MaCuMBA

Grupo de Genómica Evolutiva
Evolutionary Genomics Group

Marine Metaviromes

Carolina Megumi Mizuno

PhD. Student

Evolutionary Genomics Groups

Universidad Miguel Hernández

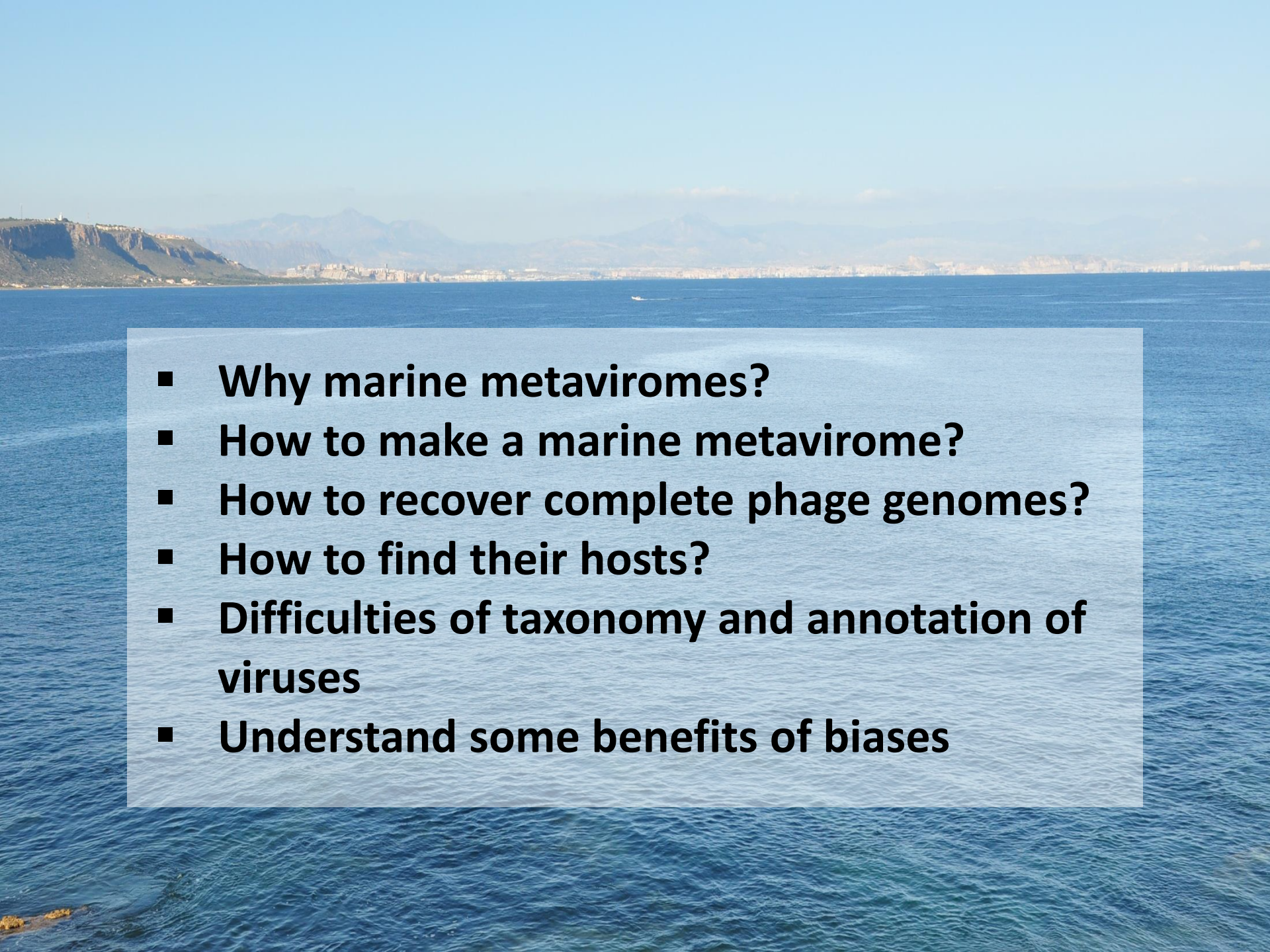
µgen

Gobierno de España
Ministerio de Ciencia e Innovación

Consolider

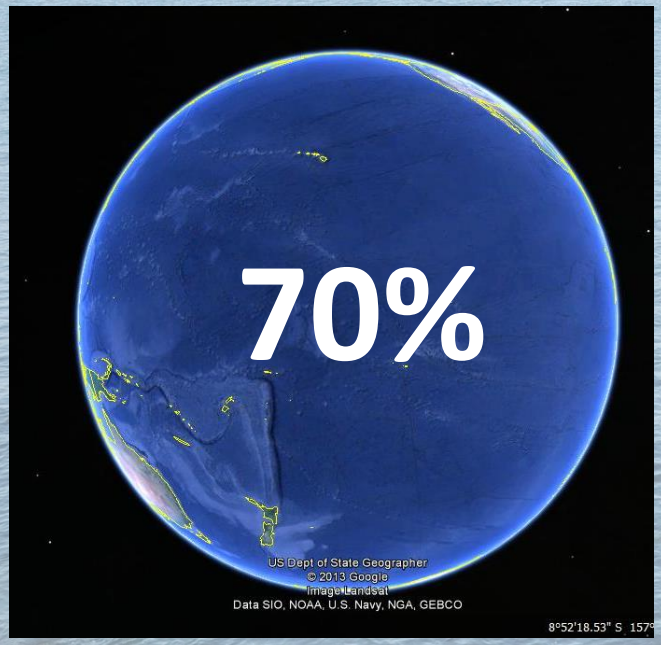
The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 311957.

