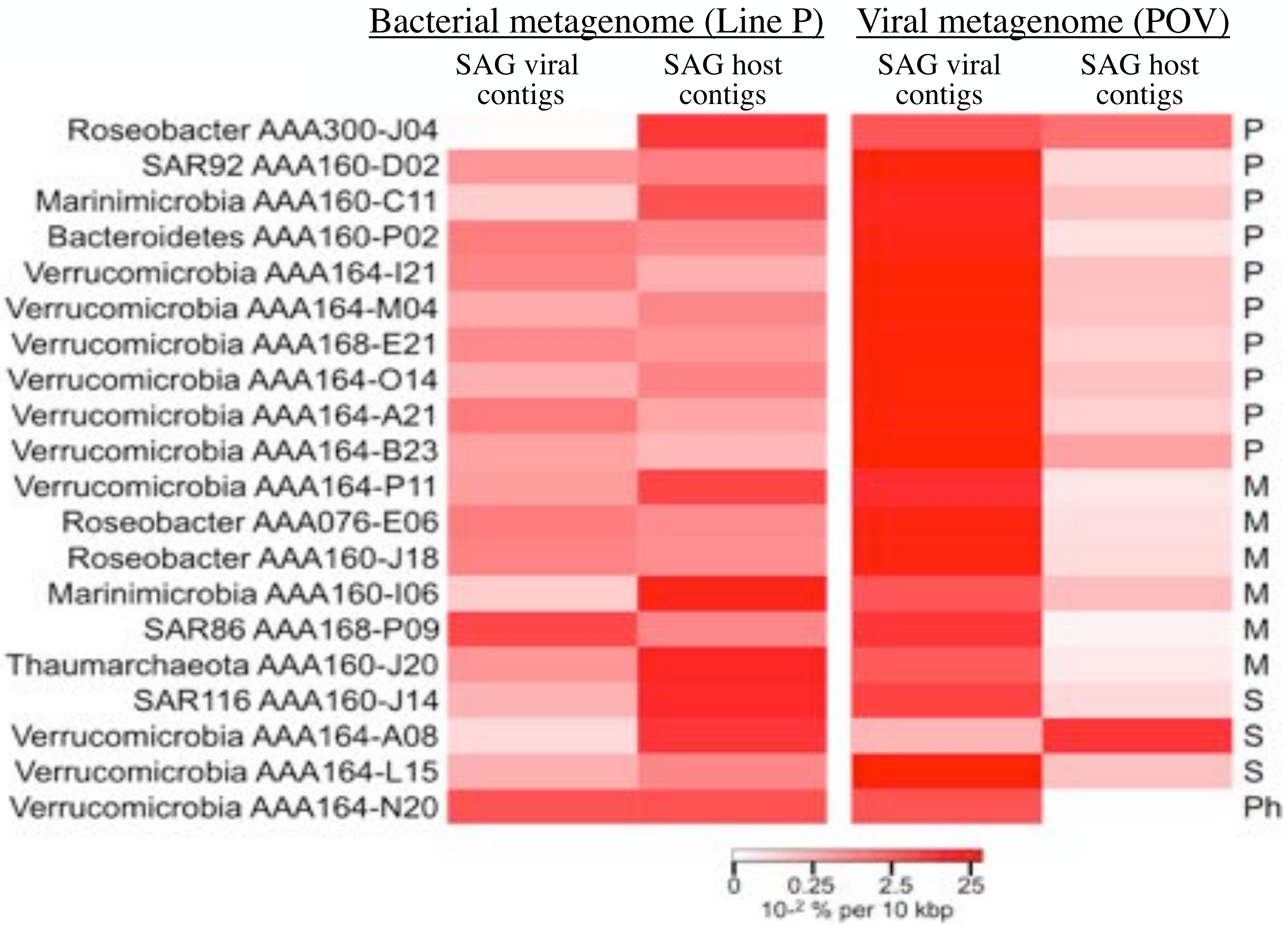
DNA polymerase A phylogeny



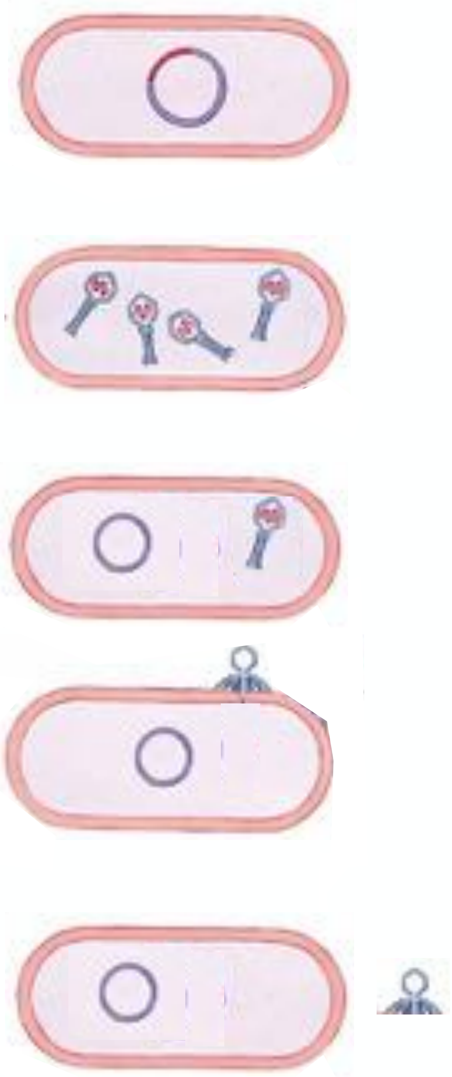
Genomic synteny



Metagenomic fragment recruitment from viral and cellular size fractions (70% nucleotide identity cutoff)



Possible sources of viral sequences in SAGs



1. **Lysogeny**
2. **Lytic infection**
3. **Chronic infection**
4. **Non-infectious attachment**
5. **Co-sort of a cell and a free viral particle**

Possible sources of viral sequences in SAGs

① **Lysogeny**

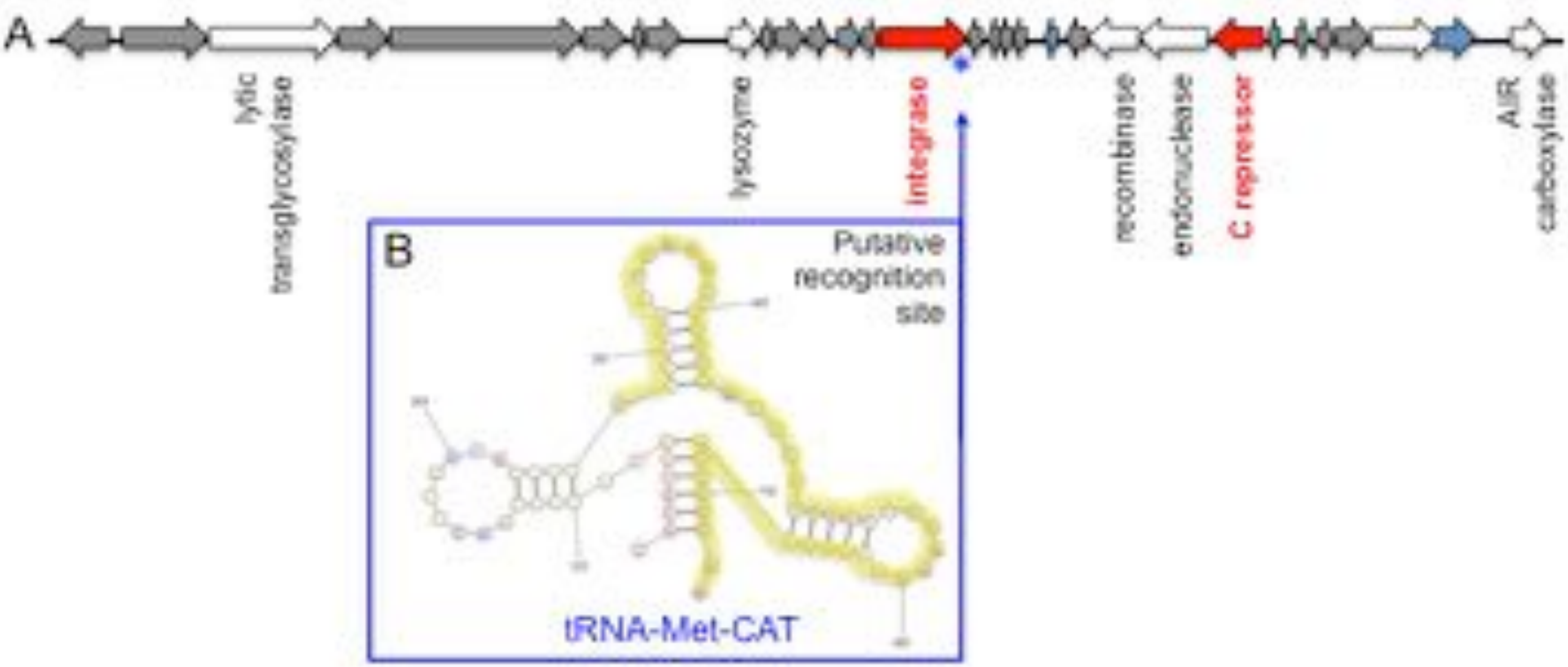
② Lytic infection

③ Chronic infection

④ Non-infectious attachment

⑤ Co-sort of a cell and a free viral particle

Potentially lysogenic phage in a SAR116 SAG



Possible sources of viral sequences in SAGs

1. Lysogeny

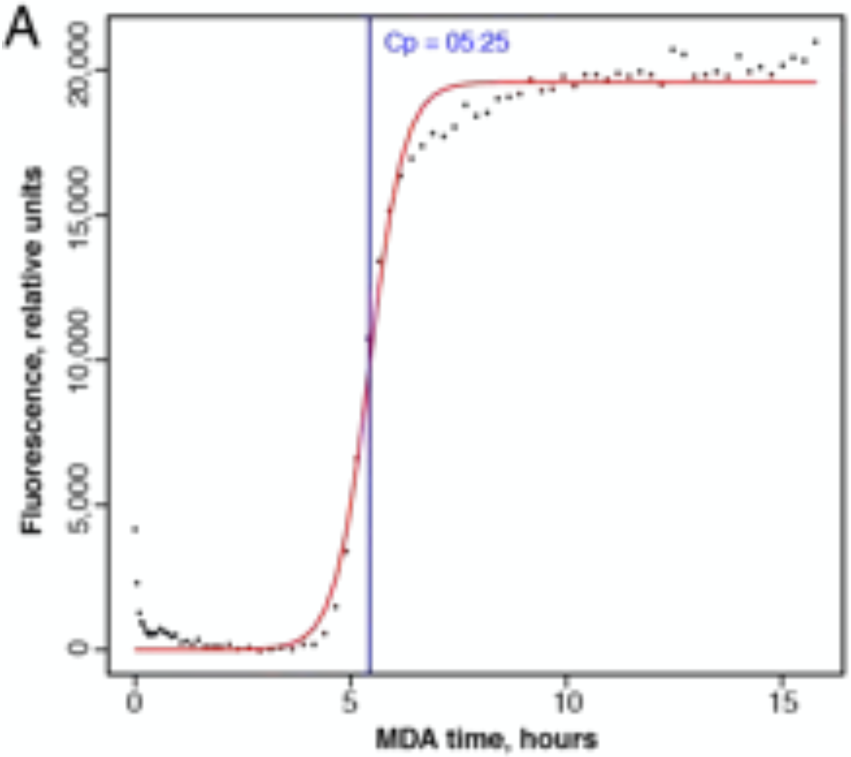
2. **Lytic infection**

3. Chronic infection

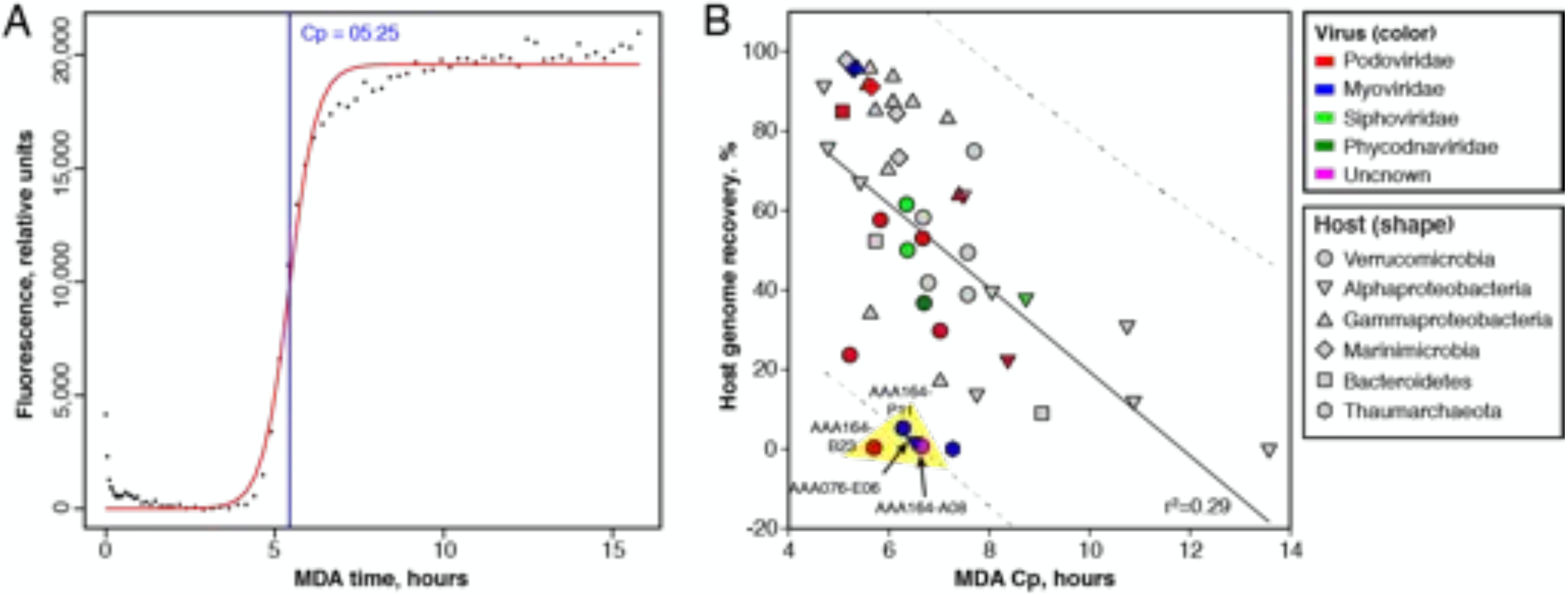
4. Non-infectious attachment

5. Co-sort of a cell and a free viral particle

Correlation between MDA Cp and genome recovery



Correlation between MDA Cp and genome recovery



Possible sources of viral sequences in SAGs

1. Lysogeny
2. Lytic infection
3. **Chronic infection**
4. Non-infectious attachment
5. Co-sort of a cell and a free viral particle

Possible sources of viral sequences in SAGs

1. Lysogeny

2. Lytic infection

3. Chronic infection

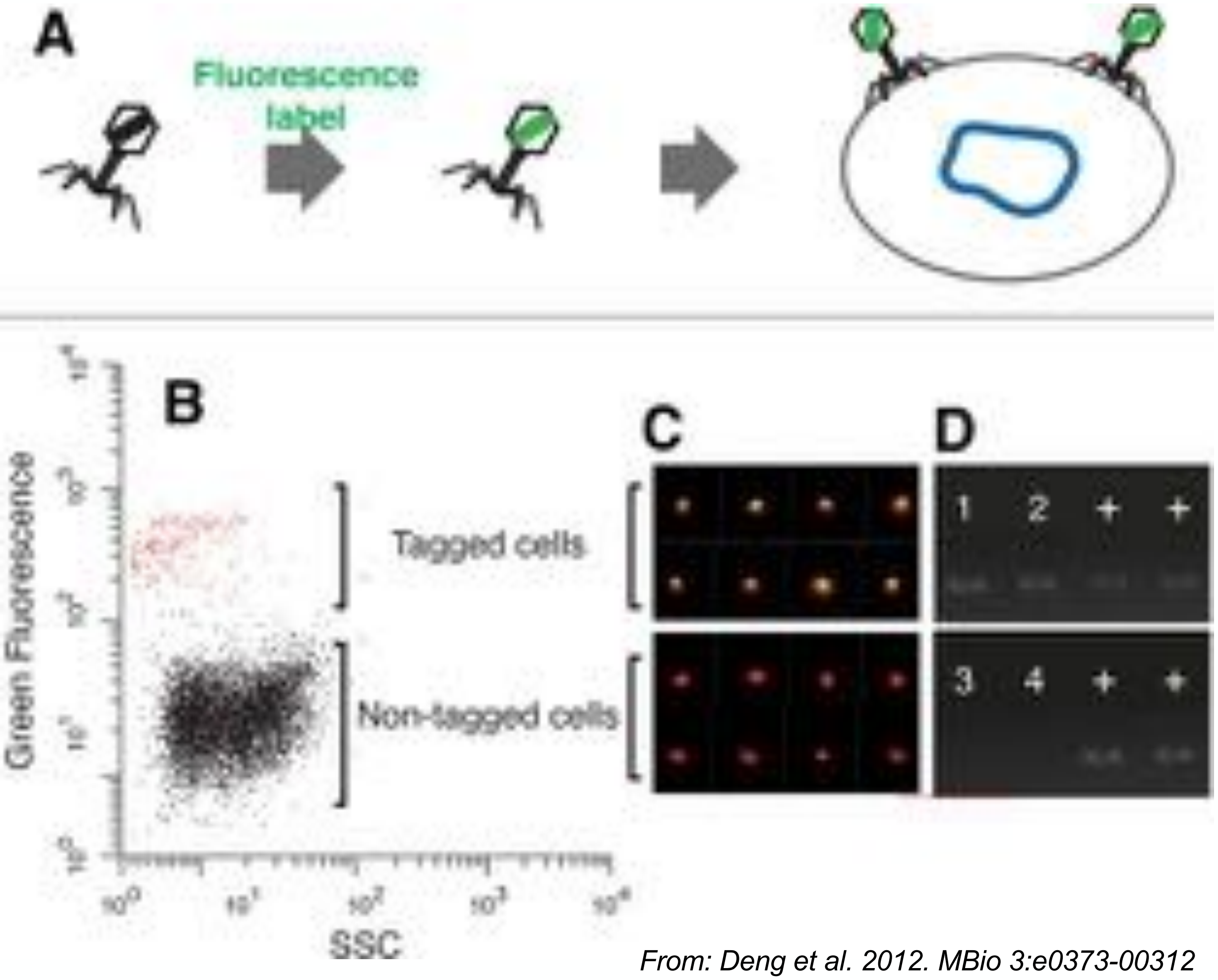
4. **Non-infectious attachment**

5. Co-sort of a cell and a free viral particle

Potential for non-infectious viral attachment to cell

This process is poorly understood

Deng et al. (2012) found that non-specific attachment is rare and that most attachment is infectious

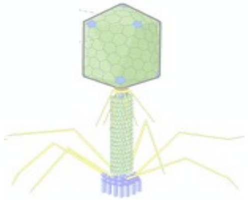


From: Deng et al. 2012. MBio 3:e0373-00312

19 out of 57 SAGs (33%) contained viral sequences

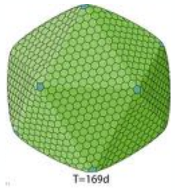
10 *Podoviridae* phages
- Marinimicrobia SAR406
- Verrucomicrobia (5)
- Gammaproteobacteria SAR92
- Bacteroidetes
- Roseobacter

5 *Myoviridae* phages
- Verrucomicrobia
- Roseobacter
- Marine Group I crenarchaeon
- Marinimicrobia SAR406
- SAR86



3 *Siphoviridae* phages
- SAR116
- Verrucomicrobia
- Flavobacteria

1 *Phycodnaviridae* virus
- Verrucomicrobia



First known viruses of Thaumarchaeota, Marinimicrobia, Verrucomicrobia and Gammaproteobacteria clusters SAR86 42 and SAR92

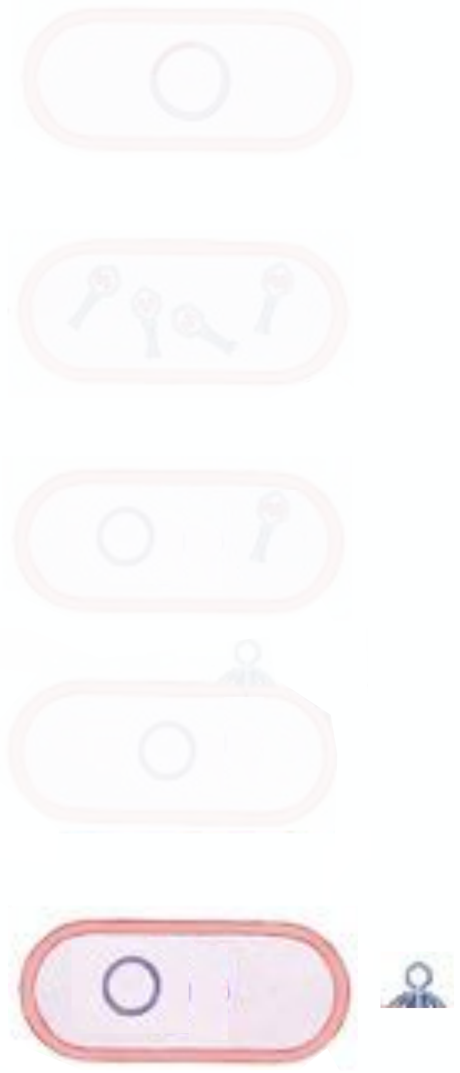
Viruses were also found in SAGs of Alphaproteobacteria and Bacteroidetes

Complete genome recovery of 3 phages

High fragment recruitment of viral metagenomic reads confirmed that most SAG-associated viruses are abundant in the ocean

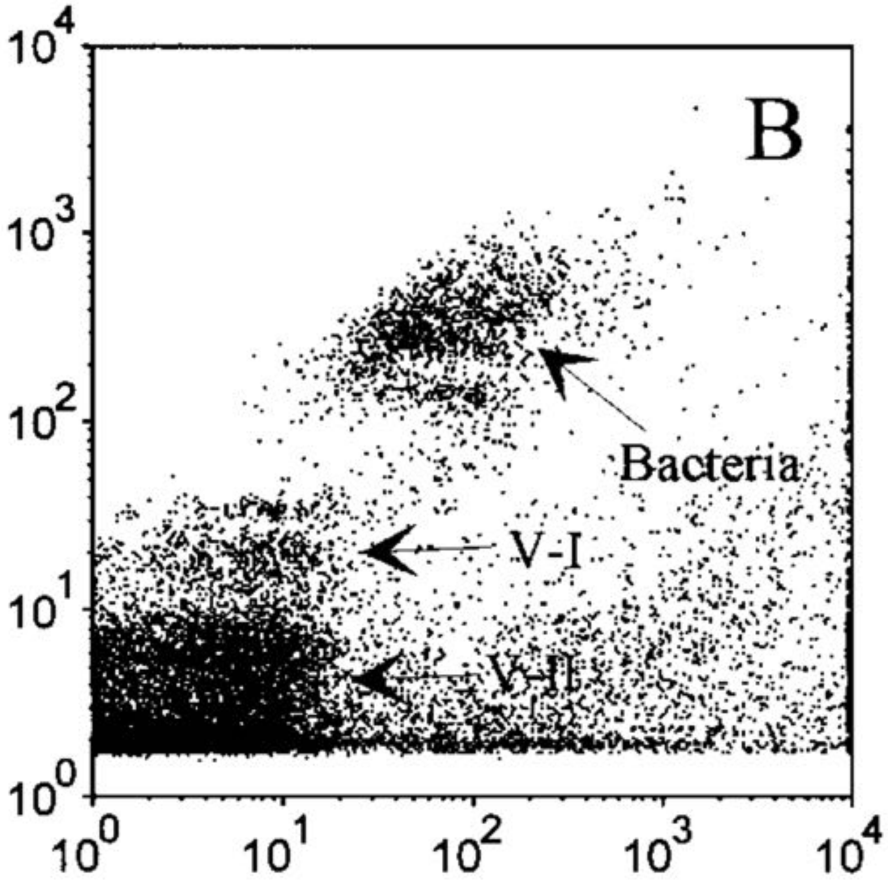
Possible sources of viral sequences in SAGs