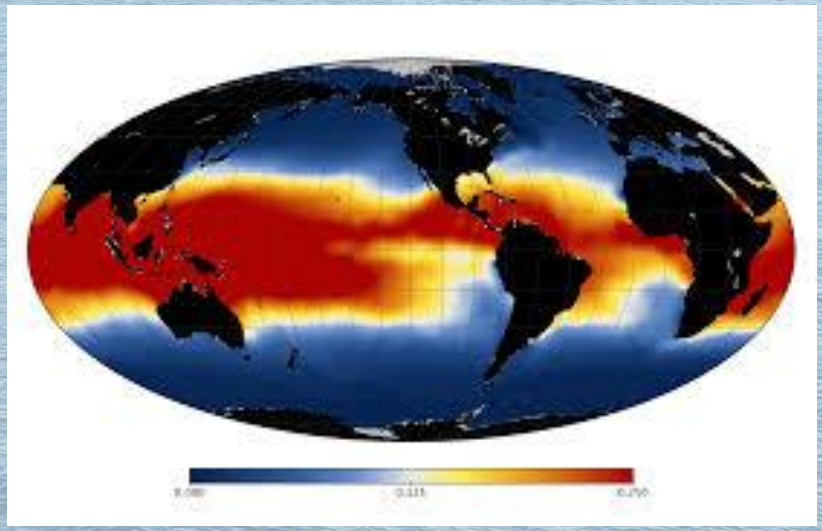
www.macumbaproject.eu



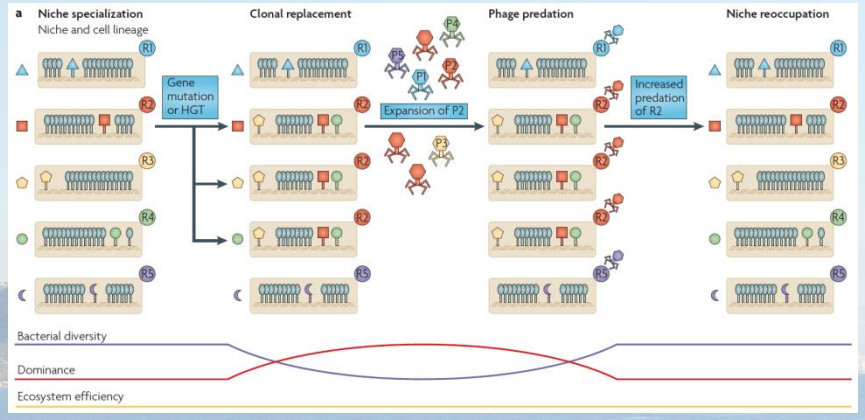
- Why marine metaviromes?
- How to make a marine metavirome?
- How to recover complete phage genomes?
- How to find their hosts?
- Difficulties of taxonomy and annotation of viruses
- Understand some benefits of biases

Why marine???



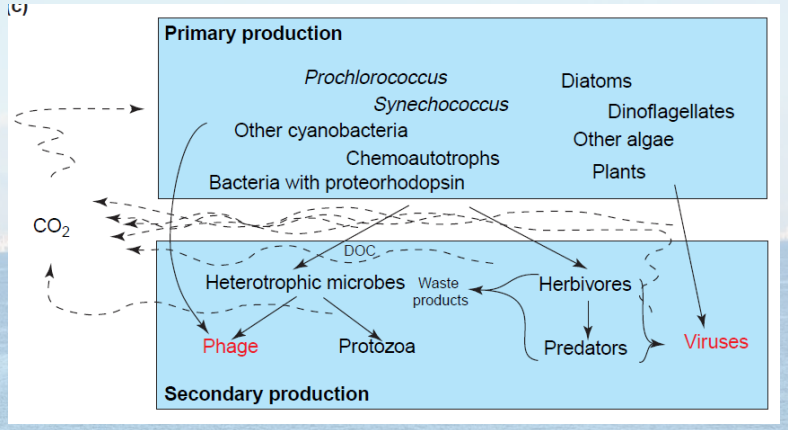


50% primary production

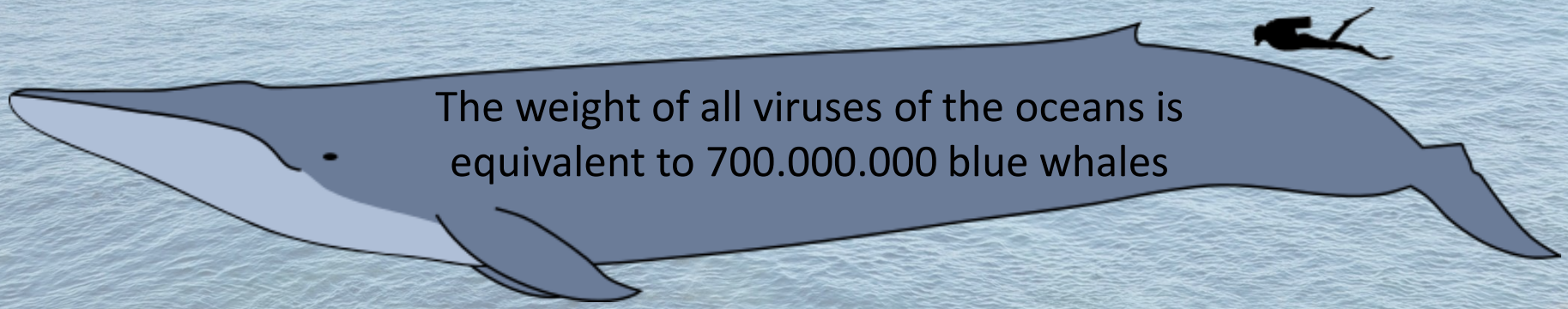


They play an important role in microbial population

and consequently in biogeochemical cycles



Why viruses?



The weight of all viruses of the oceans is equivalent to 700.000.000 blue whales

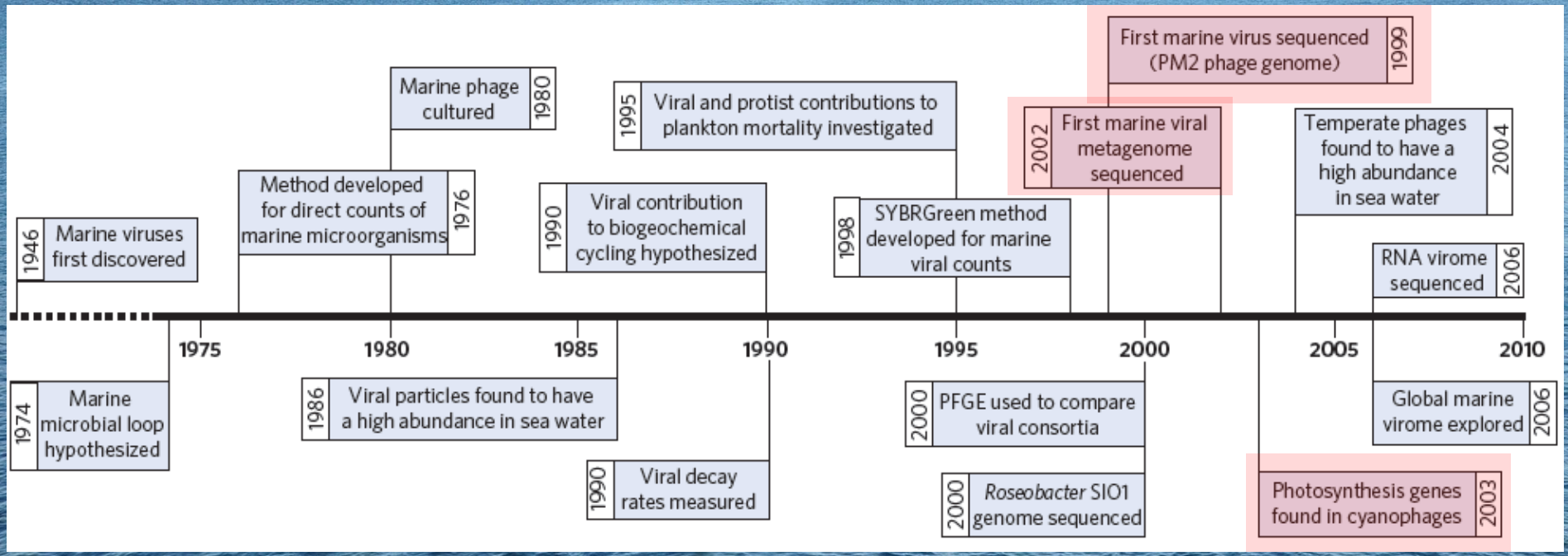
NATURE|Vol 459|14 May 2009|doi:10.1038/nature08060