1. Lysogeny
2. Lytic infection
3. Chronic infection
4. Non-infectious attachment
5. **Co-sort of a cell and a free viral particle**



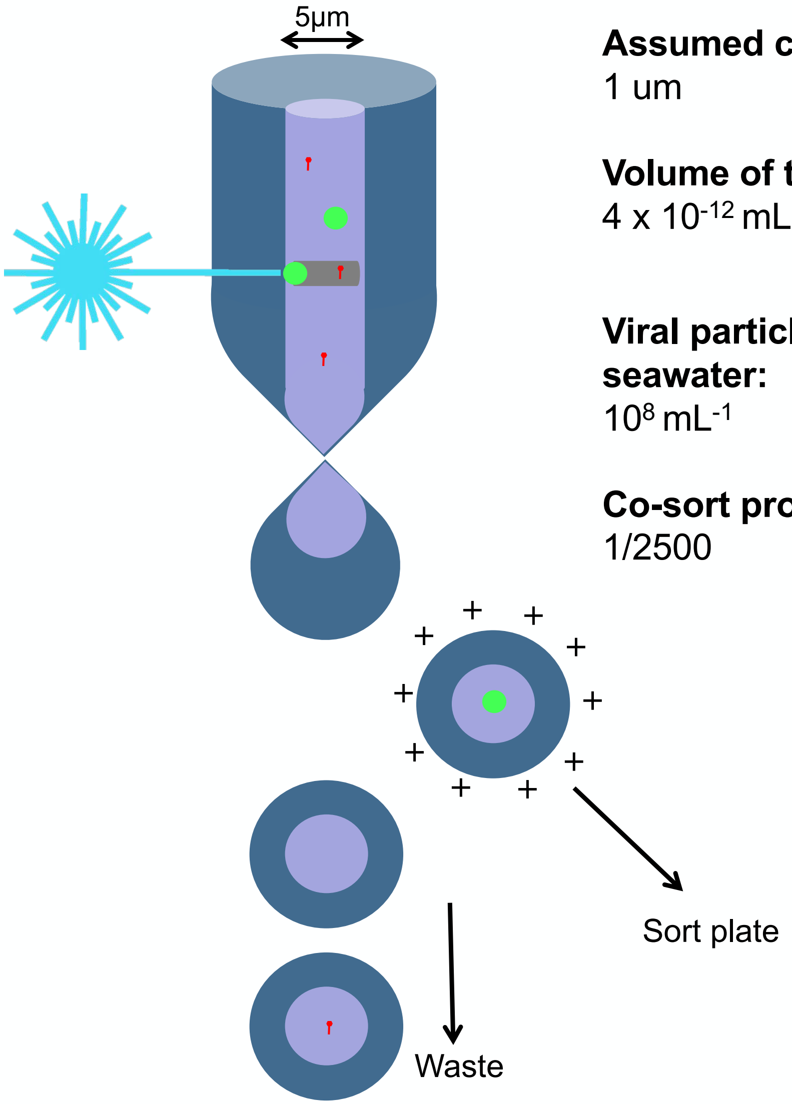
Risk of co-sorting cells free viral particles

Make viral particles visible to the instrument



From Marie et al. 1999

Use single-drop sort mode



Assumed cell diameter:
1 um

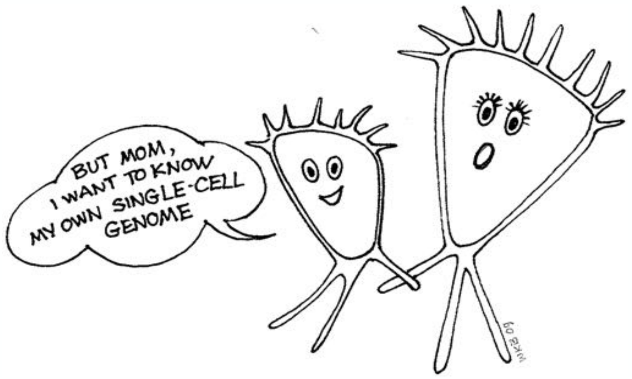
Volume of the shadow:
4 x 10^{-12} mL

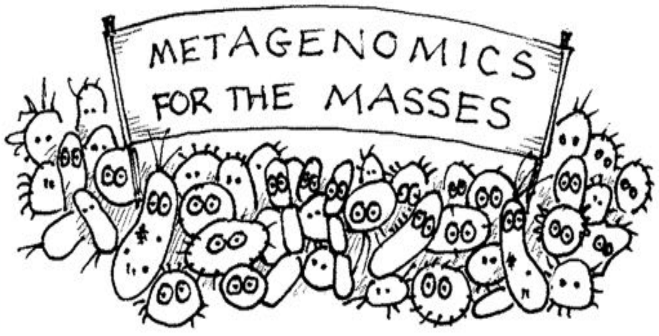
Viral particles in 1:10 dilute seawater:
10^8 mL^{-1}

Co-sort probability:
1/2500

Advantages of single cell genomics

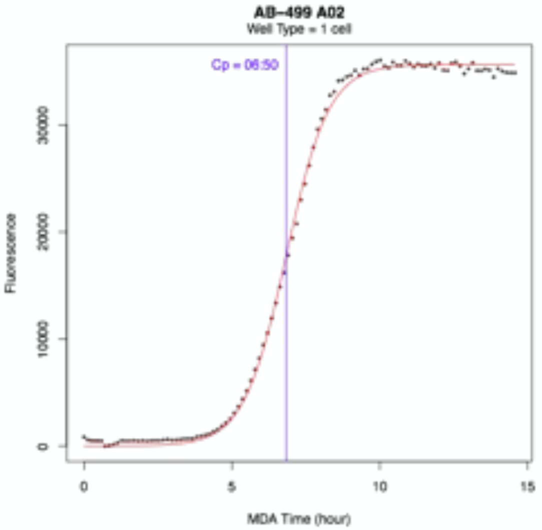
- Independent of cultivation
- Independent of unverified microevolutionary assumptions
- Compatible with high microbial community complexity
- Embraces intracellular genetic complexity
- Requires minimal field sample quantities
- Can be integrated with single cell phenomics





Whole genome amplification kinetics

Individual reaction kinetics

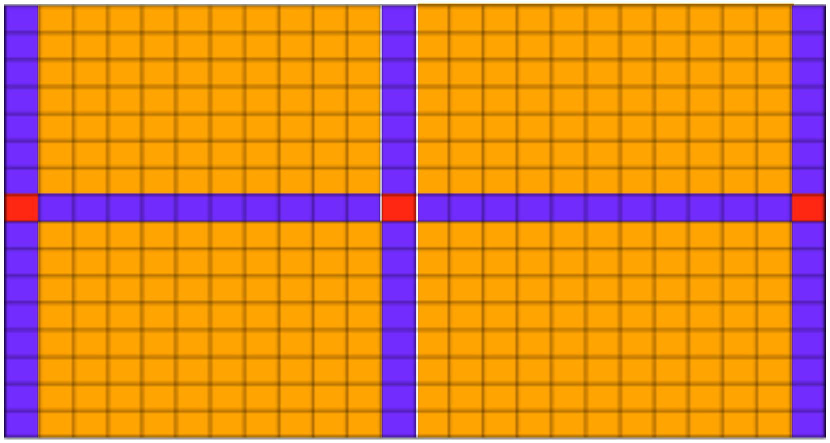


384-well plate layout

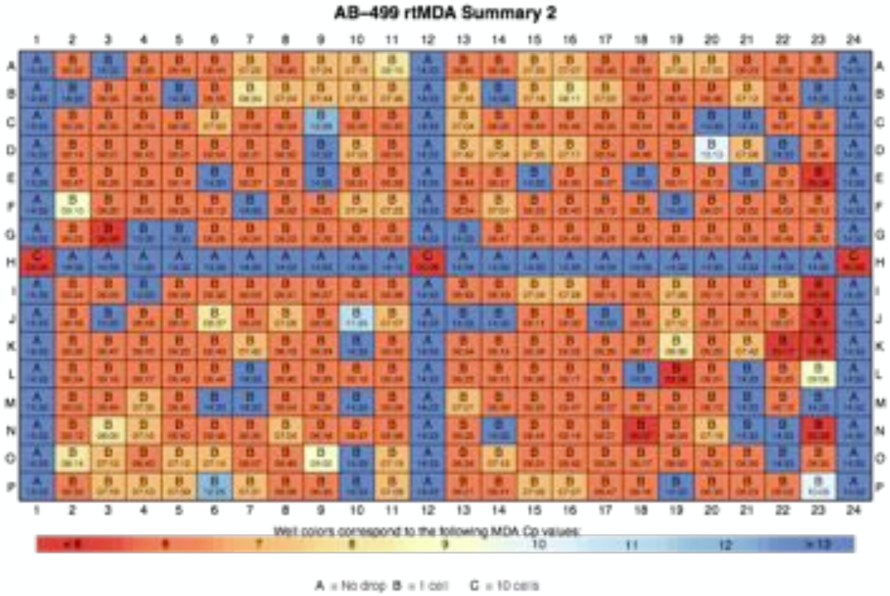
Negative controls: empty wells

Positive controls: 10 cells per well

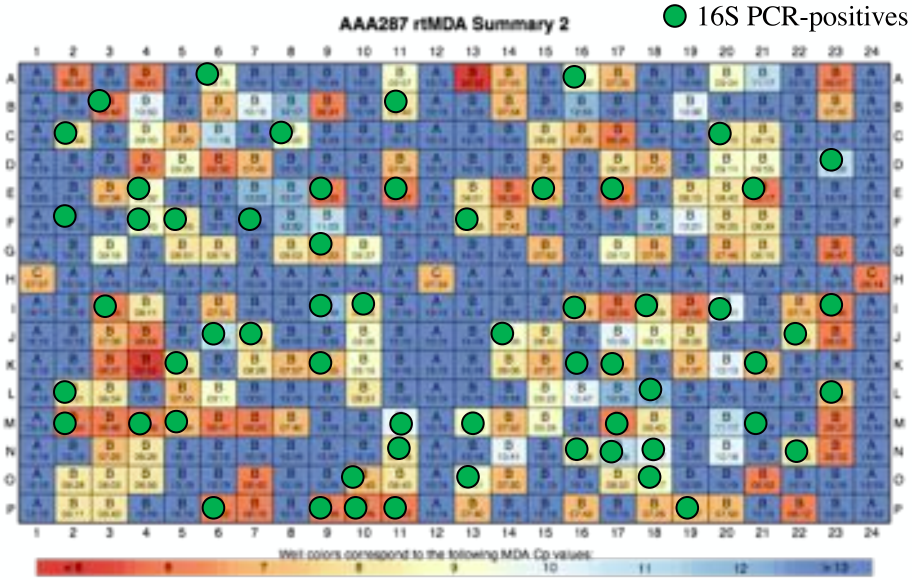
Wells containing single cells



Real data: *Tetraselmis* culture



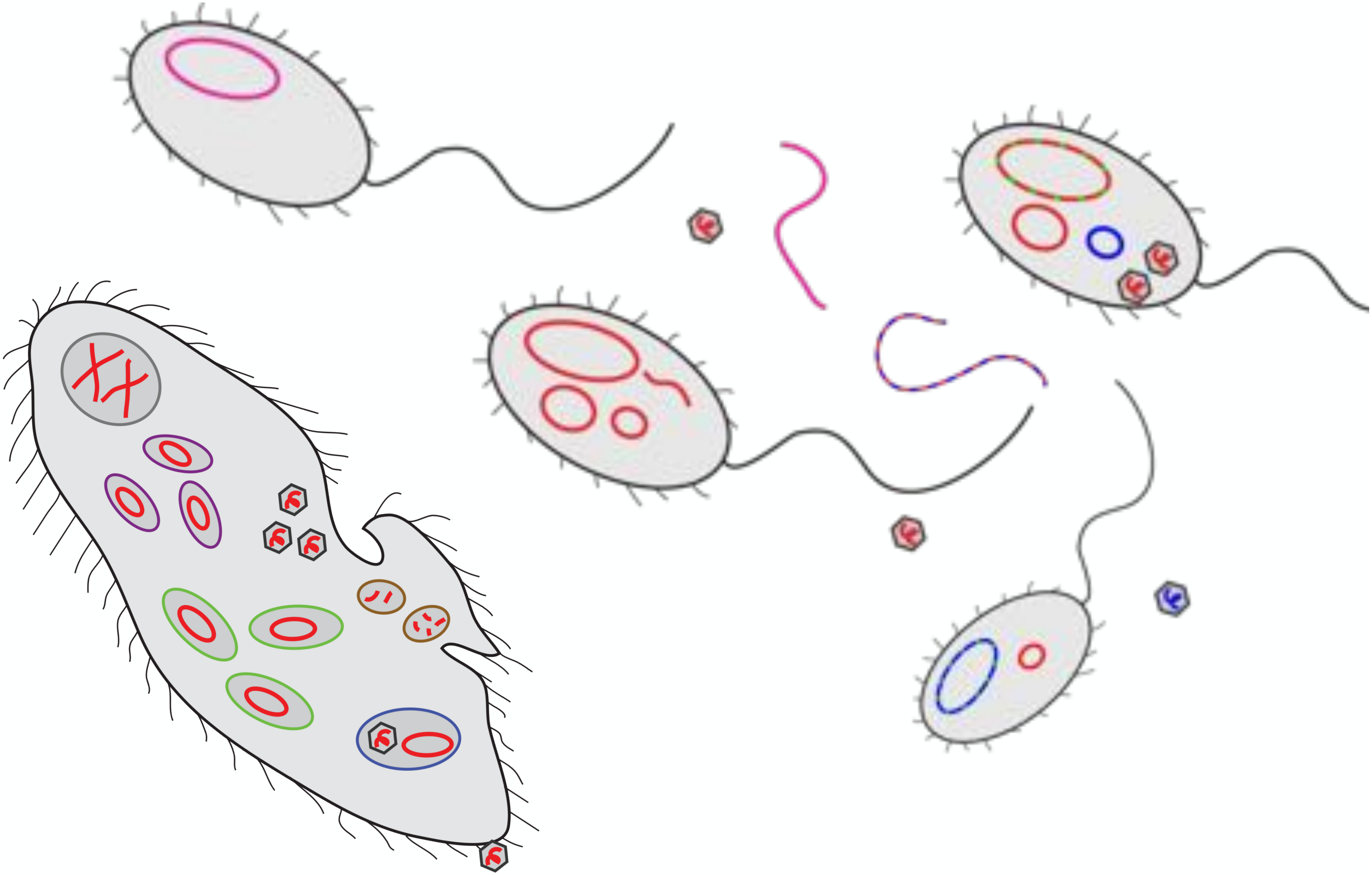
Real data: marine prokaryotes from 3000 m. depth



Propensity to microbial composition biases

Bias source	**16S Tags**	**Shotgun metagenomics**	**Single cell genomics**
Filter type/ sort gate	*Sometimes*	*Sometimes*	*Sometimes*
gDNA extraction	**High**	**High**	**High**
gDNA amplification	*Sometimes*	*Sometimes*	*Sometimes*
Targeted PCR: primer mismatches, inserts, secondary structures, multiple gene copies	**High**	*No*	*Low or No*
DNA sequencing	**High**	**High**	*Low*
Reference databases	*Low*	**High**	*Low*

What do the various techniques tell about microbiomes?

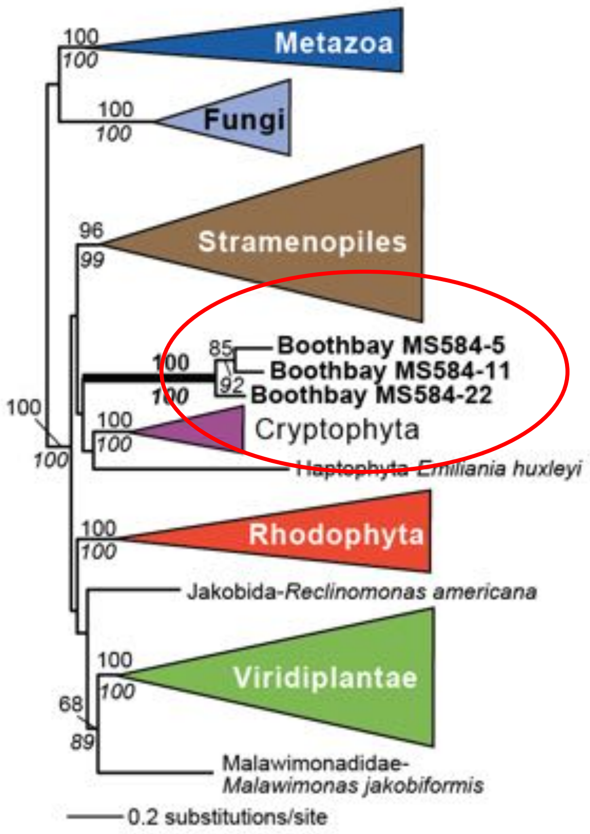


Bigelow | Laboratory for Ocean Sciences

Single cell metagenomics of Picozoa

Yoon et al., *Science 2011*

MLS phylogeny



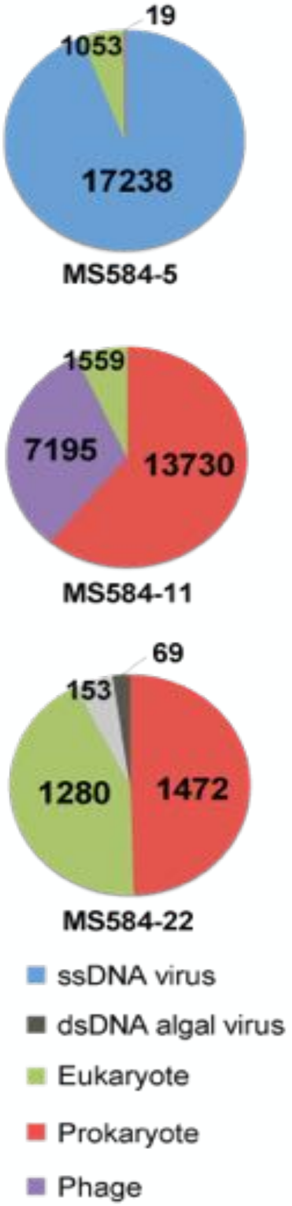
Sequenced three Picobiliphyta SAGs

No evidence for autotrophy

Feed on bacteria and large viruses

Novel, picobiliphyte-infecting nanovirus

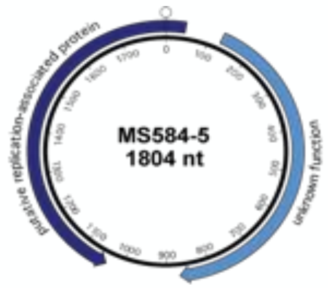
Shotgun read sources

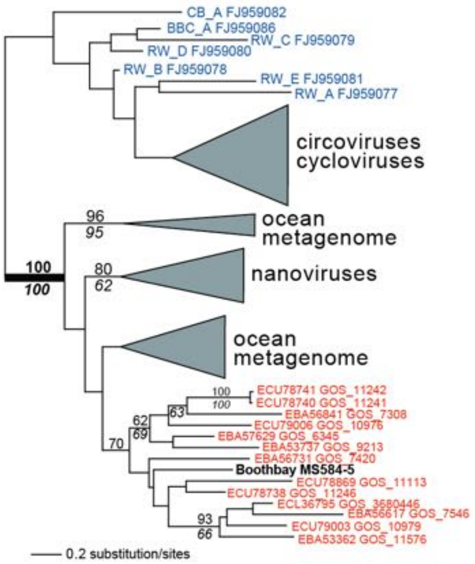


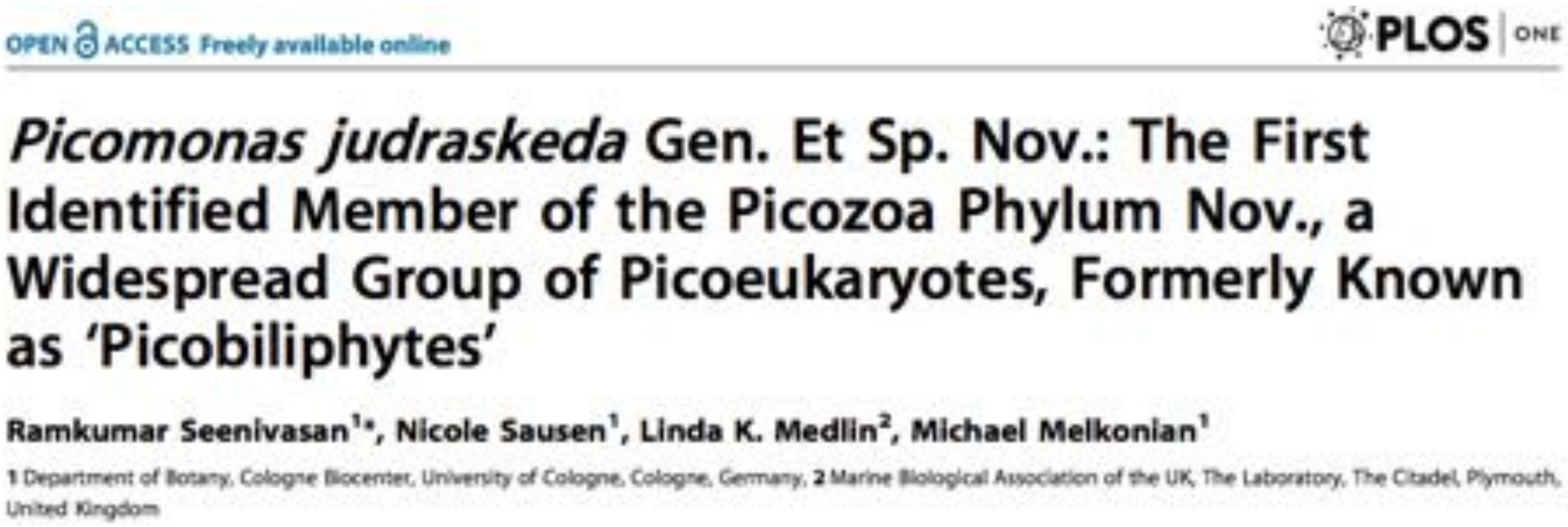
Lead PIs:

HS Yoon (Bigelow) and D Bhattacharya (Rutgers U)

Picozoa-infecting nanovirus







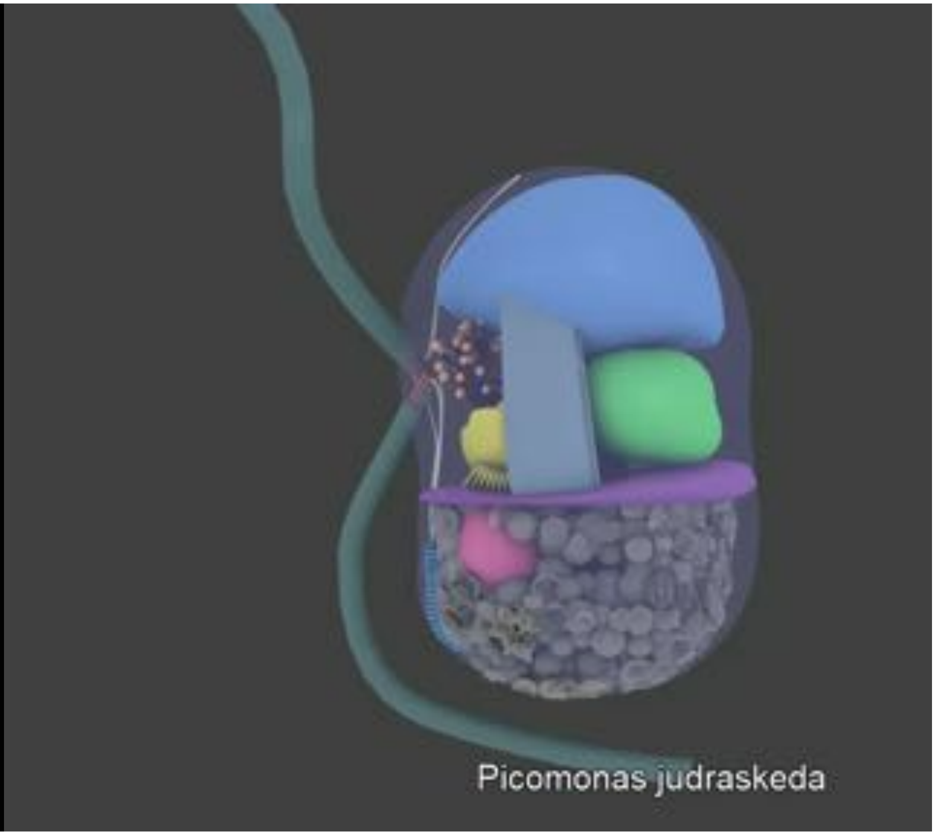
OPEN ACCESS Freely available online

PLOS ONE

Picomonas judraskeda Gen. Et Sp. Nov.: The First Identified Member of the Picozoa Phylum Nov., a Widespread Group of Picoeukaryotes, Formerly Known as 'Picobiliphytes'

Ramkumar Seenivasan[1]*, Nicole Sausen[1], Linda K. Medlin[2], Michael Melkonian[1]

1 Department of Botany, Cologne Biocenter, University of Cologne, Cologne, Germany, 2 Marine Biological Association of the UK, The Laboratory, The Citadel, Plymouth, United Kingdom



Take-home message

- We are still only scratching the surface. Current methods provide only partial and biased information about the genomic composition of microbial communities (and multicellular organisms?)