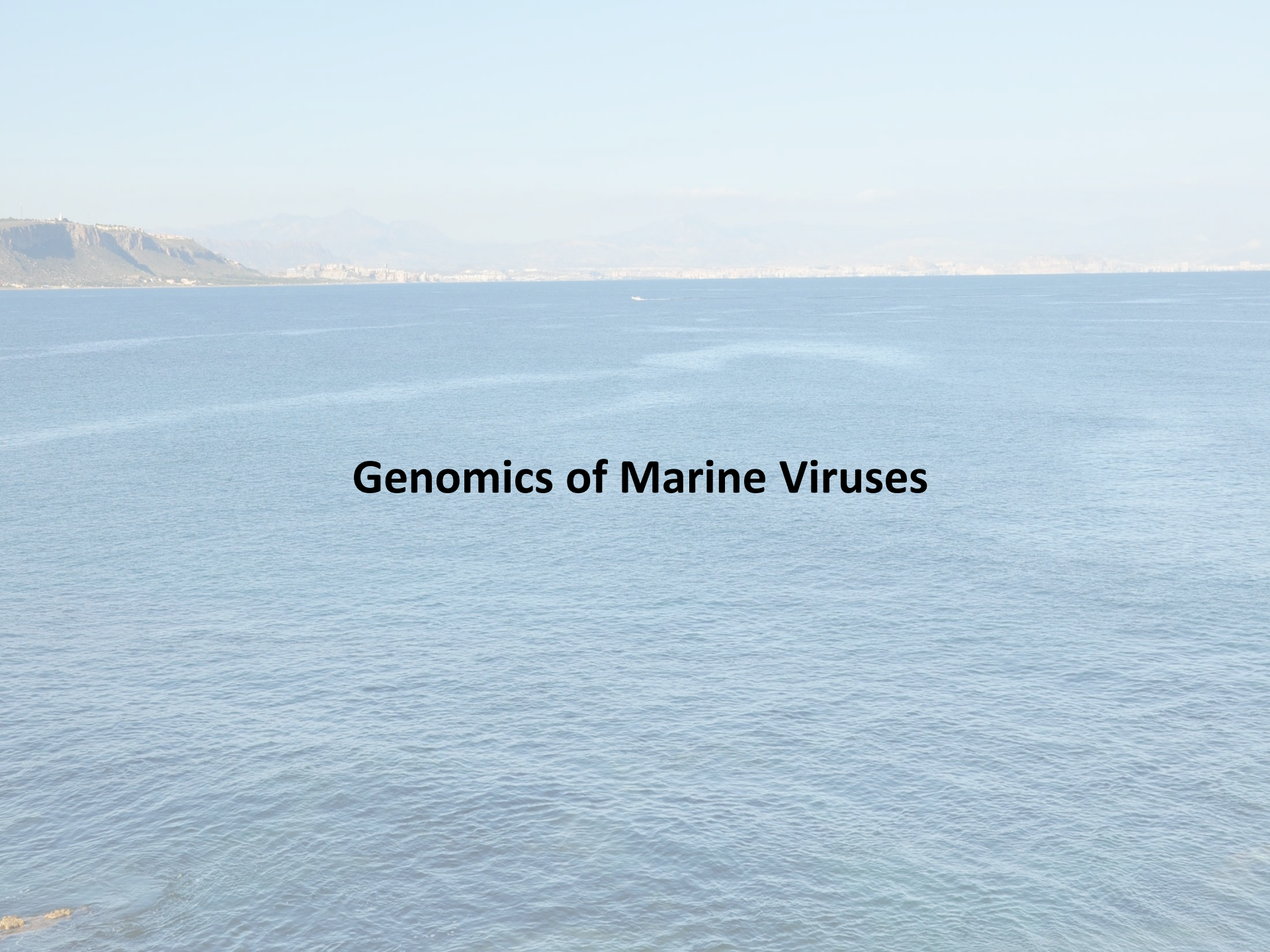
Viruses manipulate the marine environment

Forest Rohwer[1] & Rebecca Vega Thurber[1,2]

Highlights in marine virology research

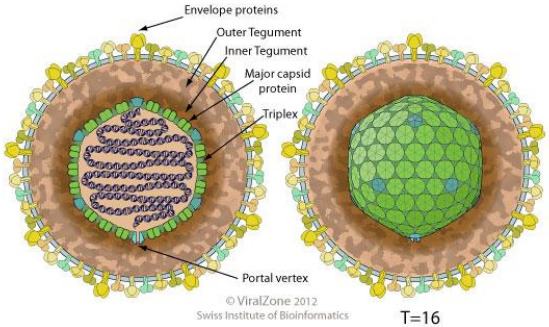
Genomic and metagenomics of marine viruses

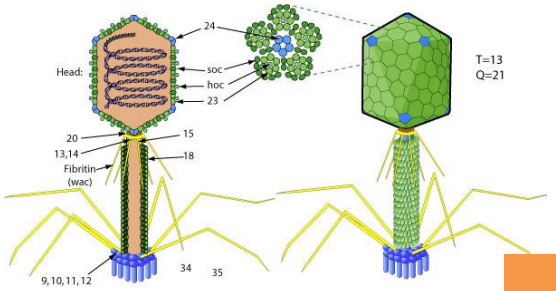




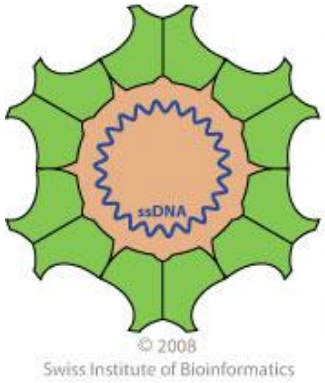
Genomics of Marine Viruses

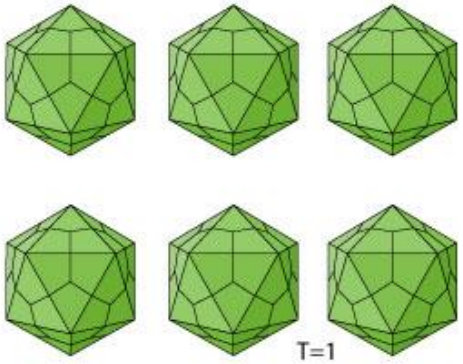
dsDNA

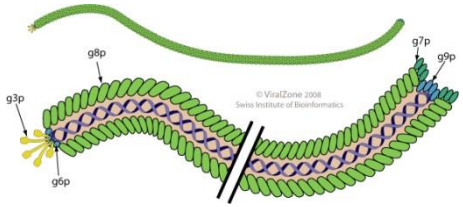


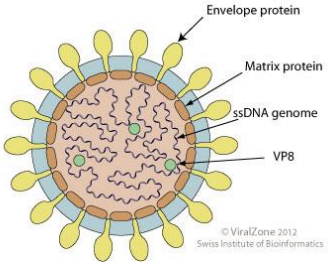