- Reduced reliance on unverified assumptions is one of the key advantages of single cell genomics, as compared to other techniques

Bigelow Laboratory Single Cell Genomics Center

scgc.*bigelow.org*

SCGC
SINGLE CELL GENOMICS CENTER
Reading nature's genetic tales, one cell at a time

Mission: make single cell genomics accessible to the broad research community; serve as an engine for discoveries in microbial ecology, evolution, bioprospecting and human health.

- **First center of its kind, established 2009**
- **Diverse samples: aquatic, soil, organismal microbiomes, etc.**
- **>1,000,000 cells analyzed, representing >70 phyla**
- **60+ publications since 2011, 8 in *Science*, *Nature* & *PNAS***

Locations of SCGC customers

My lab in 2005

- **Rudimentary knowledge of genomics**
- **No sequencing capacity**
- **”Not ideal” buildings**
- **No resources for major technology development or facility setup**

What got SCGC working?

- **Placing research questions first**
- **Team of motivated, smart people**
- **Tons of advice and collaborations**
- **Method development, continuous since 2005, ~$3 M**
- **Laboratory and IT equipment, ~$3 M**
- **Core facility business model and implementation**
- **QC and benchmarking of entire workflow**
- **Proficiency in and integration of technologies:**
  - Fluorescence-activated cell sorting (FACS)
  - Whole genome amplification
  - Robotic liquid handling
  - Cleanroom environment
  - DNA sequencing
  - *De novo* assembly and other bioinformatics
  - Laboratory information management systems (LIMS)

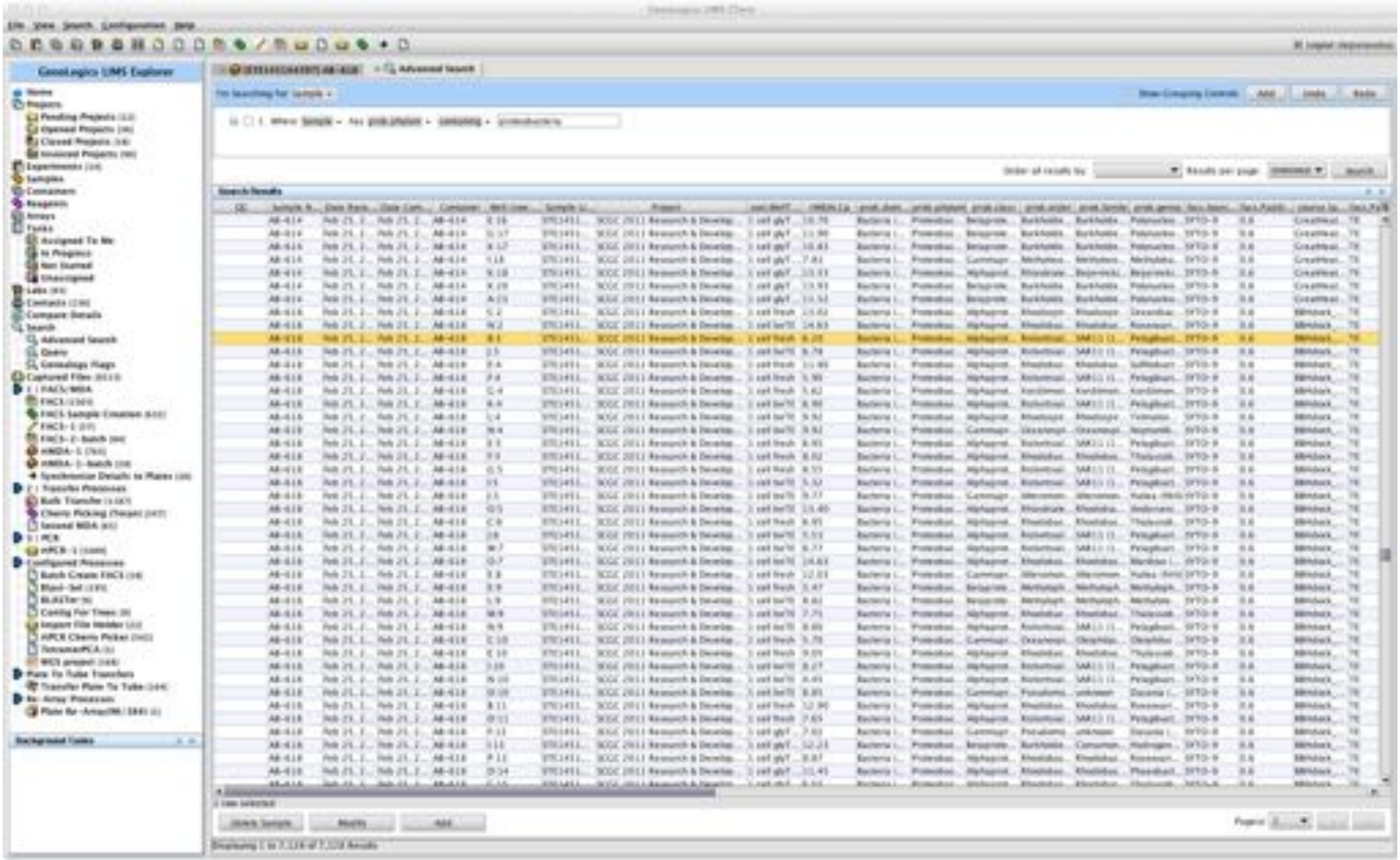
What got SCGC working?

- **Placing research questions first**
- **Motivated, qualified team**
- **A lot of advice and collaborations**
- **Method development, continuous since 2005, ~$3 M**
- **Laboratory and IT equipment, ~$3 M**
- **Core facility business model and implementation**
- **QC and benchmarking of entire workflow**
- **Proficiency in and integration of technologies:**
  - Fluorescence-activated cell sorting (FACS)
  - Whole genome amplification
  - Robotic liquid handling
  - Cleanroom environment
  - DNA sequencing
  - *De novo* assembly and other bioinformatics
  - Laboratory information management systems (LIMS)

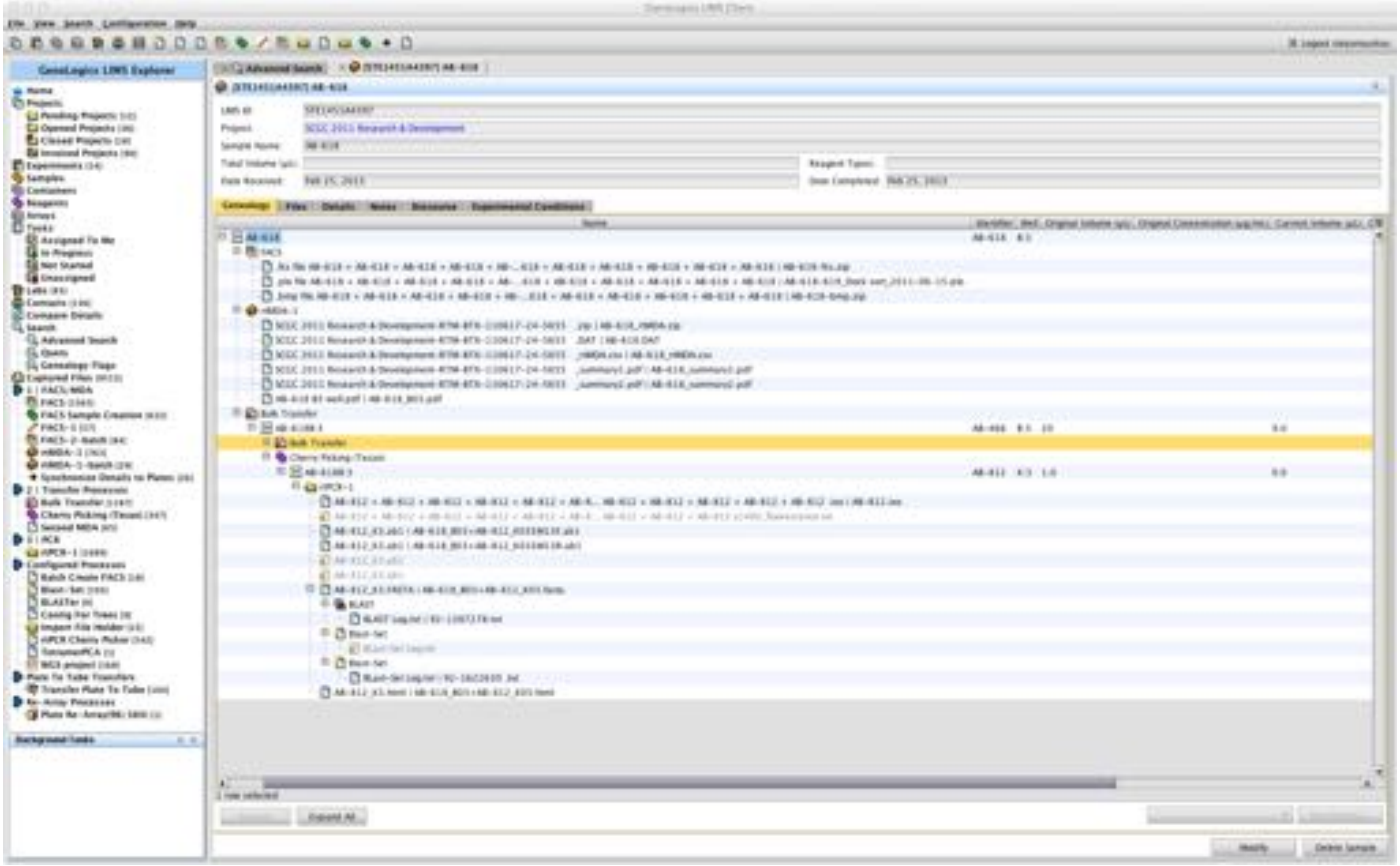
What got SCGC working?

- **Placing research questions first**
- **Motivated, qualified team**
- **A lot of advice and collaborations**
- **Method development, continuous since 2005, ~$3 M**
- **Laboratory and IT equipment, ~$3 M**
- **Core facility business model and implementation**
- **QC and benchmarking of entire workflow**
- **Proficiency in and integration of technologies:**
  - Fluorescence-activated cell sorting (FACS)
  - Whole genome amplification
  - Robotic liquid handling
  - Cleanroom environment
  - DNA sequencing
  - *De novo* assembly and other bioinformatics
  - Laboratory information management systems (LIMS)

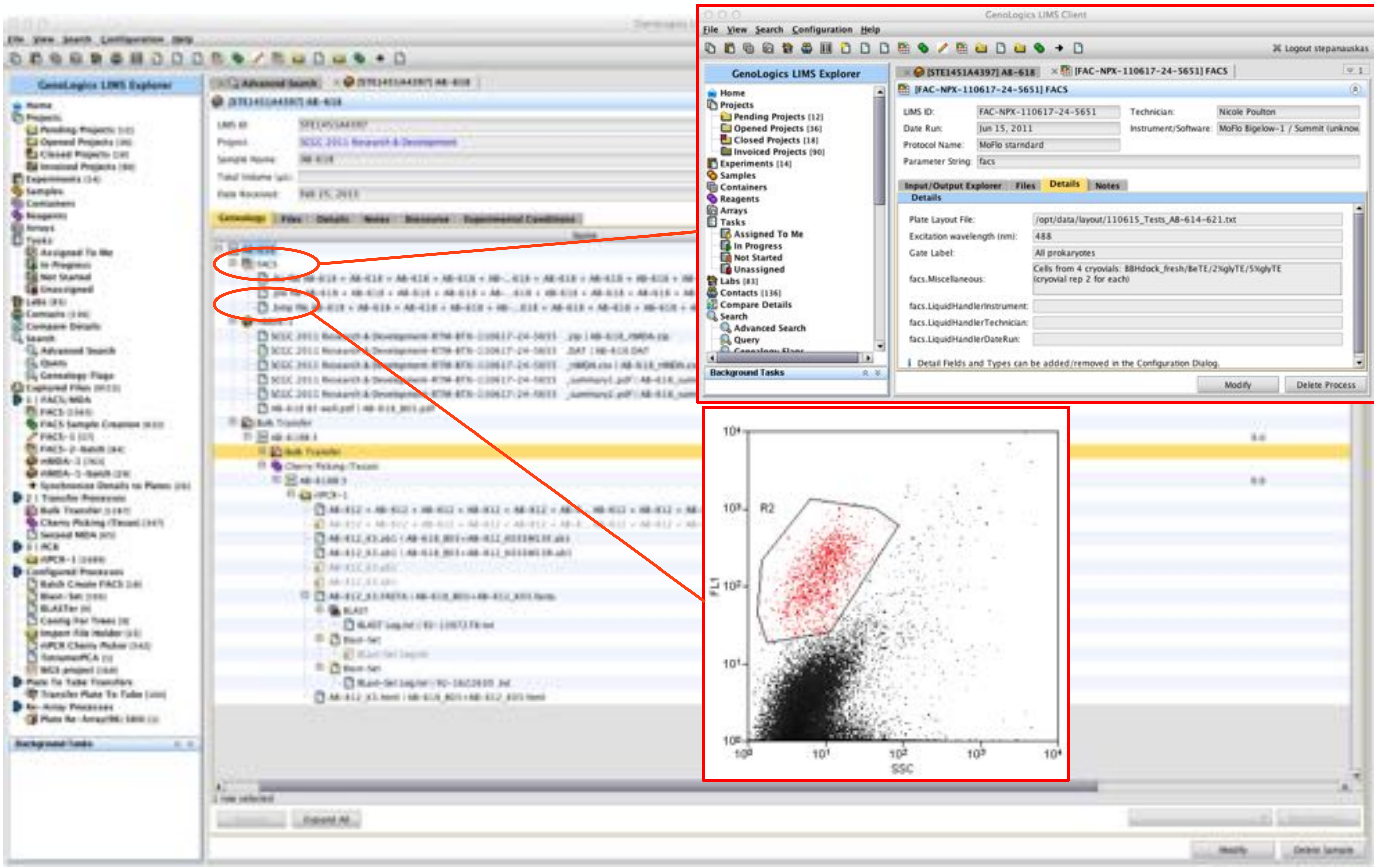
SCGC LIMS – taxonomy search



SCGC LIMS – genealogy of one SAG



SCGC LIMS – FACS data



[FAC-NPX-110617-24-5651] FACS

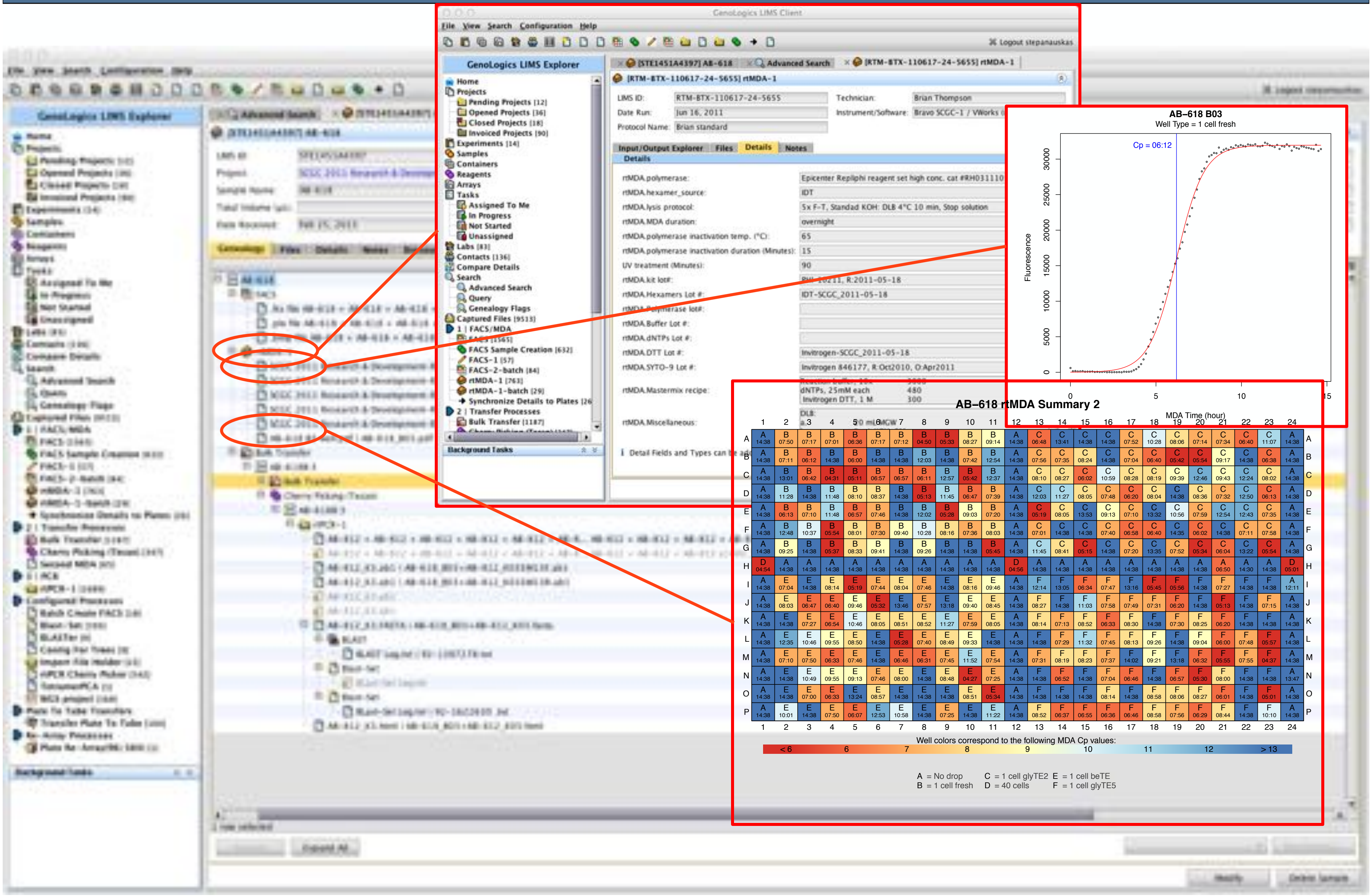
LIMS ID: FAC-NPX-110617-24-5651
Technician: Nicole Poulton
Date Run: Jun 15, 2011
Instrument/Software: MoFlo Bigelow-1 / Summit (unknow
Protocol Name: MoFlo starndard
Parameter String: facs

Details

Plate Layout File: /opt/data/layout/110615_Tests_AB-614-621.txt
Excitation wavelength (nm): 488
Gate Label: All prokaryotes
facs.Miscellaneous: Cells from 4 cryovials: BBHdock_fresh/BeTE/2%glyTE/5%glyTE (cryovial rep 2 for each)
facs.LiquidHandlerInstrument:
facs.LiquidHandlerTechnician:
facs.LiquidHandlerDateRun:

Detail Fields and Types can be added/removed in the Configuration Dialog.

SCGC LIMS – MDA data



[RTM-BTX-110617-24-5655] rtMDA-1

LIMS ID: RTM-BTX-110617-24-5655
Date Run: Jun 16, 2011
Protocol Name: Brian standard
Technician: Brian Thompson
Instrument/Software: Bravo SCGC-1 / VWorks

Details

rtMDA.polymerase: Epicenter Repliphi reagent set high conc. cat #RH031110
rtMDA.hexamer_source: IDT
rtMDA.lysis protocol: 5x F-T, Standad KOH: DLB 4°C 10 min, Stop solution
rtMDA.MDA duration: overnight
rtMDA.polymerase inactivation temp. (°C): 65
rtMDA.polymerase inactivation duration (Minutes): 15
UV treatment (Minutes): 90
rtMDA.kit lot#: RH-10211, R:2011-05-18
rtMDA.Hexamers Lot #: IDT-SCGC_2011-05-18
rtMDA.Polymerase lot#:
rtMDA.Buffer Lot #:
rtMDA.dNTPs Lot #:
rtMDA.DTT Lot #: Invitrogen-SCGC_2011-05-18
rtMDA.SYTO-9 Lot #: Invitrogen 846177, R:Oct2010, O:Apr2011
rtMDA.Mastermix recipe: Reaction buffer, 10x 3000; dNTPs, 25mM each 480; Invitrogen DTT, 1 M 300
rtMDA.Miscellaneous: DLB:

Detail Fields and Types can be ad

AB–618 rtMDA Summary 2

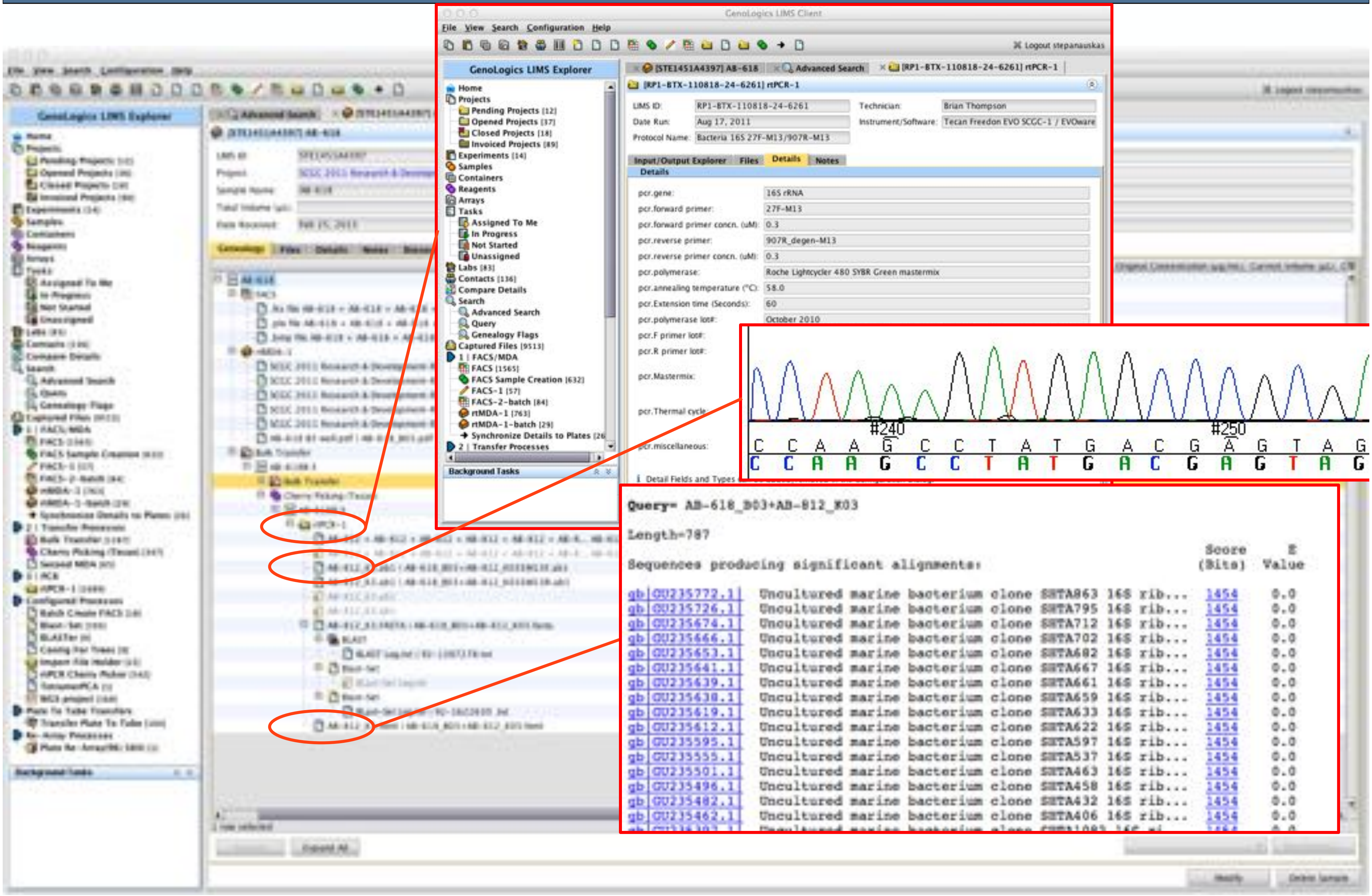
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	
A	A 14:38	B 07:50	B 07:17	B 07:01	B 06:36	B 07:17	B 07:12	B 04:50	B 05:33	B 08:27	B 09:14	A 14:38	C 06:48	C 13:41	C 14:38	C 14:38	C 07:52	C 10:28	C 08:06	C 07:14	C 07:34	C 06:40	C 11:07	A 14:38	A
B	A 14:38	B 07:11	B 06:12	B 14:38	B 06:00	B 14:38	B 14:38	B 12:03	B 14:38	B 07:42	B 12:54	A 14:38	C 07:56	C 07:35	C 08:24	C 14:38	C 07:04	C 06:40	C 05:42	C 05:54	C 09:17	C 14:38	C 06:38	A 14:38	B
C	A 14:38	B 13:01	B 06:42	B 04:31	B 05:11	B 06:57	B 06:57	B 06:11	B 12:57	B 05:42	B 12:37	A 14:38	C 08:10	C 08:27	C 06:02	C 10:59	C 08:28	C 08:19	C 09:39	C 12:46	C 09:43	C 12:24	C 08:02	A 14:38	C
D	A 14:38	B 11:28	B 14:38	B 11:48	B 08:10	B 08:37	B 14:38	B 05:13	B 11:45	B 06:47	B 07:39	A 14:38	C 12:03	C 11:27	C 08:05	C 07:48	C 06:20	C 08:04	C 14:38	C 08:36	C 07:32	C 12:50	C 06:13	A 14:38	D
E	A 14:38	B 06:13	B 07:10	B 11:48	B 06:57	B 07:46	B 14:38	B 12:02	B 05:28	B 09:03	B 07:20	A 14:38	C 05:19	C 08:05	C 13:53	C 09:13	C 07:10	C 13:32	C 10:56	C 07:59	C 12:54	C 12:43	C 07:35	A 14:38	E
F	A 14:38	B 12:48	B 10:37	B 05:54	B 08:01	B 07:30	B 09:40	B 10:28	B 08:16	B 07:36	B 08:03	A 14:38	C 07:01	C 14:38	C 14:38	C 07:40	C 06:58	C 06:40	C 14:35	C 06:02	C 14:38	C 07:11	C 07:58	A 14:38	F
G	A 14:38	B 09:25	B 14:38	B 05:37	B 08:33	B 09:41	B 14:38	B 09:26	B 14:38	B 14:38	B 05:45	A 14:38	C 11:45	C 08:41	C 05:15	C 14:38	C 07:20	C 13:35	C 07:52	C 05:34	C 06:04	C 13:22	C 05:54	A 14:38	G
H	D 04:54	A 14:38	A 14:38	A 14:38	A 14:38	A 14:38	A 14:38	A 14:38	A 14:38	A 14:38	A 14:38	D 04:56	A 14:38	A 14:38	A 14:38	A 14:38	A 14:38	A 14:38	A 14:38	A 14:38	A 06:50	A 14:30	A 14:38	D 05:01	H
I	A 14:38	E 07:04	E 14:38	E 08:14	E 05:19	E 07:44	E 08:04	E 07:46	E 14:38	E 08:16	E 09:46	A 14:38	F 12:14	F 13:05	F 06:34	F 07:47	F 13:16	F 05:45	F 05:56	F 14:38	F 07:27	F 14:38	F 14:38	A 12:11	I
J	A 14:38	E 08:03	E 06:47	E 06:40	E 09:46	E 05:32	E 13:46	E 07:57	E 13:18	E 09:40	E 08:45	A 14:38	F 08:27	F 14:38	F 11:03	F 07:58	F 07:49	F 07:31	F 06:20	F 14:38	F 05:13	F 14:38	F 07:15	A 14:38	J
K	A 14:38	E 14:38	E 07:27	E 06:54	E 10:46	E 08:05	E 08:51	E 08:52	E 11:27	E 07:59	E 08:05	A 14:38	F 08:14	F 07:13	F 08:52	F 06:33	F 08:30	F 14:38	F 07:30	F 08:25	F 06:20	F 14:38	F 14:38	A 14:38	K
L	A 14:38	E 12:35	E 10:46	E 09:55	E 08:50	E 14:38	E 05:28	E 07:40	E 08:49	E 09:33	E 14:38	A 14:38	F 14:38	F 07:29	F 11:32	F 07:45	F 08:13	F 09:26	F 14:38	F 09:04	F 06:00	F 07:48	F 05:57	A 14:38	L
M	A 14:38	E 07:10	E 07:50	E 06:33	E 07:46	E 14:38	E 06:46	E 06:31	E 07:45	E 11:52	E 07:54	A 14:38	F 07:31	F 08:19	F 08:23	F 07:37	F 14:02	F 09:21	F 13:18	F 06:32	F 05:55	F 07:55	F 04:37	A 14:38	M
N	A 14:38	E 14:38	E 10:49	E 09:55	E 09:13	E 07:46	E 08:00	E 14:38	E 08:48	E 04:27	E 07:25	A 14:38	F 14:38	F 06:52	F 14:38	F 07:04	F 06:46	F 14:38	F 06:57	F 05:30	F 08:00	F 14:38	F 14:38	A 13:47	N
O	A 14:38	E 14:38	E 07:00	E 06:33	E 13:24	E 08:57	E 14:38	E 14:38	E 14:38	E 08:51	E 05:34	A 14:38	F 14:38	F 14:38	F 14:38	F 08:14	F 14:38	F 08:58	F 08:06	F 08:27	F 06:01	F 14:38	F 05:01	A 14:38	O
P	A 14:38	E 10:01	E 14:38	E 07:50	E 06:07	E 12:53	E 10:58	E 14:38	E 07:25	E 14:38	E 11:22	A 14:38	F 08:52	F 06:37	F 06:55	F 06:36	F 06:46	F 08:58	F 07:56	F 06:29	F 08:44	F 14:38	F 10:10	A 14:38	P
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	

Well colors correspond to the following MDA Cp values:

< 6 · 6 · 7 · 8 · 9 · 10 · 11 · 12 · > 13

A = No drop
B = 1 cell fresh
C = 1 cell glyTE2
D = 40 cells
E = 1 cell beTE
F = 1 cell glyTE5

SCGC LIMS – SSU rRNA PCR-sequencing data



[RP1-BTX-110818-24-6261] rtPCR-1

LIMS ID: RP1-BTX-110818-24-6261

Technician: Brian Thompson

Date Run: Aug 17, 2011

Instrument/Software: Tecan Freedon EVO SCGC-1 / EVOware

Protocol Name: Bacteria 16S 27F-M13/907R-M13

Details

pcr.gene: 16S rRNA

pcr.forward primer: 27F-M13

pcr.forward primer concn. (uM): 0.3

pcr.reverse primer: 907R_degen-M13

pcr.reverse primer concn. (uM): 0.3

pcr.polymerase: Roche Lightcycler 480 SYBR Green mastermix

pcr.annealing temperature (°C): 58.0

pcr.Extension time (Seconds): 60

pcr.polymerase lot#: October 2010

pcr.F primer lot#:

pcr.R primer lot#:

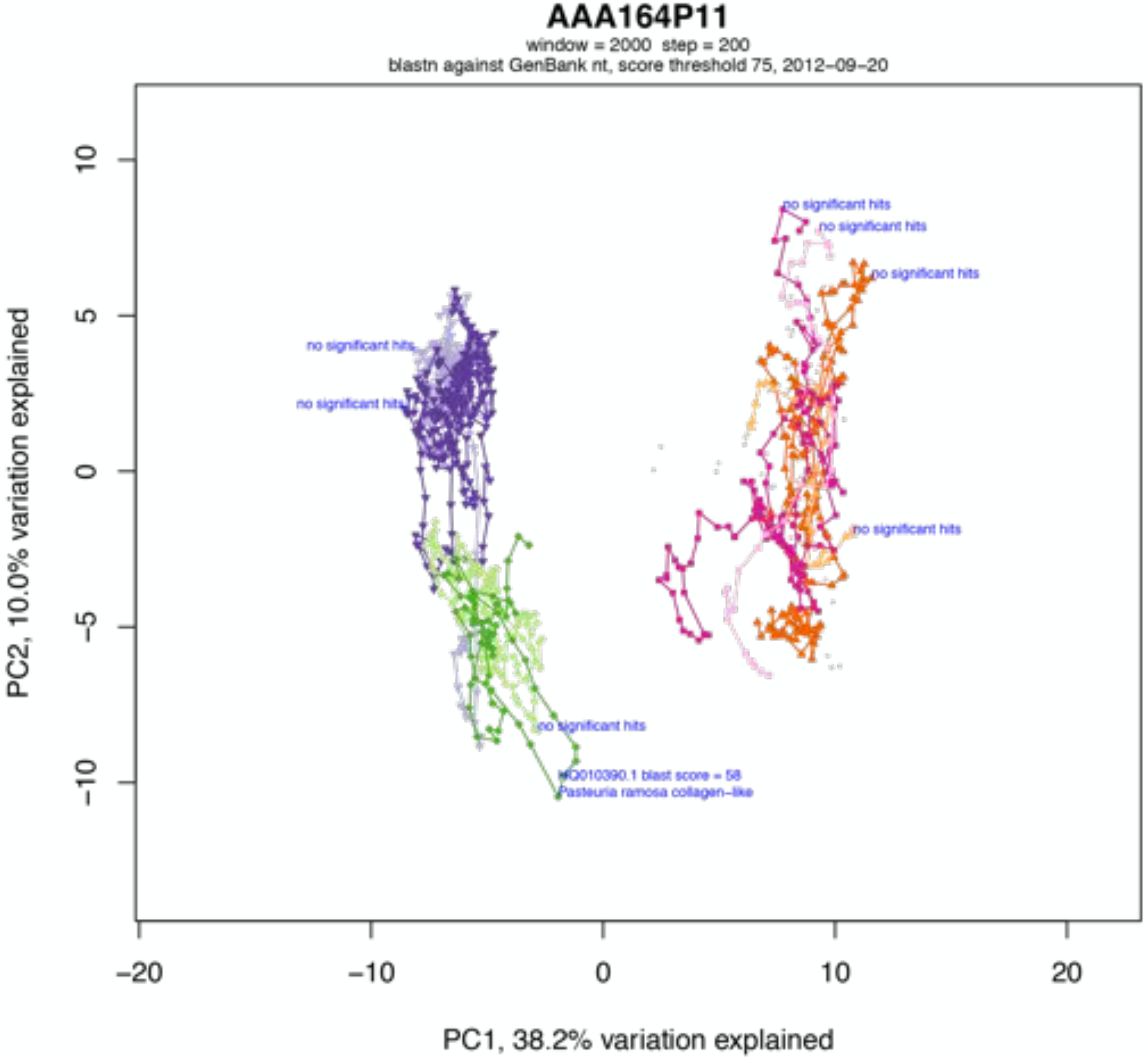
pcr.Mastermix:

pcr.Thermal cycle:

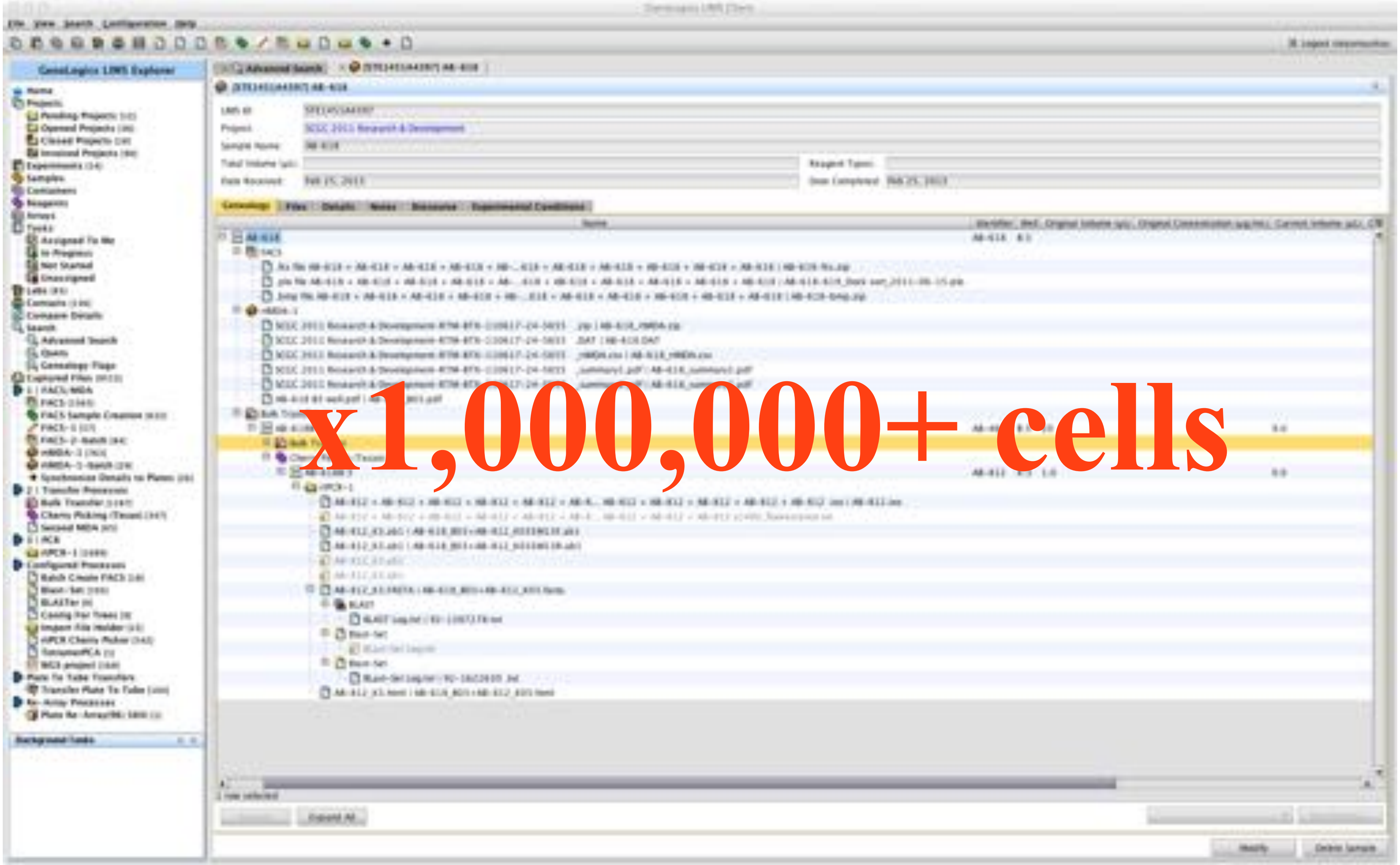
pcr.miscellaneous:

```
Query= AB-618_D03+AB-812_E03

Length=787
                                                                      Score     E
Sequences producing significant alignments:                          (Bits)  Value

gb|GU235772.1|  Uncultured marine bacterium clone SHTA863 16S rib...  1454    0.0
gb|GU235726.1|  Uncultured marine bacterium clone SHTA795 16S rib...  1454    0.0
gb|GU235674.1|  Uncultured marine bacterium clone SHTA712 16S rib...  1454    0.0
gb|GU235666.1|  Uncultured marine bacterium clone SHTA702 16S rib...  1454    0.0
gb|GU235653.1|  Uncultured marine bacterium clone SHTA682 16S rib...  1454    0.0
gb|GU235641.1|  Uncultured marine bacterium clone SHTA667 16S rib...  1454    0.0
gb|GU235639.1|  Uncultured marine bacterium clone SHTA661 16S rib...  1454    0.0
gb|GU235638.1|  Uncultured marine bacterium clone SHTA659 16S rib...  1454    0.0
gb|GU235619.1|  Uncultured marine bacterium clone SHTA633 16S rib...  1454    0.0
gb|GU235612.1|  Uncultured marine bacterium clone SHTA622 16S rib...  1454    0.0
gb|GU235595.1|  Uncultured marine bacterium clone SHTA597 16S rib...  1454    0.0
gb|GU235555.1|  Uncultured marine bacterium clone SHTA537 16S rib...  1454    0.0
gb|GU235501.1|  Uncultured marine bacterium clone SHTA463 16S rib...  1454    0.0
gb|GU235496.1|  Uncultured marine bacterium clone SHTA458 16S rib...  1454    0.0
gb|GU235482.1|  Uncultured marine bacterium clone SHTA432 16S rib...  1454    0.0
gb|GU235462.1|  Uncultured marine bacterium clone SHTA406 16S rib...  1454    0.0
```

SCGC QC: Tetramer PCA



SCGC LIMS



x1,000,000+ cells

What got SCGC working?

- **Placing research questions first**
- **Motivated, qualified team**
- **A lot of advice and collaborations**
- **Method development, continuous since 2005, ~$3 M**
- **Laboratory and IT equipment, ~$3 M**
- **Core facility business model and implementation**
- **QC and benchmarking of entire workflow**
- **Proficiency in and integration of technologies:**
  - Fluorescence-activated cell sorting (FACS)
  - Whole genome amplification
  - Robotic liquid handling
  - Cleanroom environment
  - DNA sequencing
  - *De novo* assembly and other bioinformatics
  - Laboratory information management systems (LIMS)

Benchmarking SAG assemblies (QUAST, 5 replicates)

Genome characteristics

Organism	Genome size, bp	GC, %
Prochlorooccus marinus	1,751,080	37
Escherichia coli	4,630,707	51
Meiothermus ruber	3,097,457	63

Standard QC metrics

Genome recovery, %	N50	# N's per 100 kbp
78	54,686	0
52	21,687	0
50	15,097	0

Assessment of assembly accuracy

Organism	# unaligned bases	# misassemblies	# local misassemblies	# mismatches per 100 kbp	# indels per 100 kbp
Prochlorooccus marinus	0	2	3	4	1
Escherichia coli	68	16	5	5	1
Meiothermus ruber	28	20	3	5	3

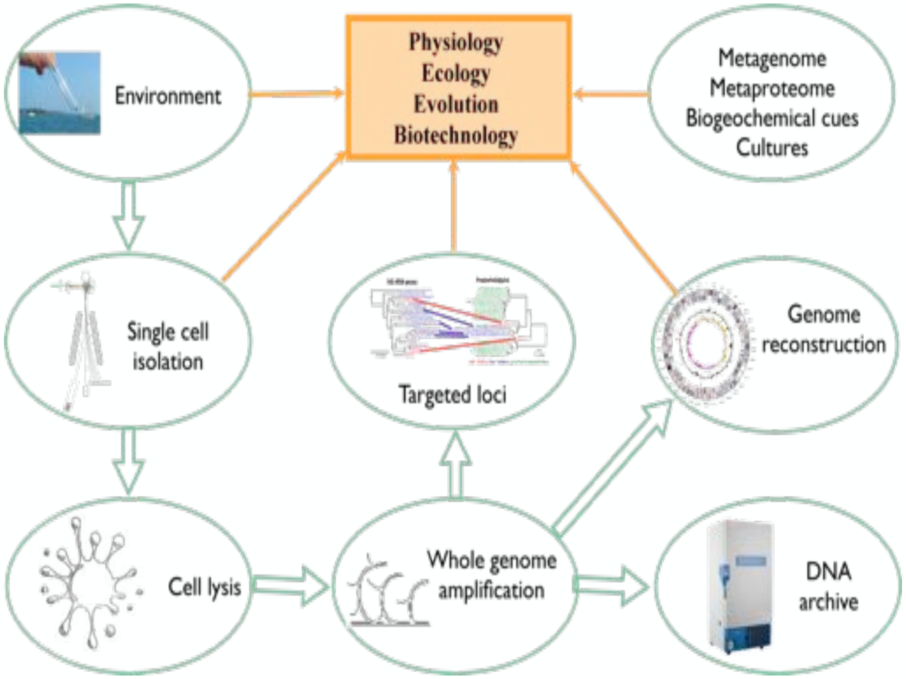
Assumptions:

1. Original assembly is correct.
2. No genetic changes since

Benefits:

1. Accurate interpretation of results
2. Continued method improvements

Factors that may impair SAG genome recovery



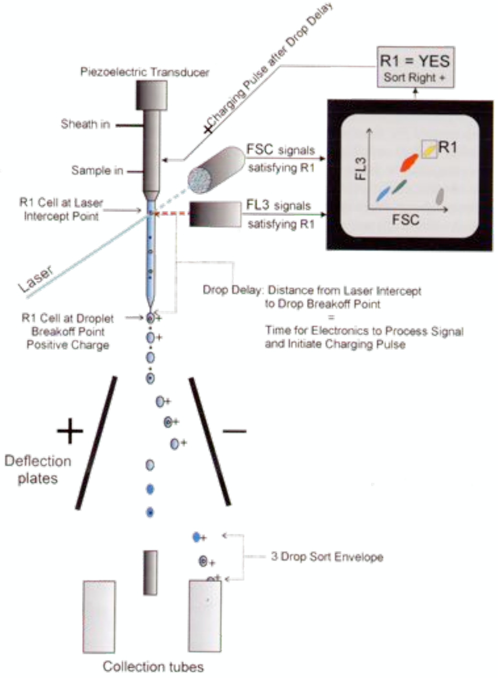
Technical:

- DNA degradation during storage/shipping
- Failure to deposit a cell into a well
- Failure to lyse the cell
- Failure to denature DNA
- Uneven WGA
- Sequencing artifacts
- Assembly artifacts
- Pipetting error at any lab step
- Computational error

Biological:

- DNA leaked from a mechanically damaged cell
- Cell is in a dormant state
- Host DNA was degraded by a lytic phage
- Polymerase was hijacked by small, circular plasmids
- DNA was fragmented, e.g. due to desiccation
- DNA was bound to proteins and/or other molecules, e.g. due to desiccation
- DNA was protected by intracellular compartmentalization

SCGC QC: Cell sorting



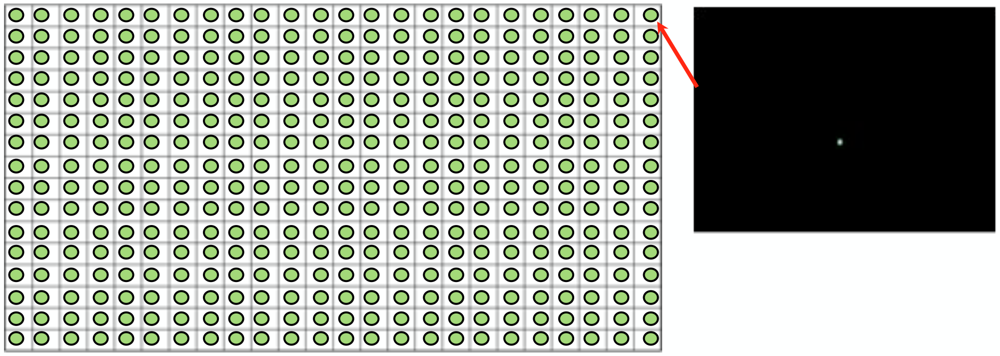
Contamination prevention:

- HEPA-filtered air, cleanroom techniques
- Decontamination of all reagents
- Single-drop sort mode
- Careful drop delay and sort alignment
- Negative controls on each sort plate

Small volumes is the key!