ssDNA



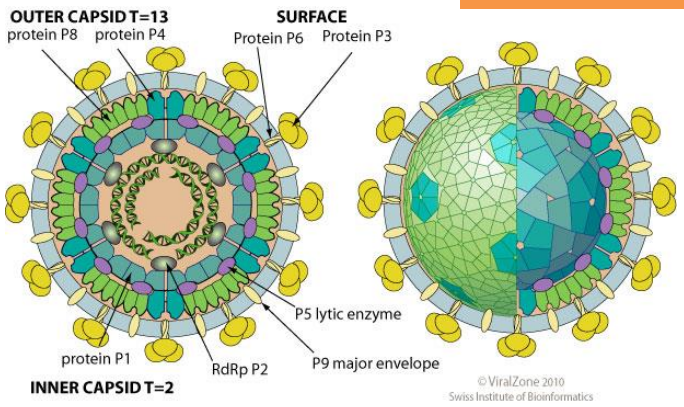




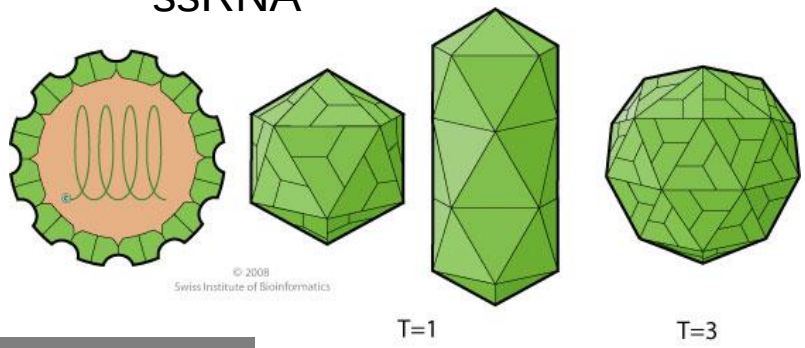


No gene as such the 16S for prokaryotes

dsRNA



ssRNA



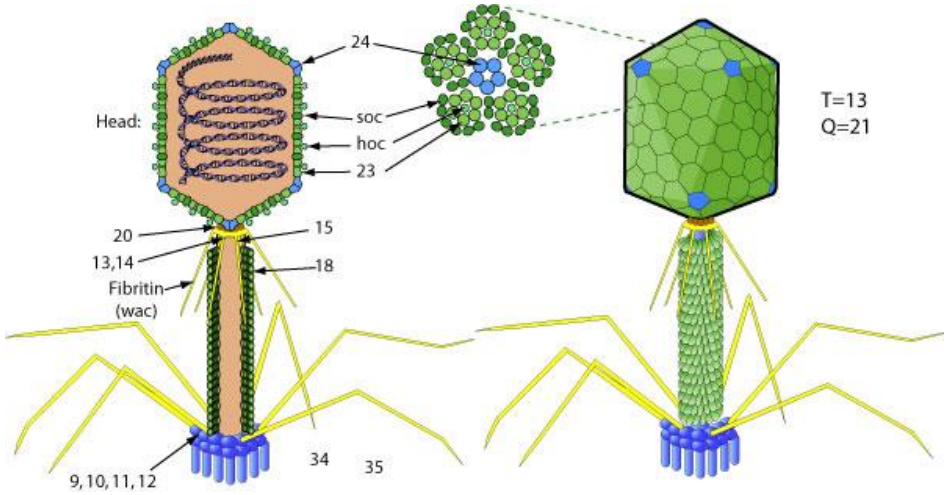
NIGHTMARE for taxonomists

ViralZone

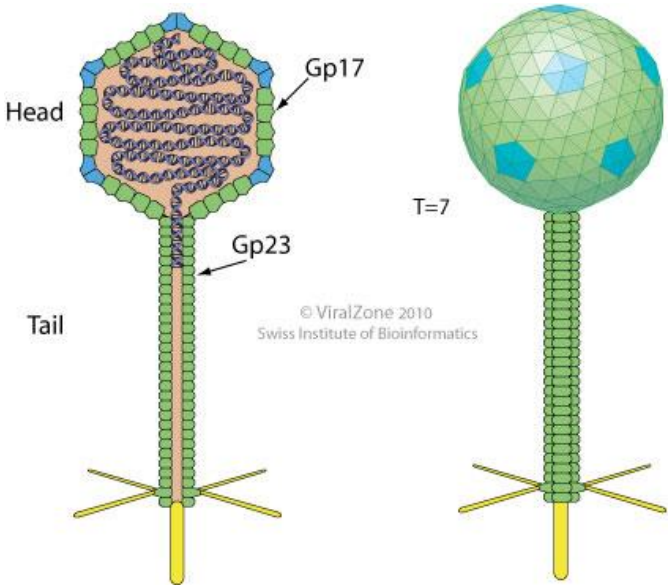
dsDNA viruses

Caudovirales (head-tail)

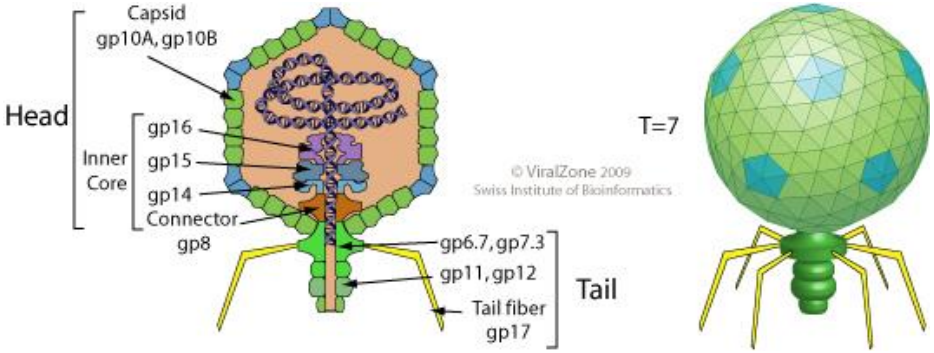
Myoviruses (T4-like)