Pressure differential	Ratio of V_s to V_d (R)	Sample stream diameter (µm)	Sample droplet volume, V_s (pL)	ESD of V_s (µm)
0.2	3.13×10^{-3}	5.6	8.7	25.9
0.4	4.98×10^{-3}	7.0	13.9	30.3
0.6	7.93×10^{-3}	8.9	22.1	35.3
0.8	9.71×10^{-3}	9.8	27.0	37.8

Microscopy of sorted fluorescent beads

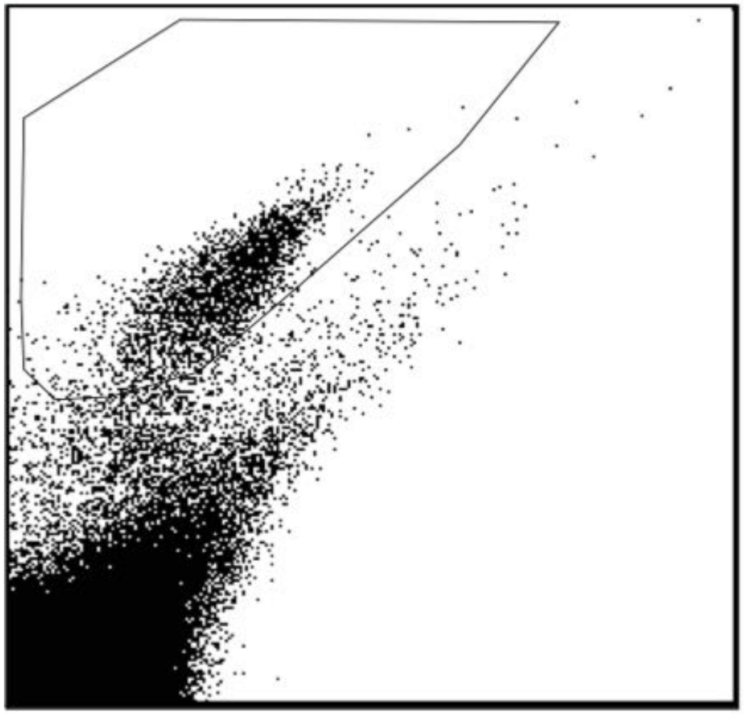


<2% wells contain no bead

<0.4% wells contain more than one bead

Preventing particle co-sort

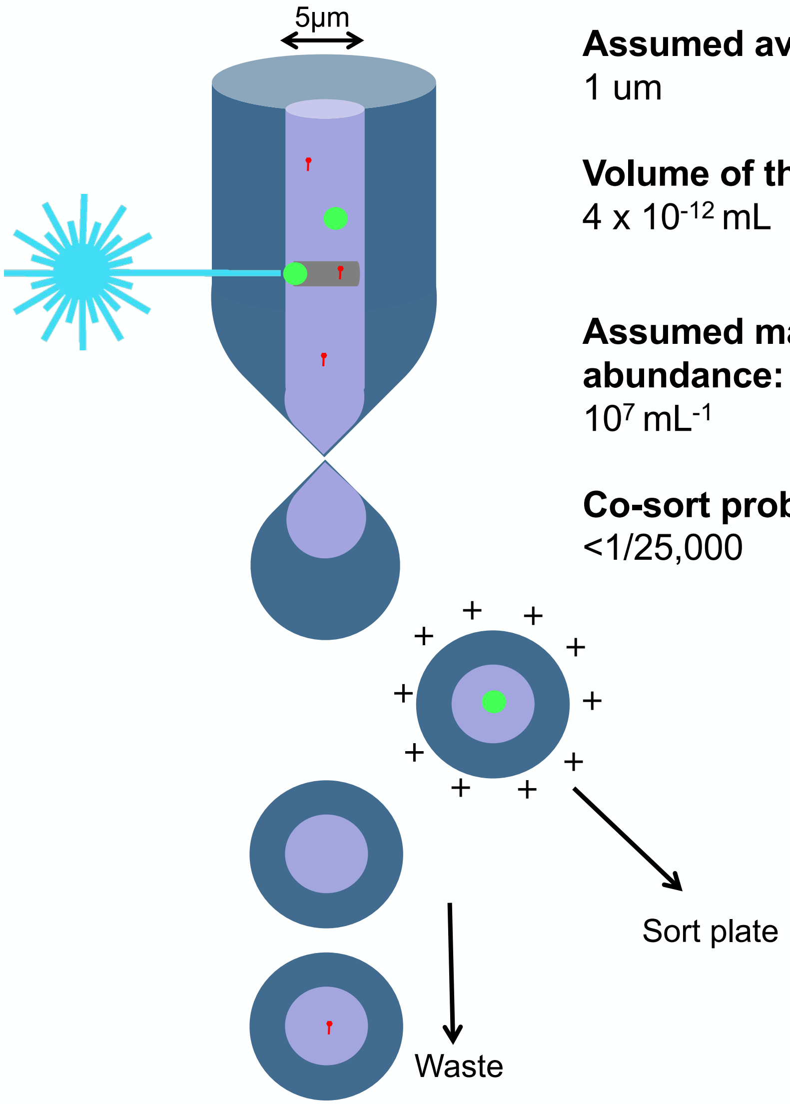
Maximize detection sensitivity



Generate small droplets

Pressure differential	Sample stream diameter (μm)	Sample droplet volume, V_s (pL)
0.2	5.6	8.7
0.4	7.0	13.9
0.6	8.9	22.1
0.8	9.8	27.0

Use single-drop FACS mode



Assumed average diameter:
1 um

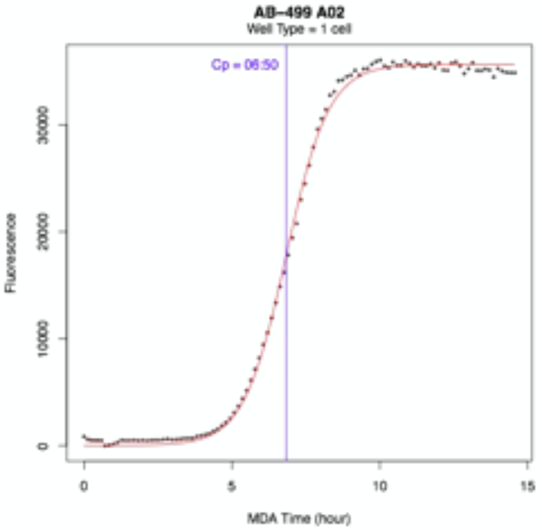
Volume of the shadow:
4 x 10^{-12} mL

Assumed max particle abundance:
10^7 mL^{-1}

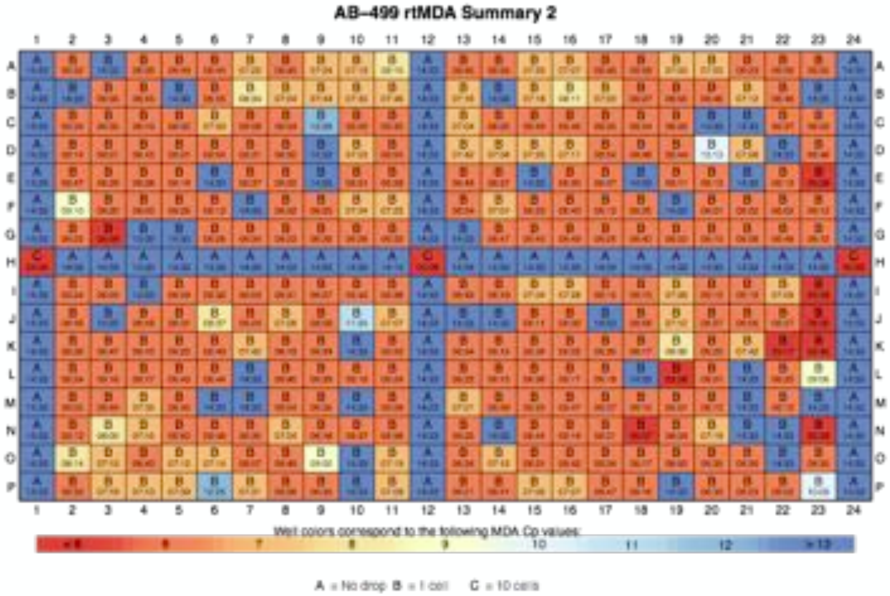
Co-sort probability:
<1/25,000

Whole genome amplification kinetics

Individual reaction kinetics



Real data: *Tetraselmis* culture

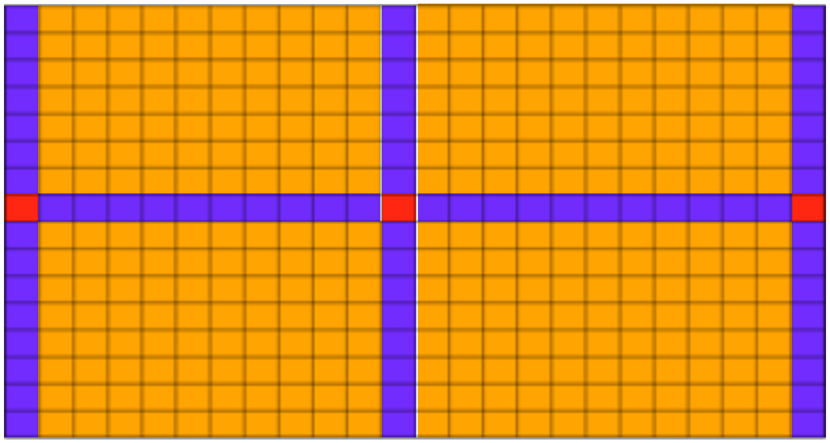


384-well plate layout

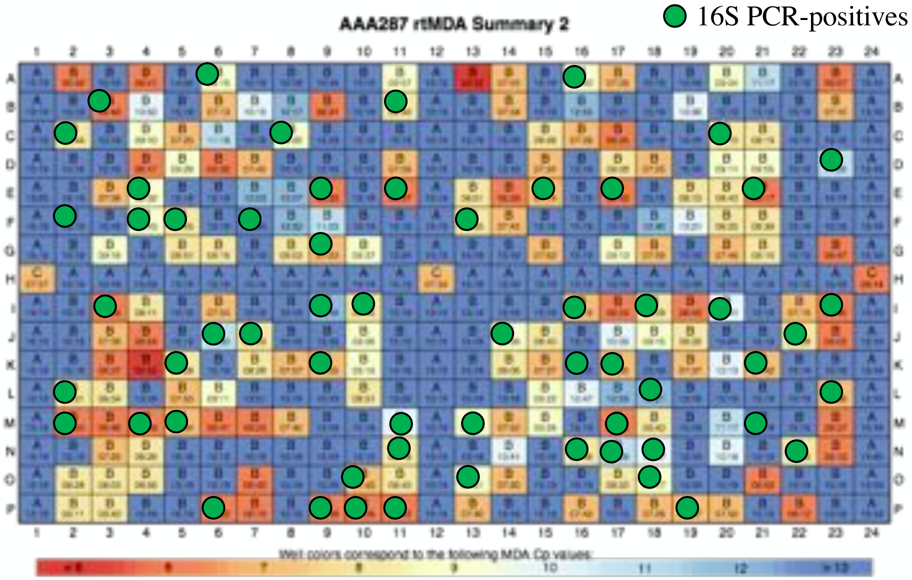
Negative controls: empty wells

Positive controls: 10 cells per well

Wells containing single cells

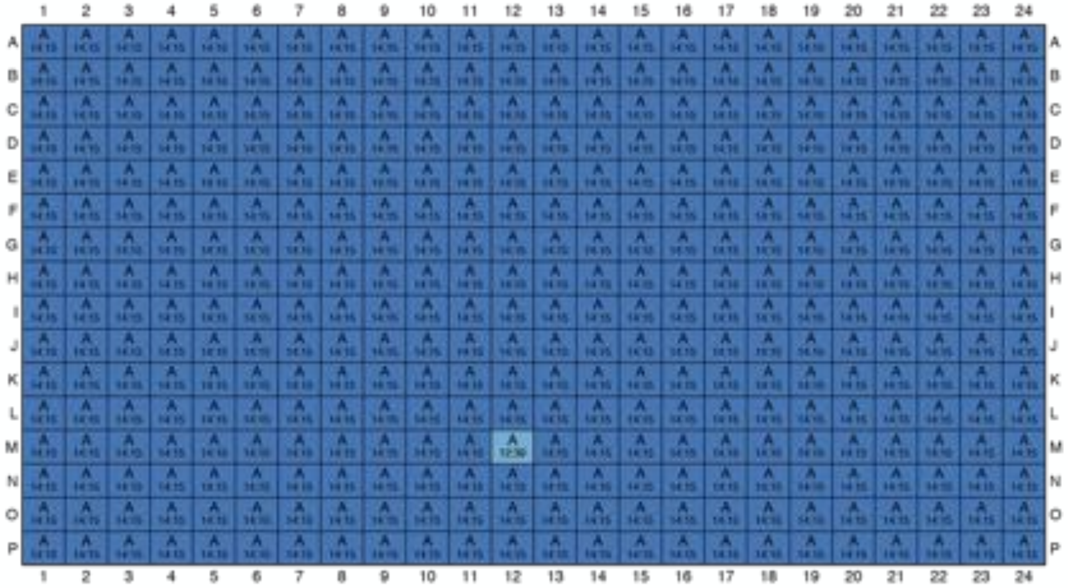


Real data: marine prokaryotes from 3000 m. depth

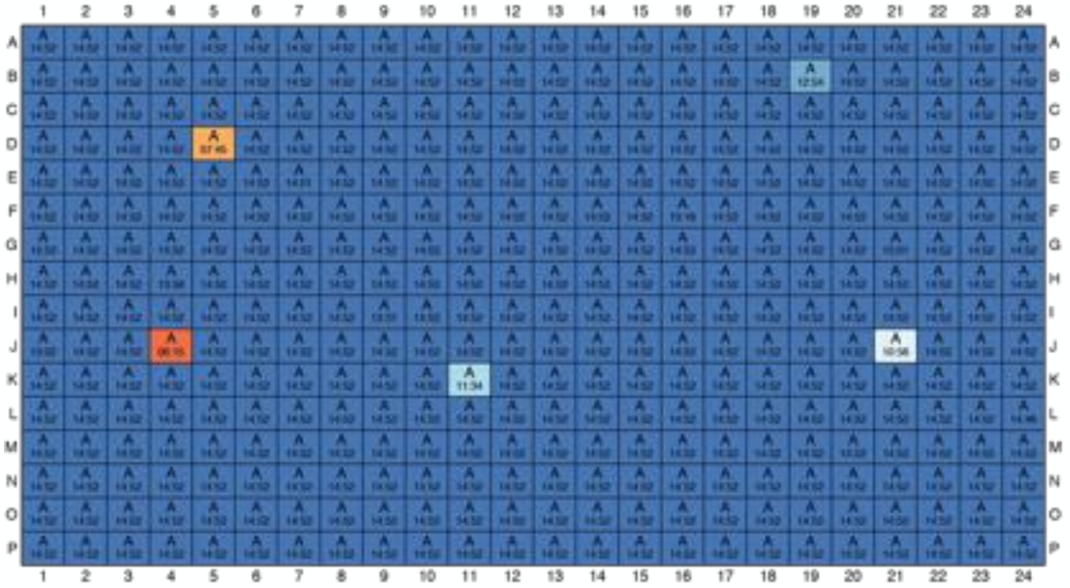


QC of lab consumables

Clean batch of robotic pipetting tips



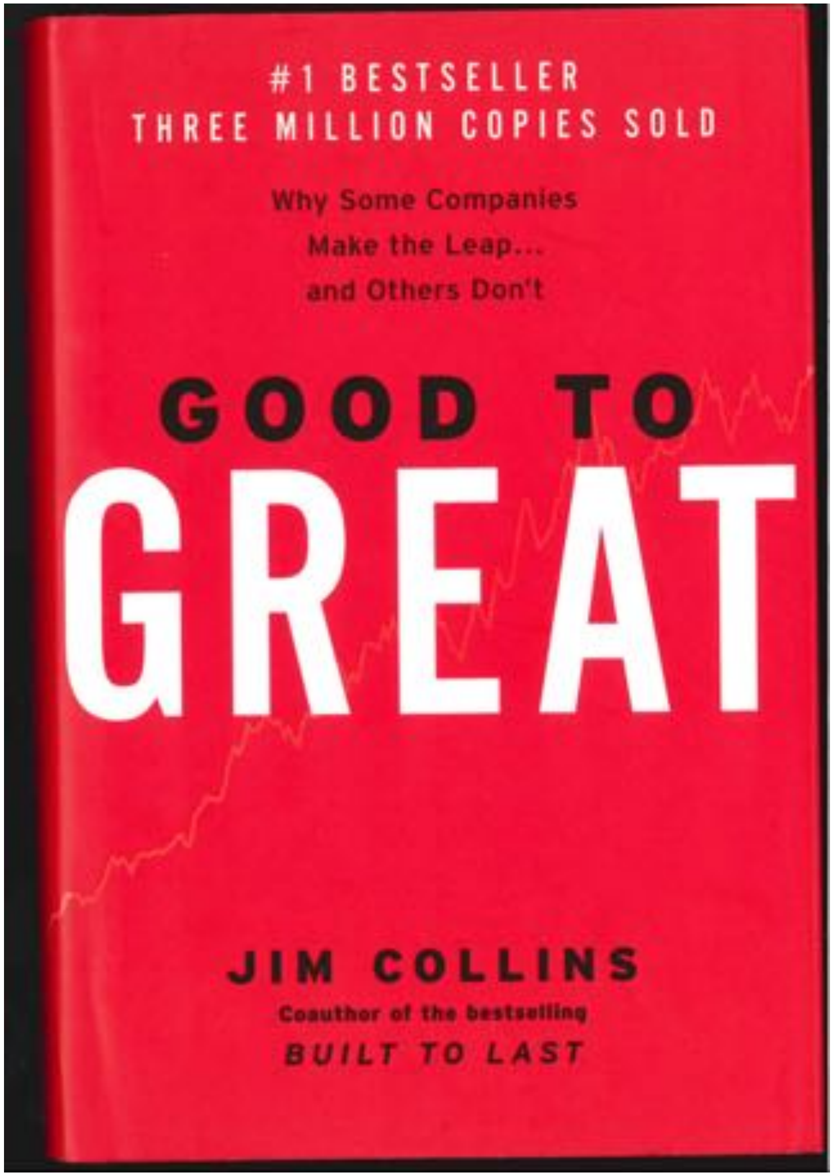
DNA-contaminated batch of tips

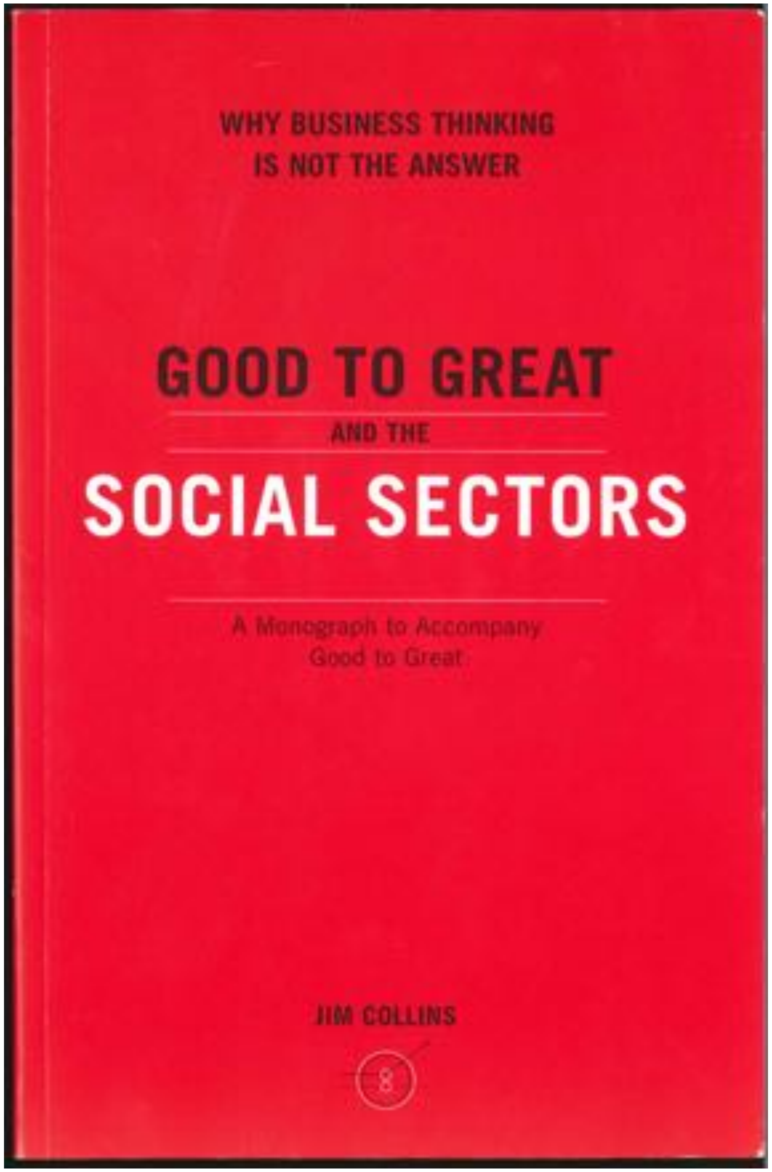


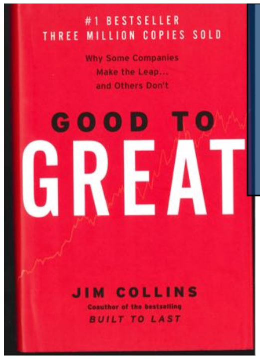
What got SCGC working?

- **Placing research questions first**
- **Motivated, qualified team**
- **A lot of advice and collaborations**
- **Method development, continuous since 2005, ~$3 M**
- **Laboratory and IT equipment, ~$3 M**
- **Core facility business model and implementation**
- **QC and benchmarking of entire workflow**
- **Proficiency in and integration of technologies:**
  - Fluorescence-activated cell sorting (FACS)
  - Whole genome amplification
  - Robotic liquid handling
  - Cleanroom environment
  - DNA sequencing
  - *De novo* assembly and other bioinformatics
  - Laboratory information management systems (LIMS)

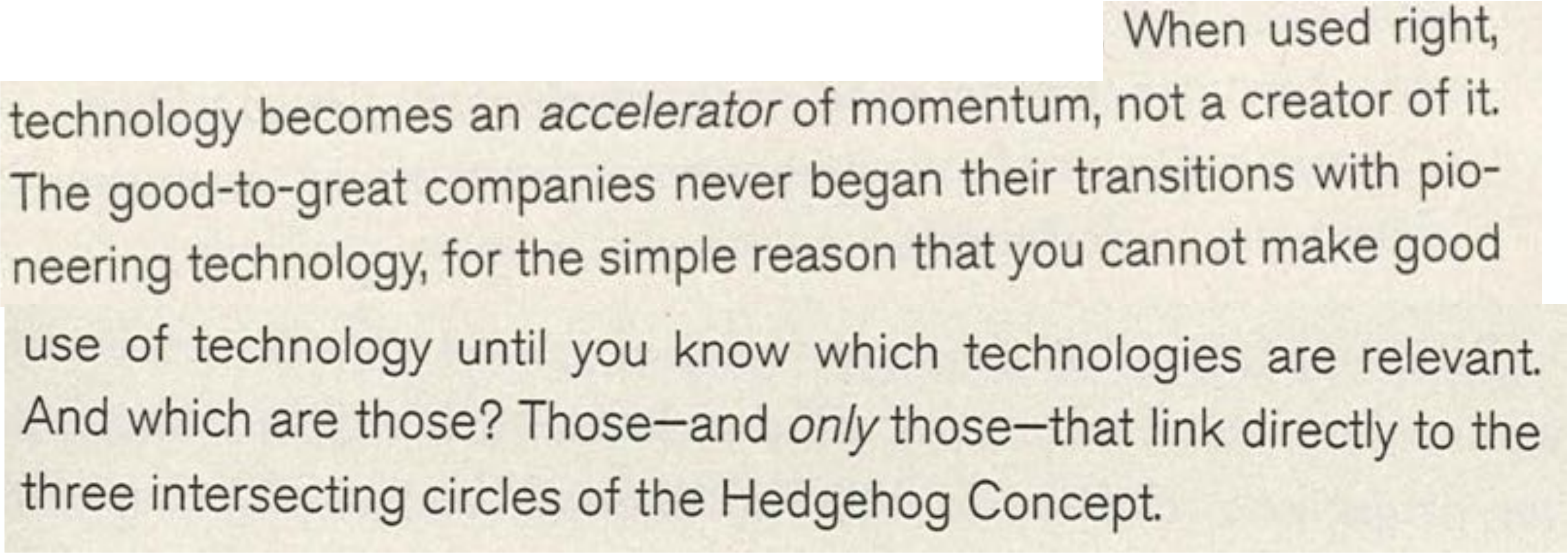
Excellent primers in management



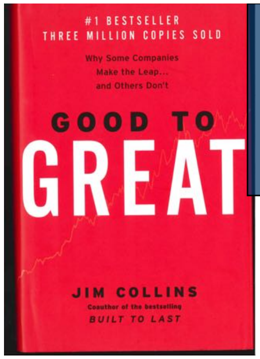




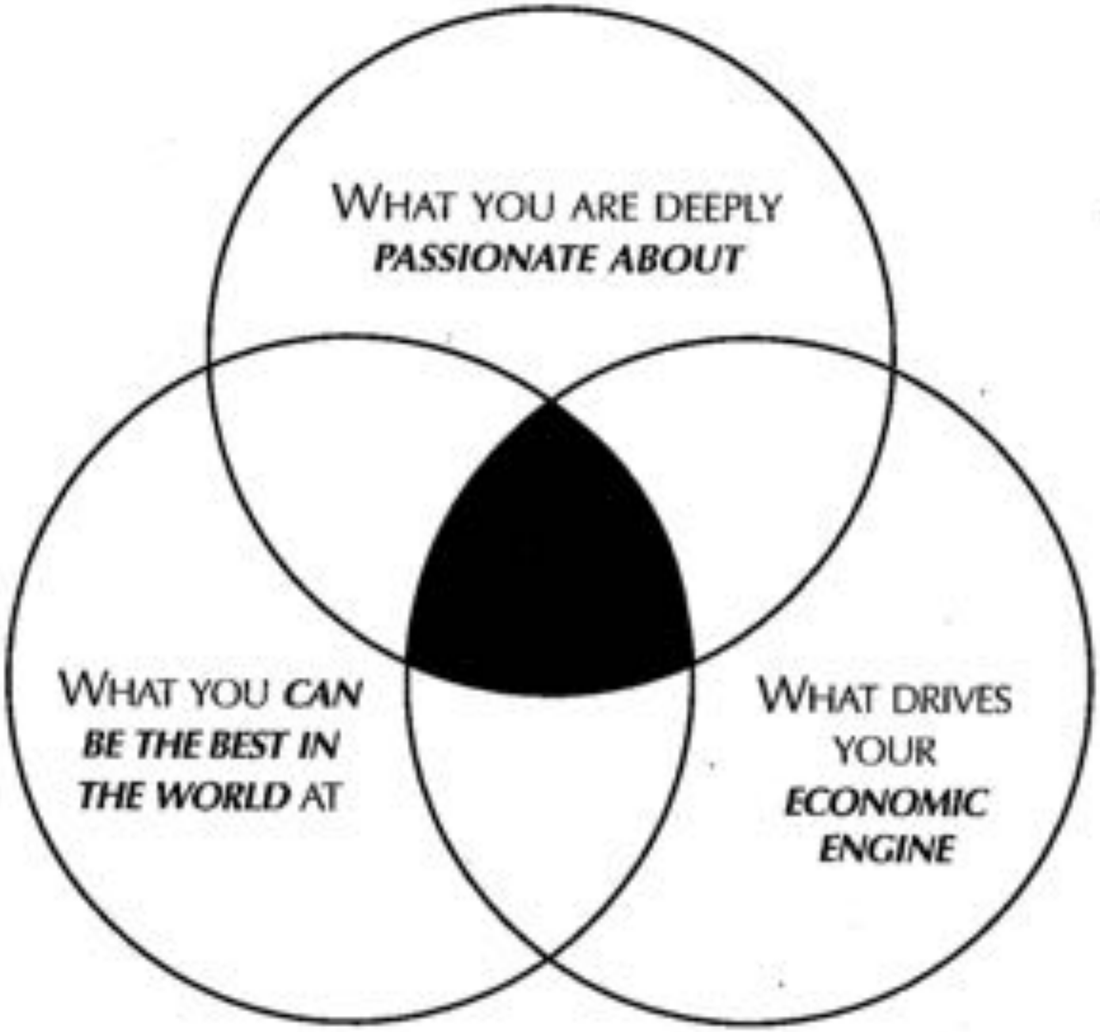
Technology does not come first



When used right, technology becomes an *accelerator* of momentum, not a creator of it. The good-to-great companies never began their transitions with pioneering technology, for the simple reason that you cannot make good use of technology until you know which technologies are relevant. And which are those? Those—and *only* those—that link directly to the three intersecting circles of the Hedgehog Concept.



The Hedgehog Concept

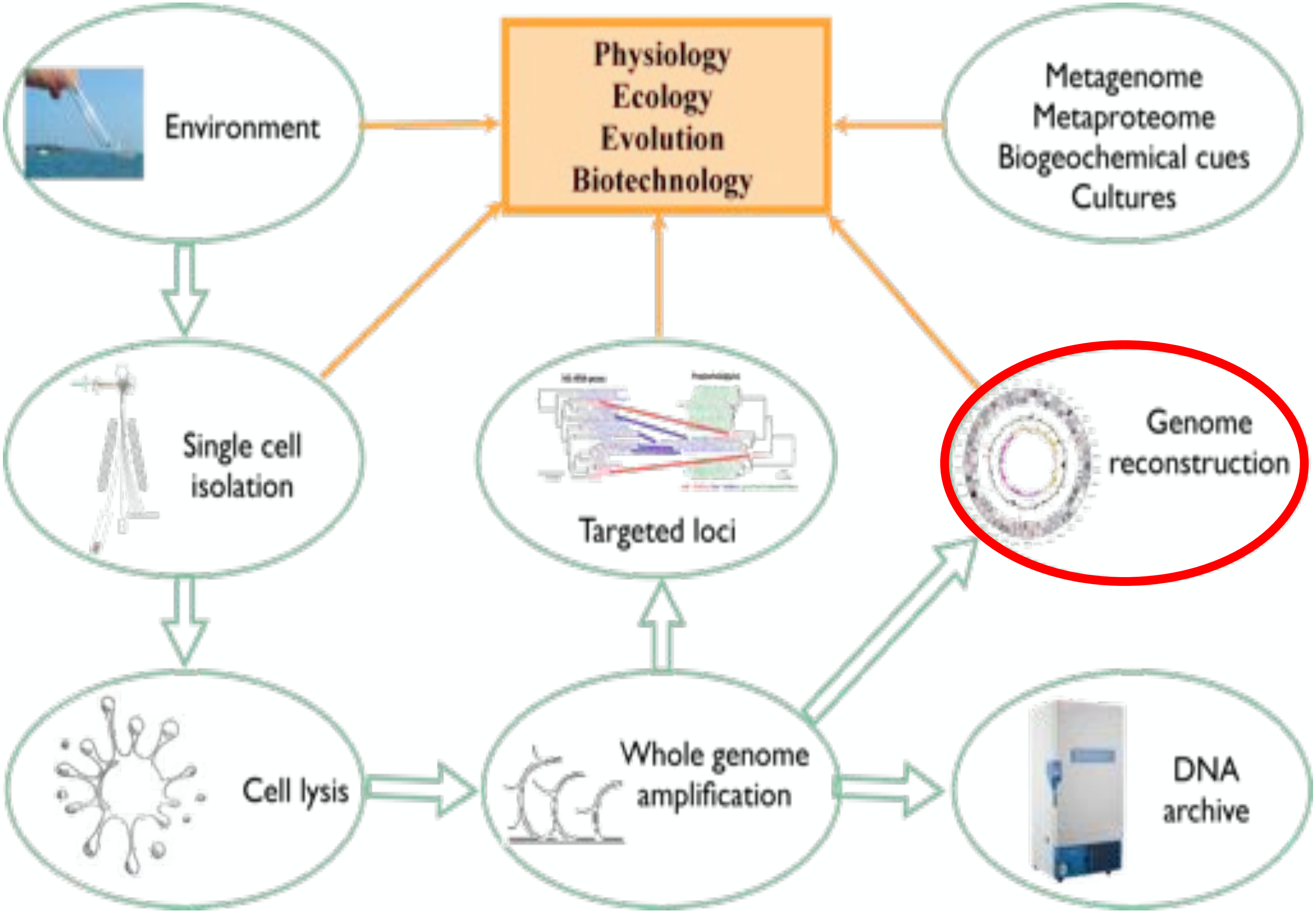


THREE CIRCLES OF THE HEDGEHOG CONCEPT

What got SCGC working?

- **Placing research questions first**
- **Motivated, qualified team**
- **A lot of advice and collaborations**
- **Method development, continuous since 2005, ~$3 M**
- **Laboratory and IT equipment, ~$3 M**
- **Core facility business model and implementation**
- **QC and benchmarking of entire workflow**
- **Proficiency in and integration of technologies:**
  - Fluorescence-activated cell sorting (FACS)
  - Whole genome amplification
  - Robotic liquid handling
  - Cleanroom environment
  - DNA sequencing
  - *De novo* assembly and other bioinformatics
  - Laboratory information management systems (LIMS)

Single Cell Genomics Pipeline



Environment

Physiology
Ecology
Evolution
Biotechnology

Metagenome
Metaproteome
Biogeochemical cues
Cultures

Single cell isolation

Targeted loci

Genome reconstruction

Cell lysis

Whole genome amplification

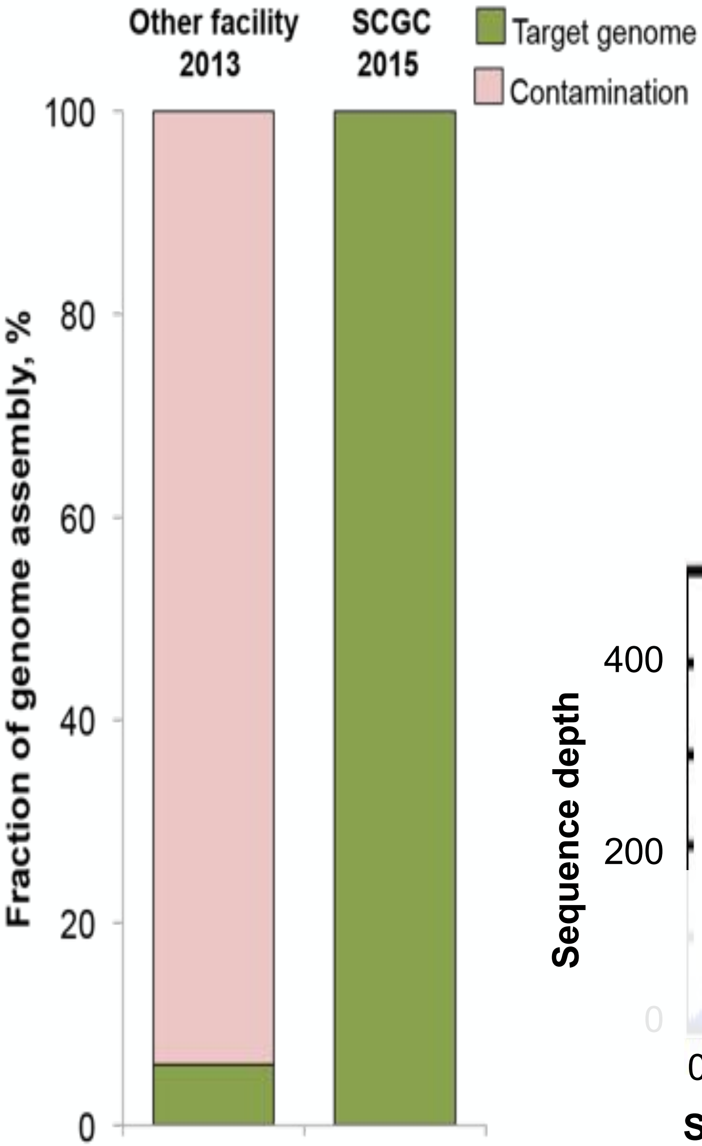
DNA archive

Bigelow | Laboratory for Ocean Sciences

SAG cross-contamination during sequencing

Illumina library cross-contamination

SAG SCGC AC-310-N17 (Firmicutes)



Mechanisms:

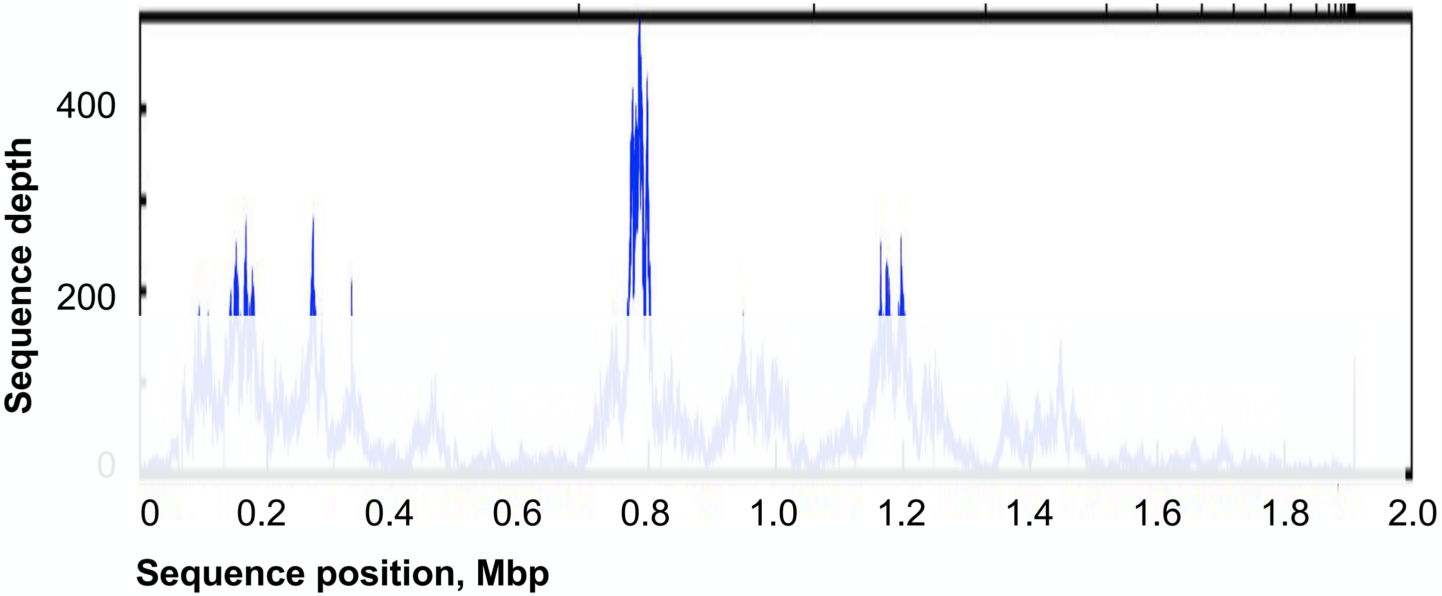
- Miss-assignment of reads among multiplexed libraries
- Sample carry-over between runs

Solution:

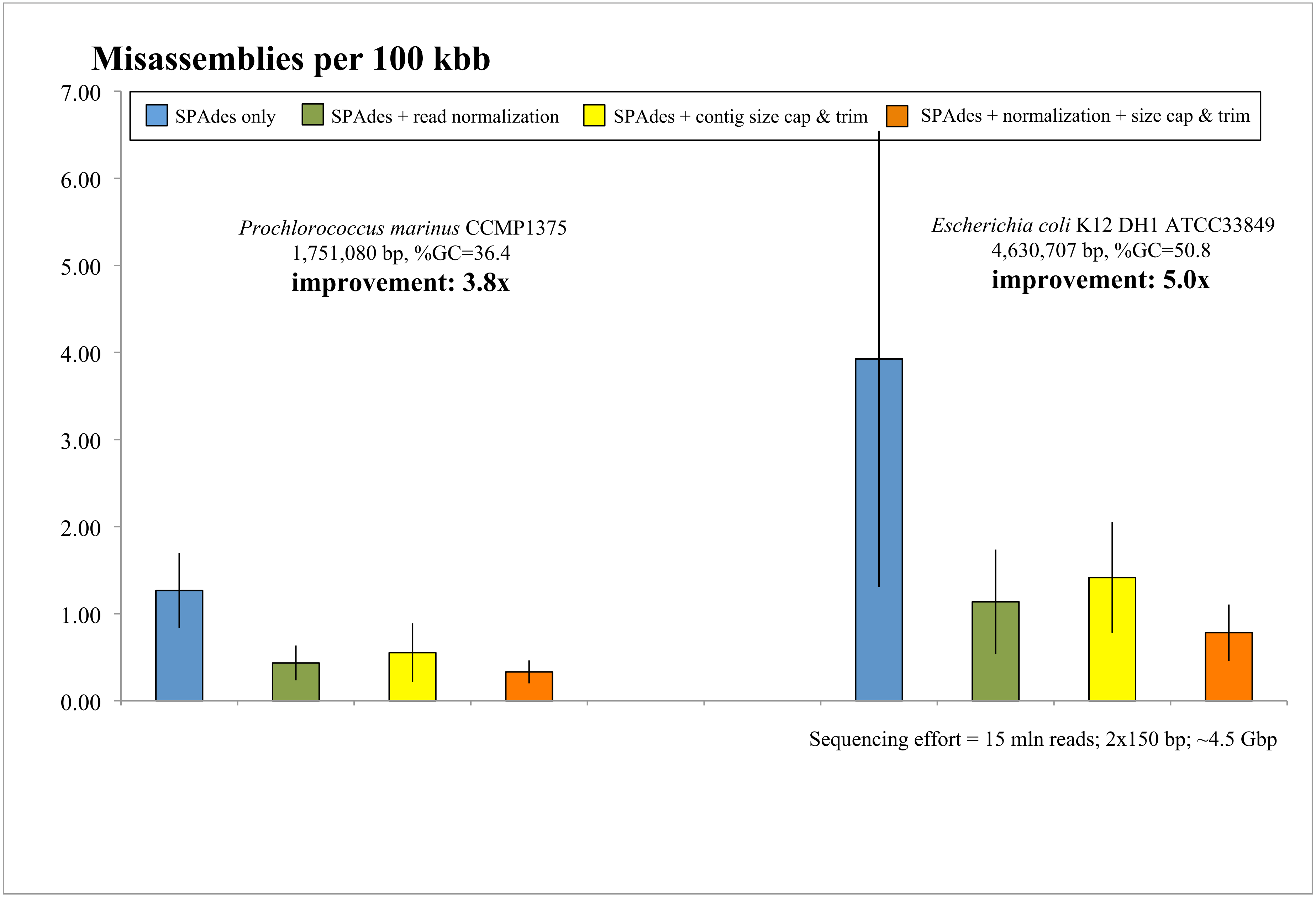
- Dual barcoding
- Use of NextSeq
- Extra care and validation of each step

Outcome:

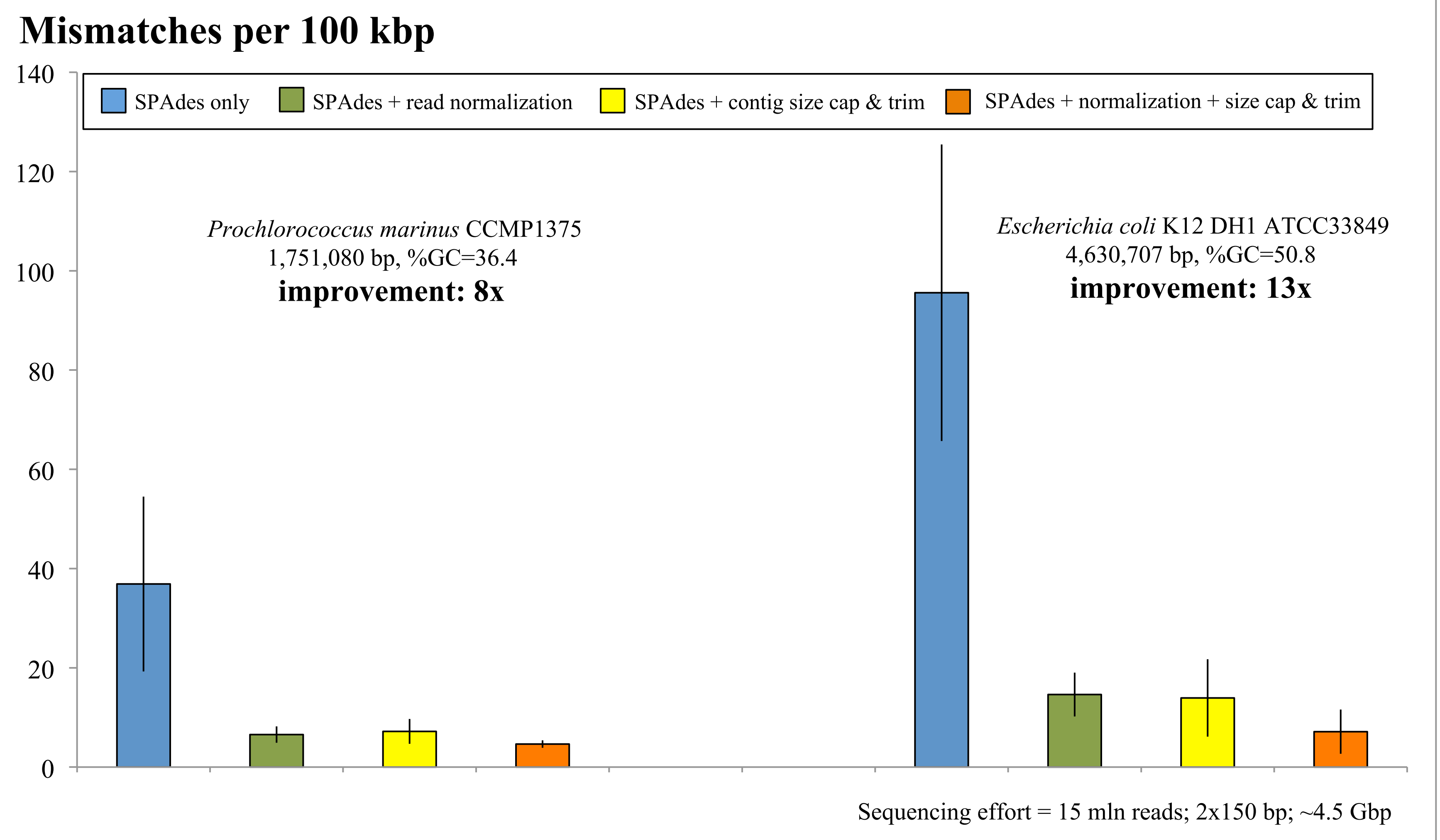
- ~50 SAGs sequenced worldwide in 2013-2014
- ~1k/10k SAGs sequenced at SCGC in 2015-2016