Siphoviruses (Lambda-like)

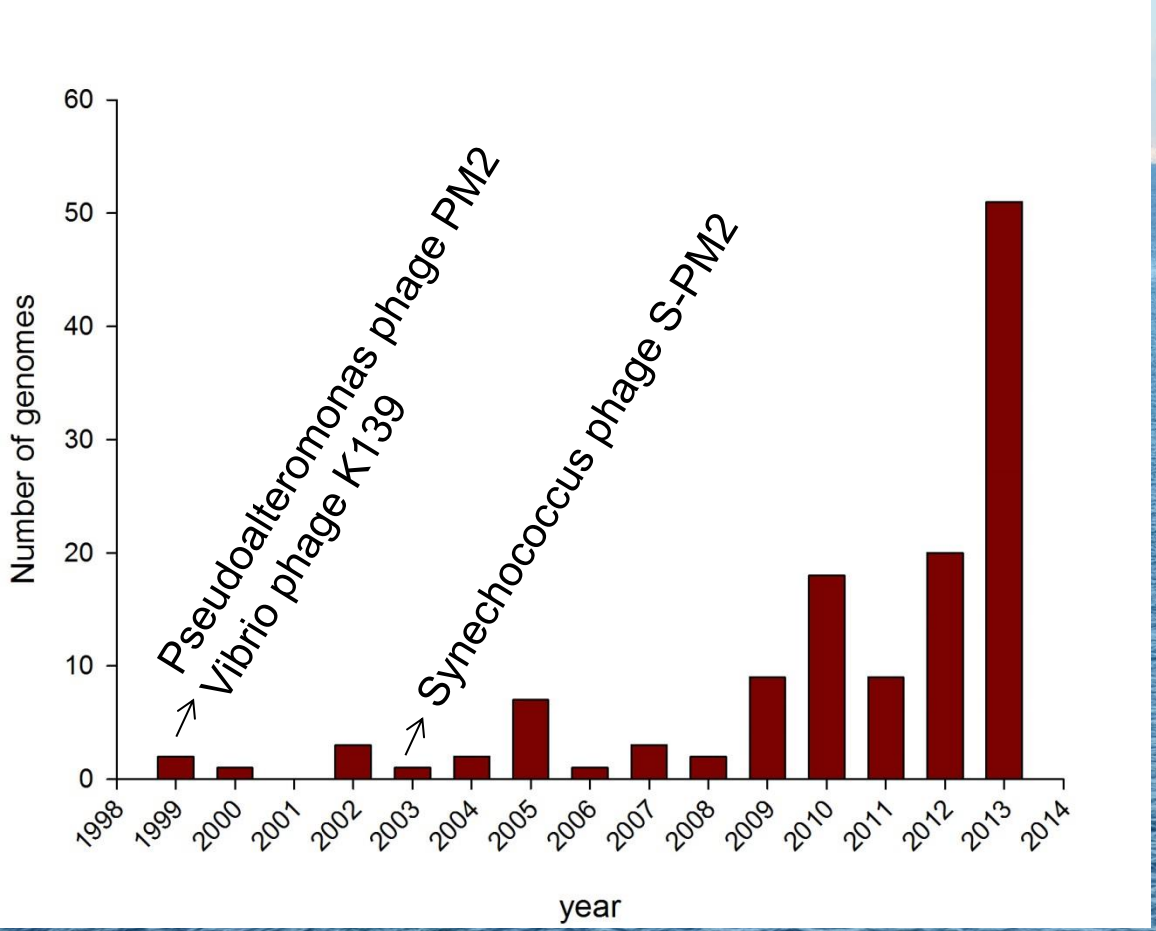


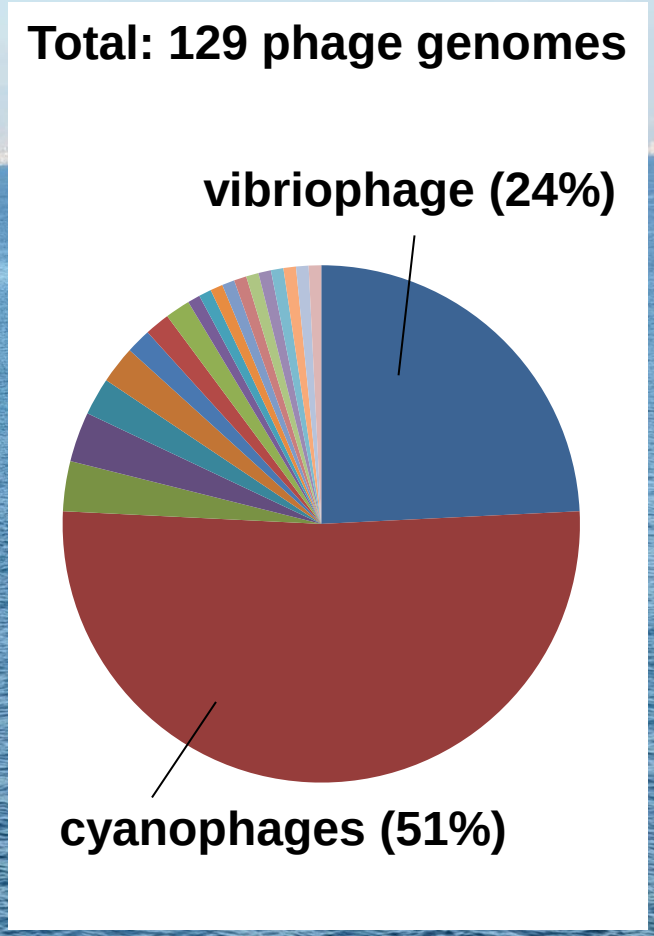
Podoviruses (T7-like)



ViralZone

Genomics of marine dsDNA viruses





Prokaryotic viruses = bacteriophages = phages

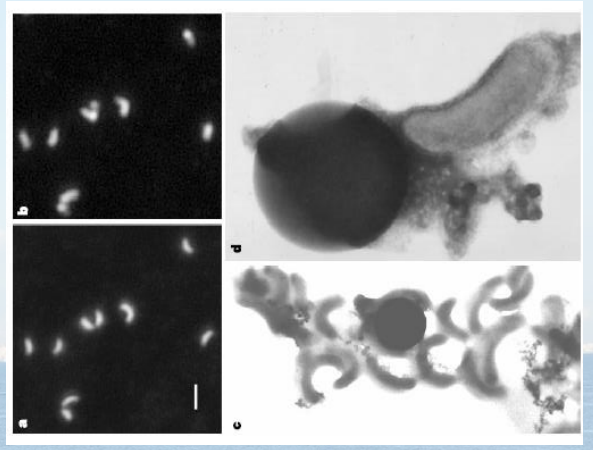
Why metaviromes???

NATURE · VOL 345 · 3 MAY 1990

Genetic diversity in Sargasso Sea bacterioplankton