Improved SAG *de novo* assembly

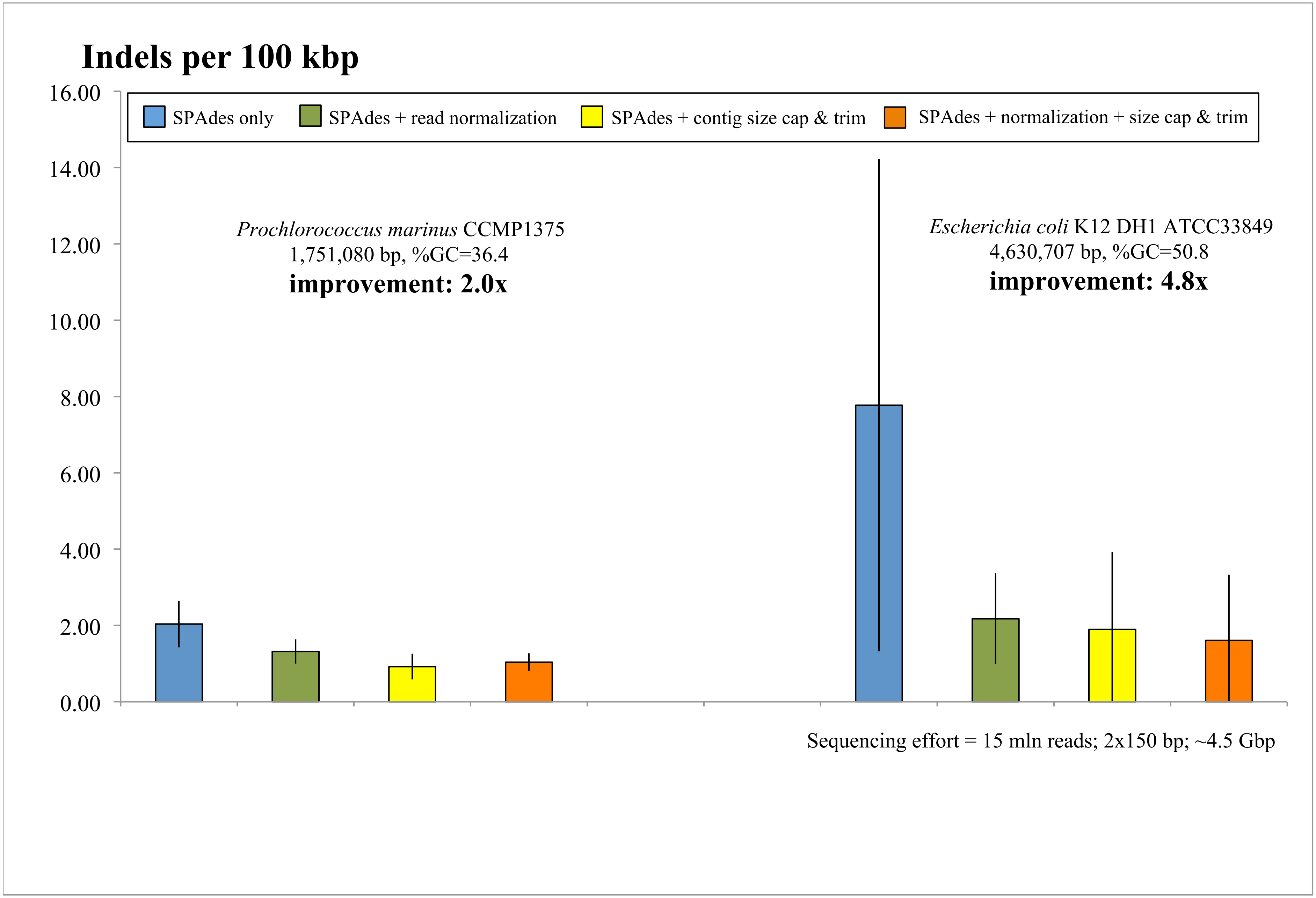


Improved SAG *de novo* assembly

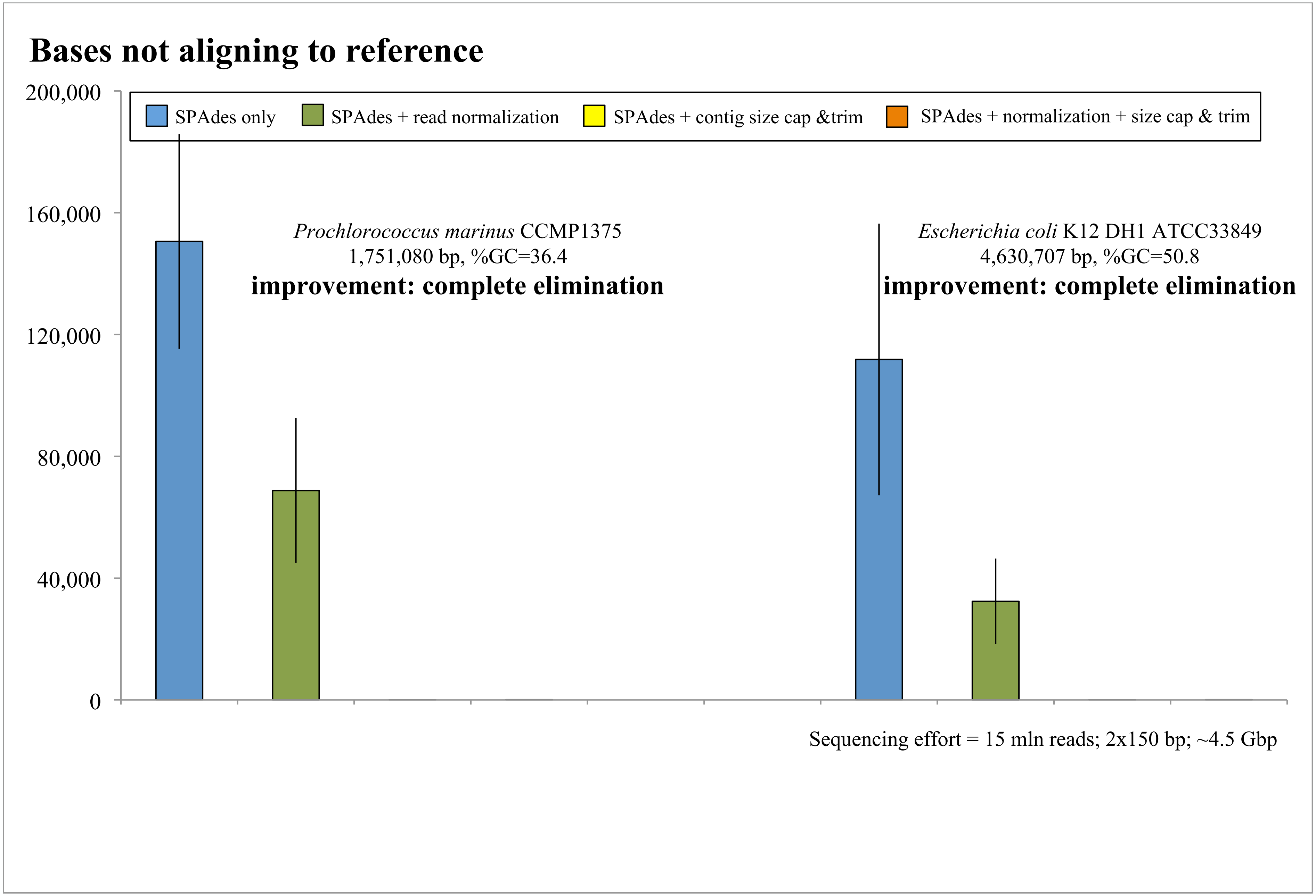


Sequencing effort = 15 mln reads; 2x150 bp; ~4.5 Gbp

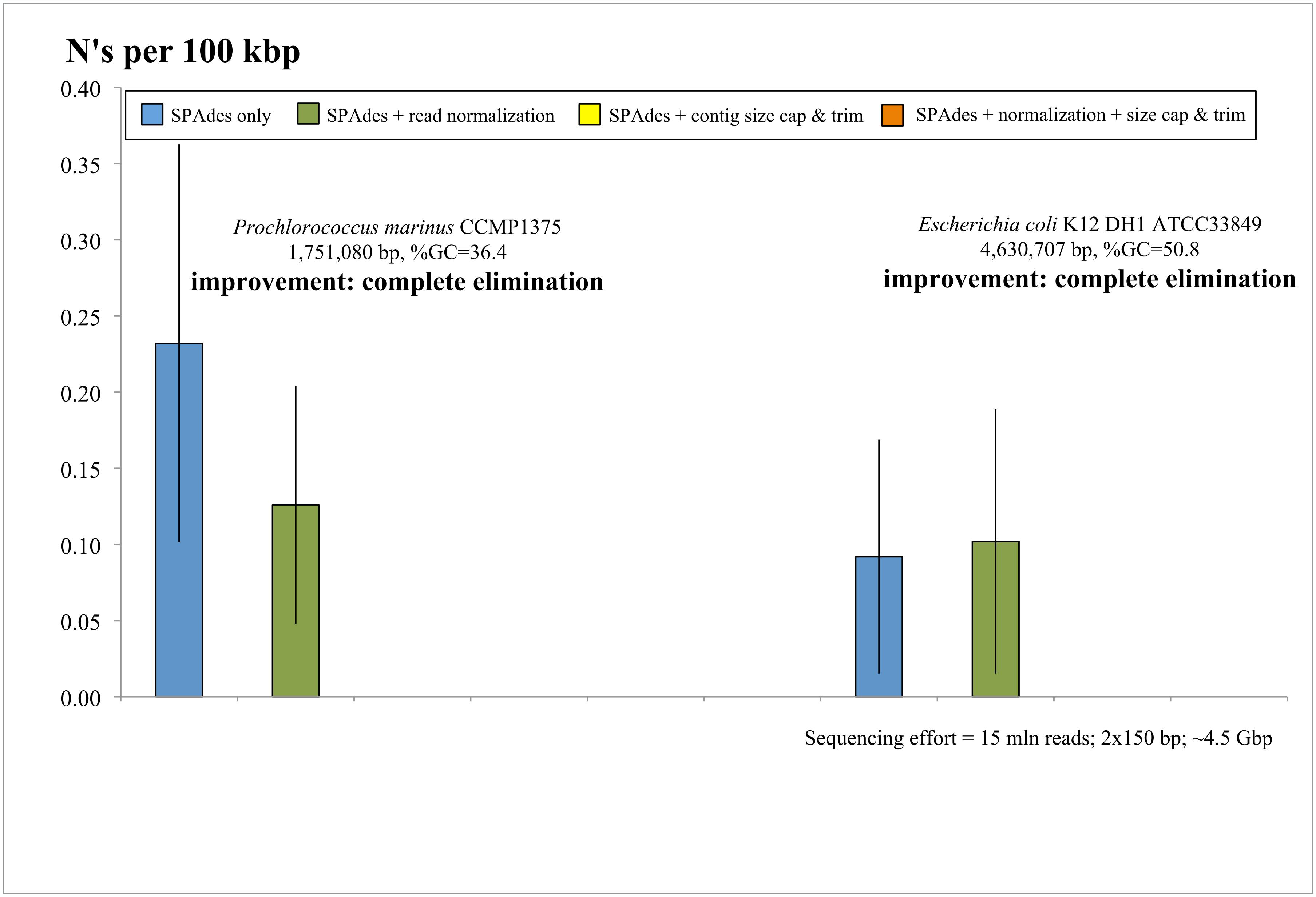
Improved SAG *de novo* assembly



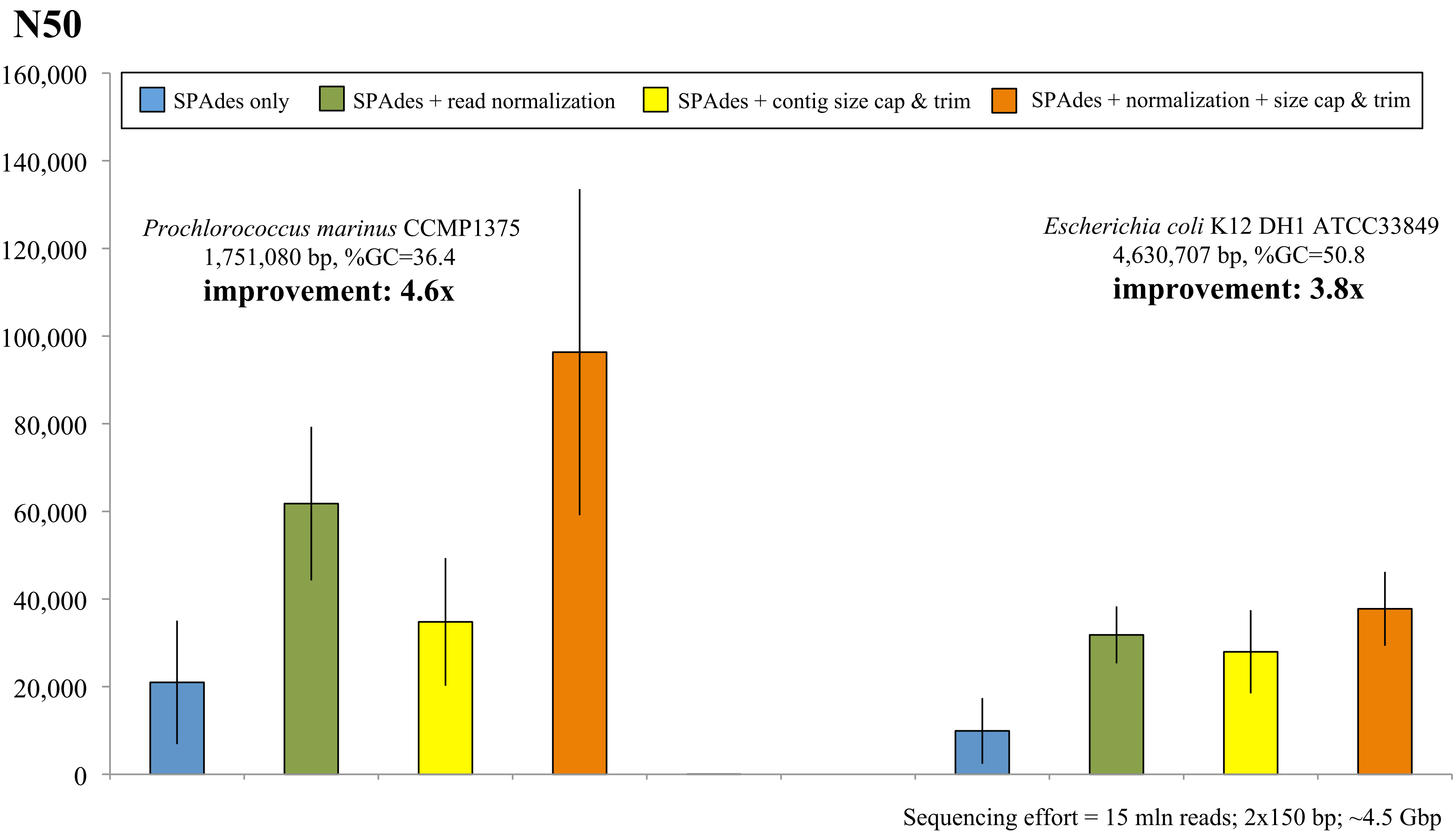
Improved SAG *de novo* assembly



Improved SAG *de novo* assembly

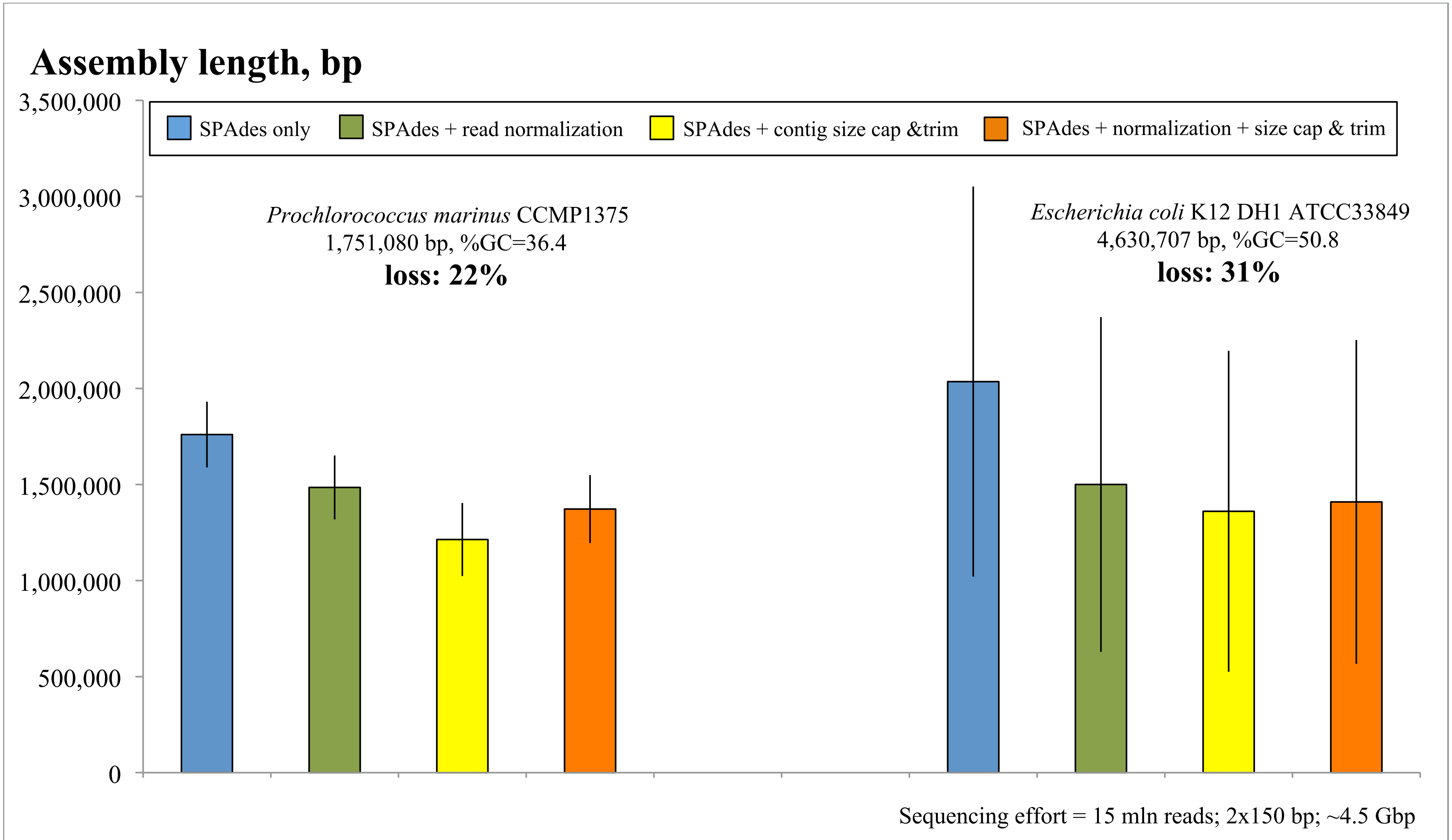


Improved SAG *de novo* assembly



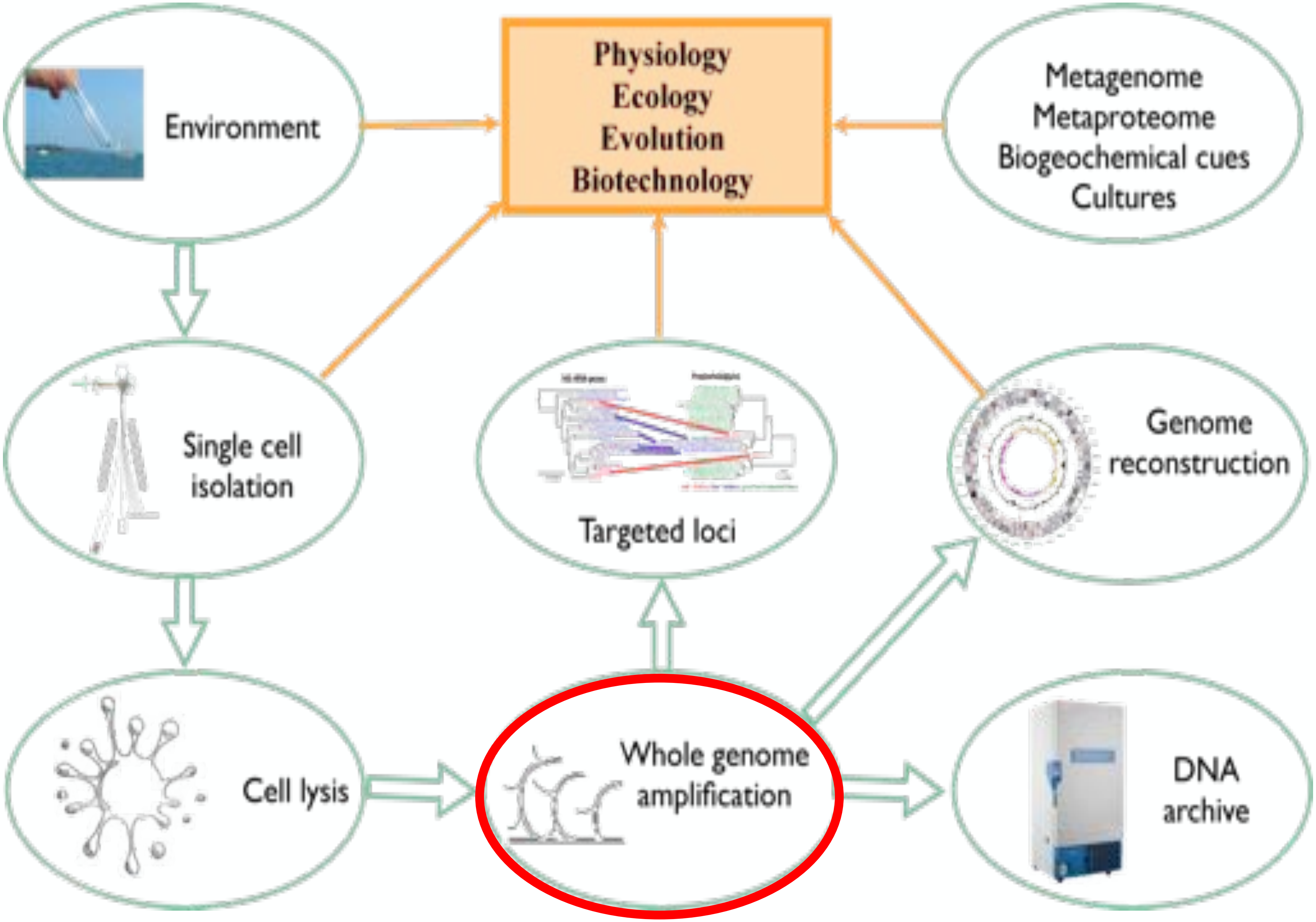
Sequencing effort = 15 mln reads; 2x150 bp; ~4.5 Gbp

Improved SAG *de novo* assembly



Sequencing effort = 15 mln reads; 2x150 bp; ~4.5 Gbp

Single Cell Genomics Pipeline



Environment

Physiology
Ecology
Evolution
Biotechnology

Metagenome
Metaproteome
Biogeochemical cues
Cultures

Single cell isolation

Targeted loci

Genome reconstruction

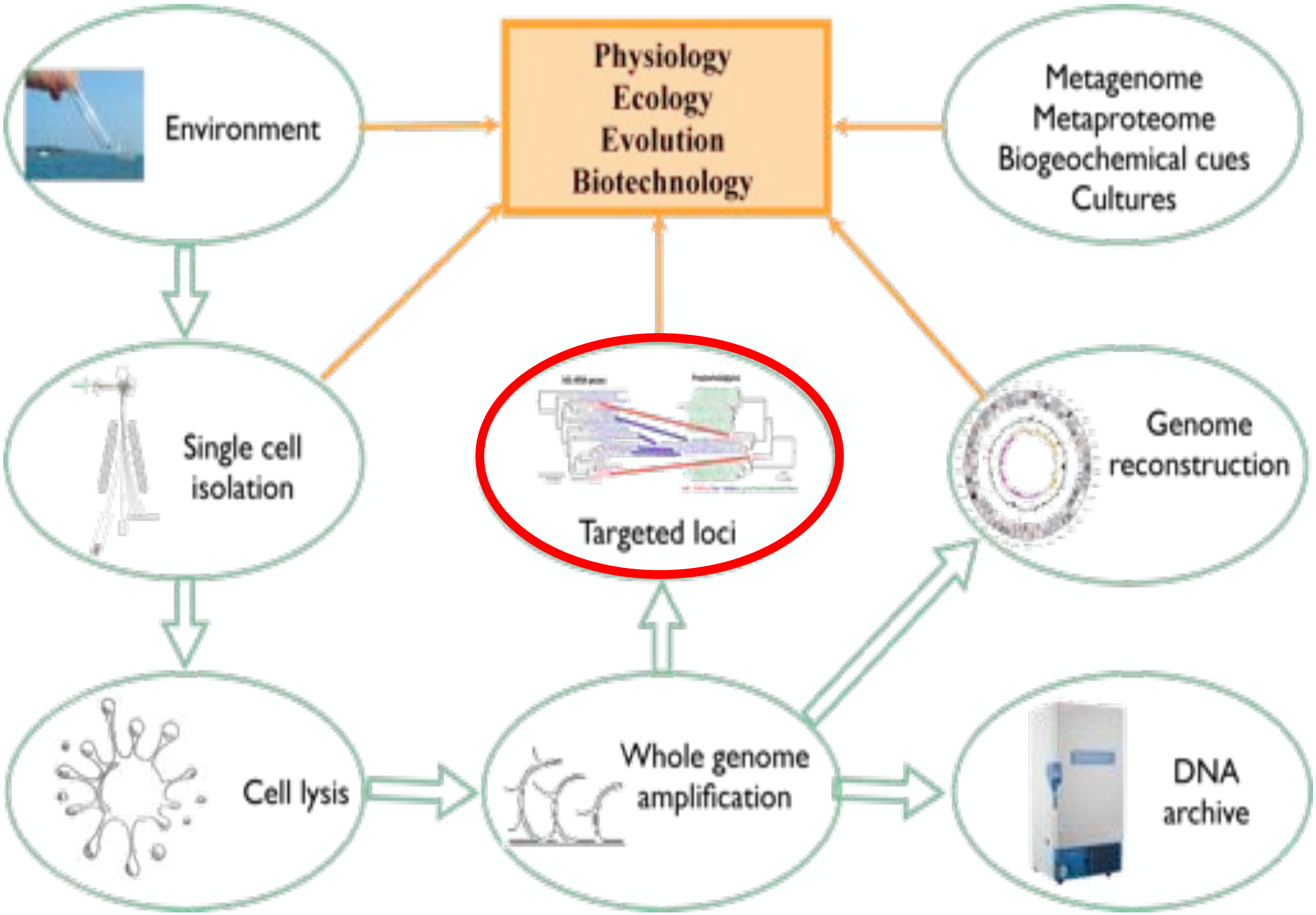
Cell lysis

Whole genome amplification

DNA archive

Bigelow | Laboratory for Ocean Sciences

Single Cell Genomics Pipeline



Environment

Physiology
Ecology
Evolution
Biotechnology

Metagenome
Metaproteome
Biogeochemical cues
Cultures

Single cell isolation

Targeted loci

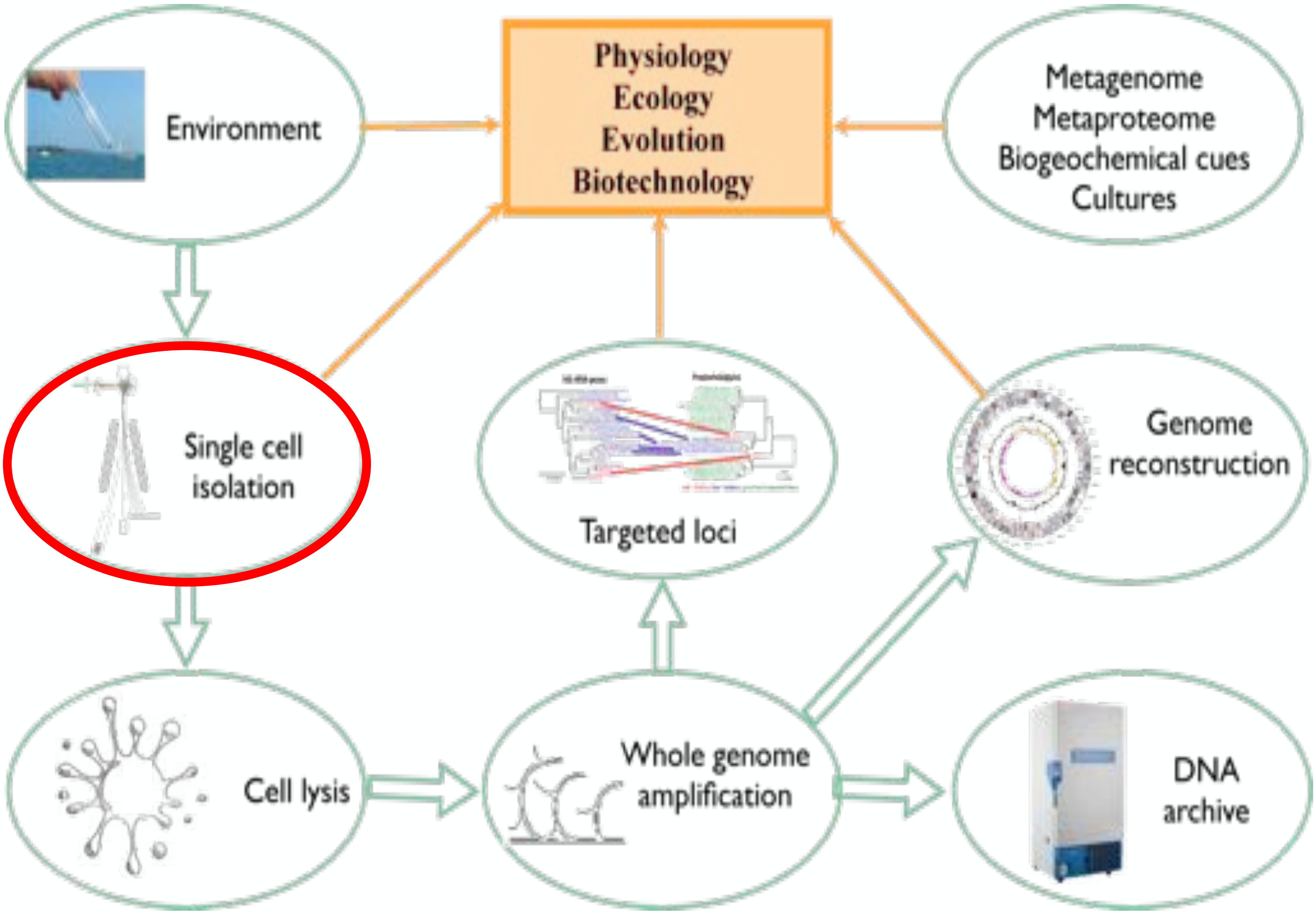
Genome reconstruction

Cell lysis

Whole genome amplification

DNA archive

Single Cell Genomics Pipeline



Environment

Physiology
Ecology
Evolution
Biotechnology

Metagenome
Metaproteome
Biogeochemical cues
Cultures

Single cell isolation

Targeted loci

Genome reconstruction

Cell lysis

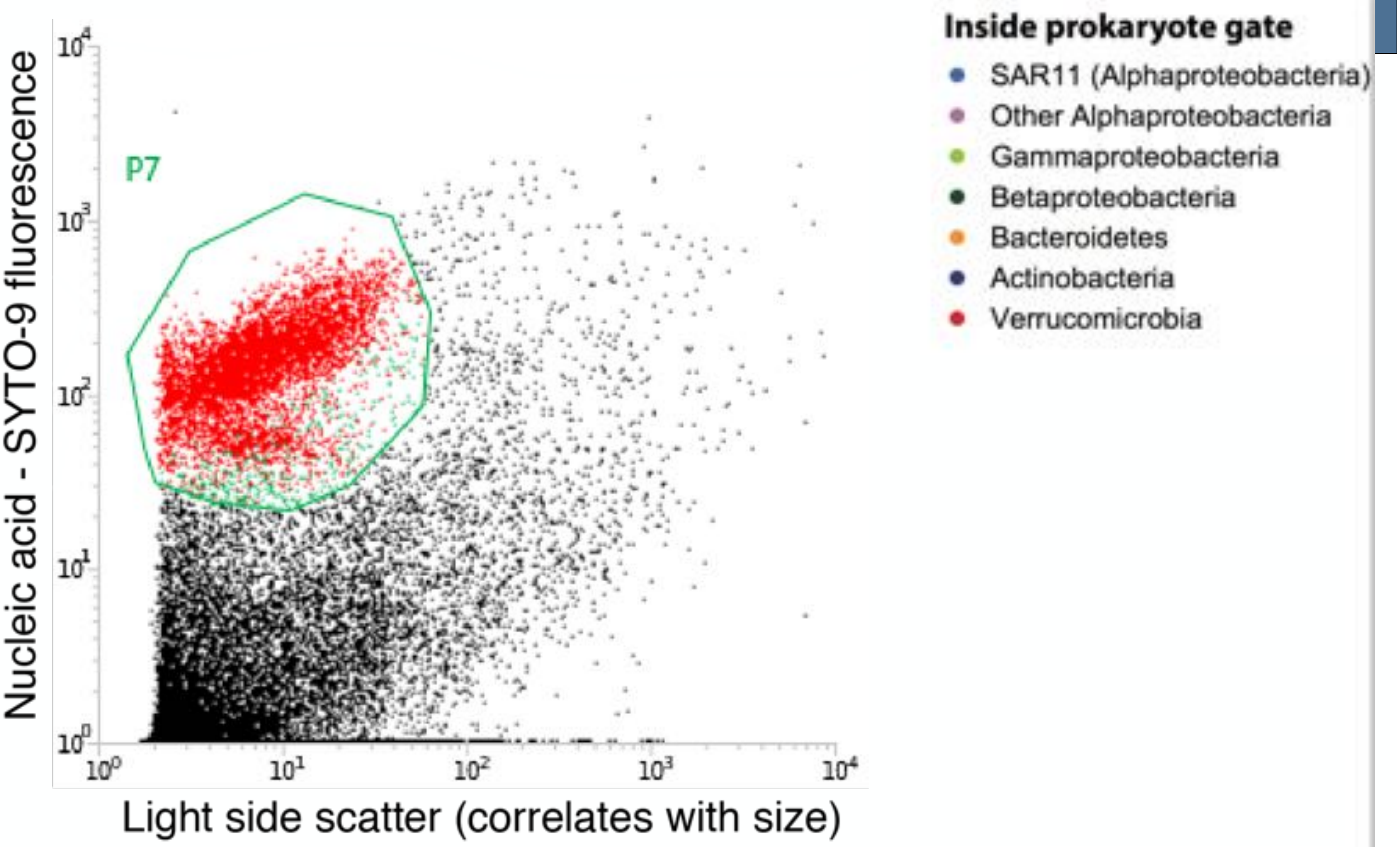
Whole genome amplification

DNA archive

Bigelow | Laboratory for Ocean Sciences

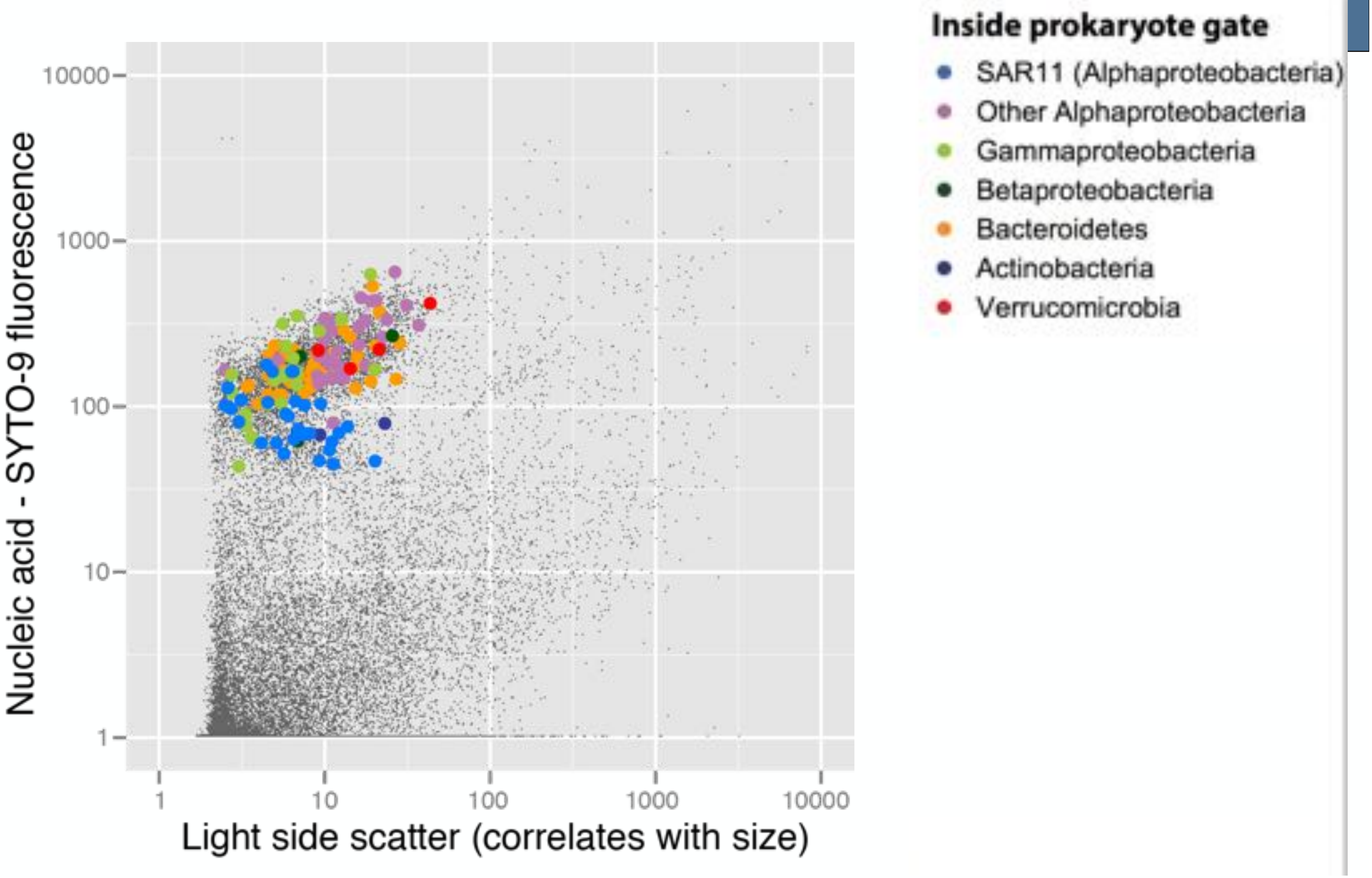
Traditional fluorescence-activated cell sorting

Gulf of Maine surface microbial community



Index cell sorting integration with single cell genomics

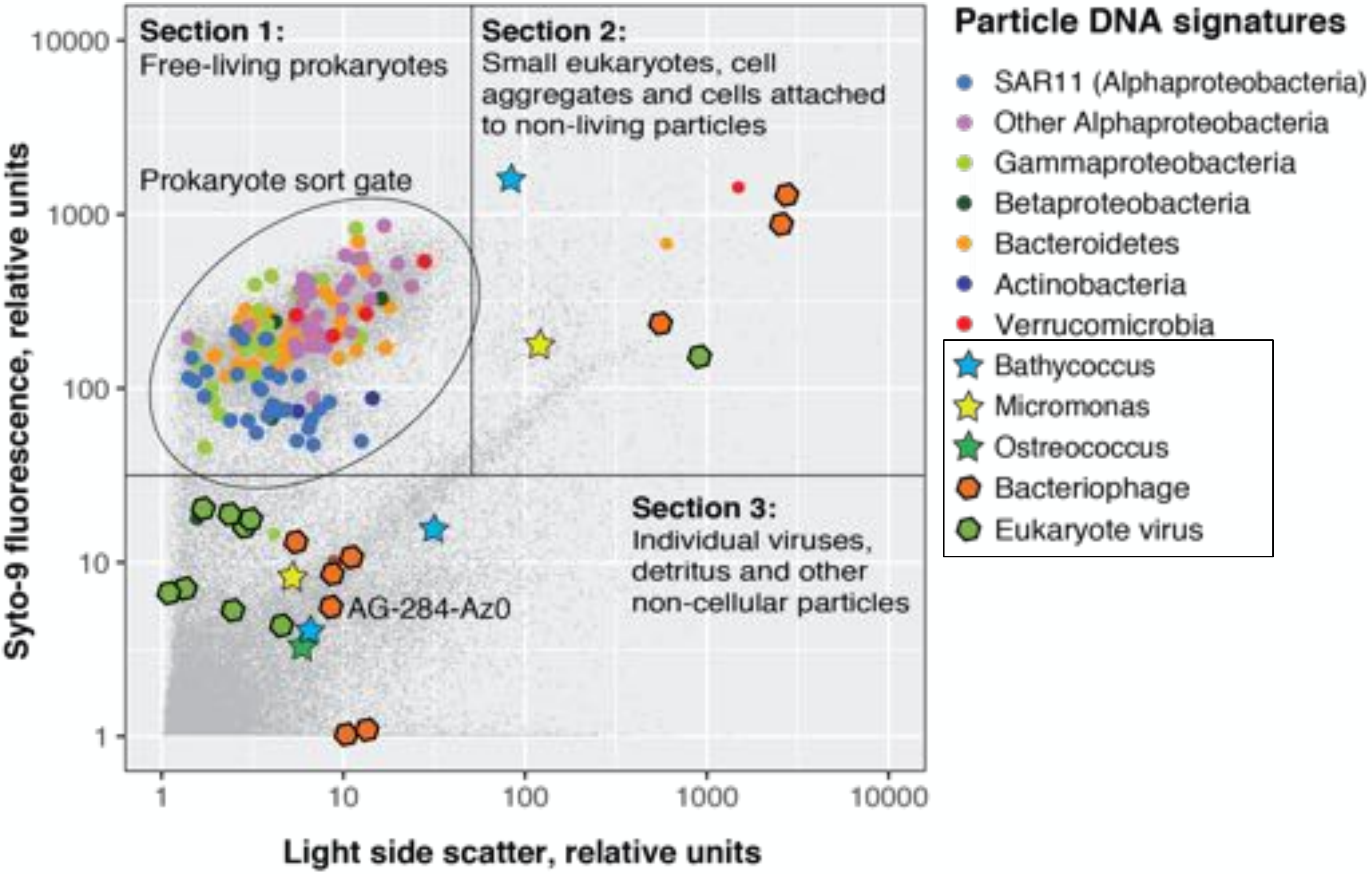
Gulf of Maine surface microbial community



Bigelow | Laboratory for Ocean Sciences

Genomics of individual extracellular particles

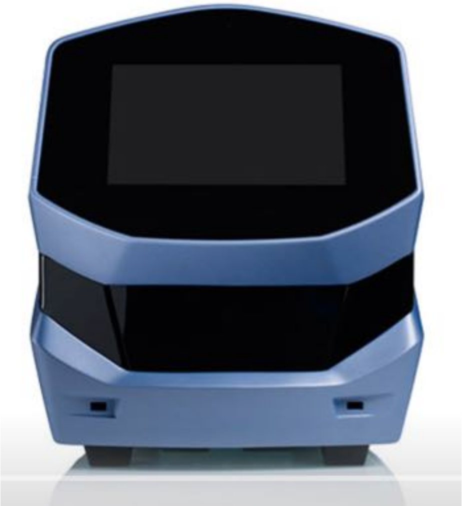
Gulf of Maine surface microbial community



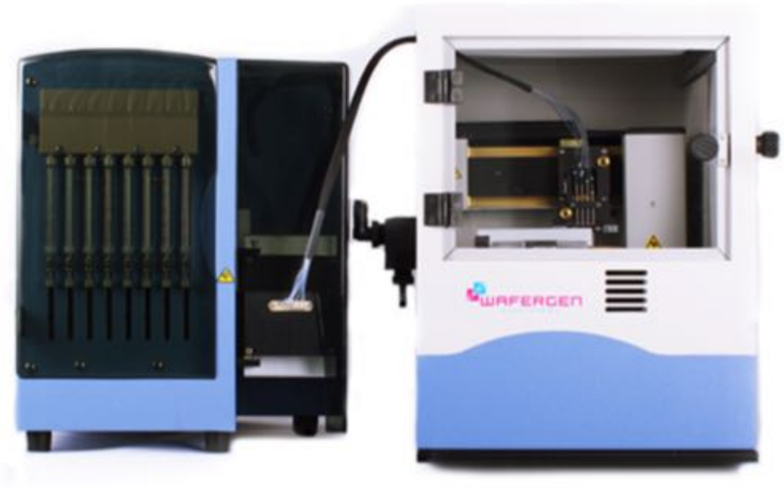
Bigelow Laboratory for Ocean Sciences

Commercialized systems for human single cell transcriptomics

C1 (Fluidigm)



ICELL8 (Wafergen)

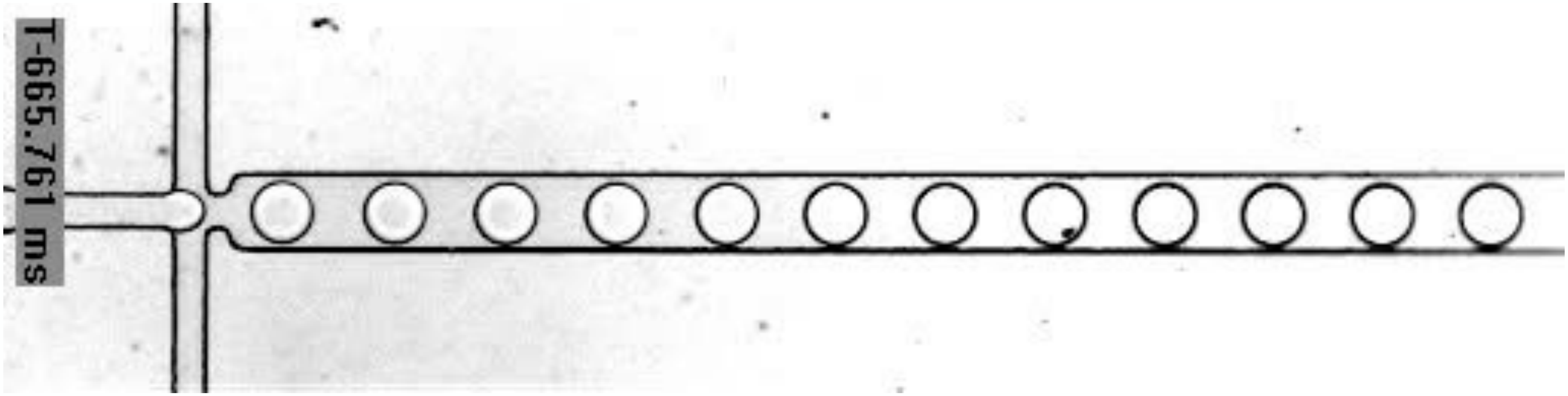


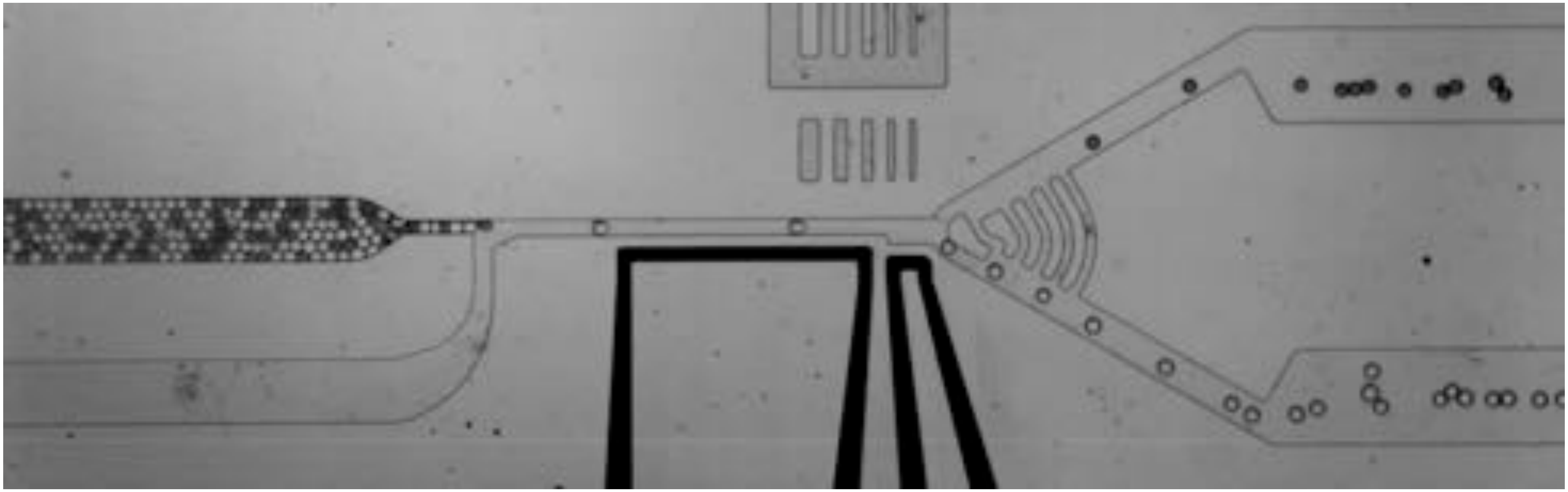
Chromium (10x)

ddSEQ (BIO-RAD & Illumina)

RNA –Seq System (Dolomite Bio)

Droplet microfluidics: Future of single cell genomics?





Video courtesy Linas Mazutis

Take-home message

- Microbial single cell genomics currently requires major investment into specialized facilities

- Methods are improving rapidly for enhanced scalability, genome recovery and integration with single cell phenomics