Stephen J. Giovannoni, Theresa B. Britschgi, Craig L. Moyer & Katharine G. Field

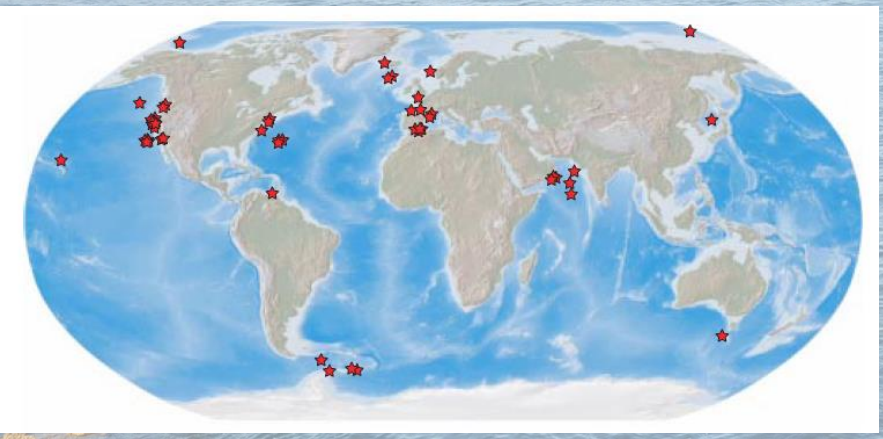
12 years



NATURE | VOL 418 | 8 AUGUST 2002

Cultivation of the ubiquitous SAR11 marine bacterioplankton clade

Michael S. Rappé, Stephanie A. Connon, Kevin L. Vergin & Stephen J. Giovannoni



NATURE | VOL 420 | 19/26 DECEMBER 2002

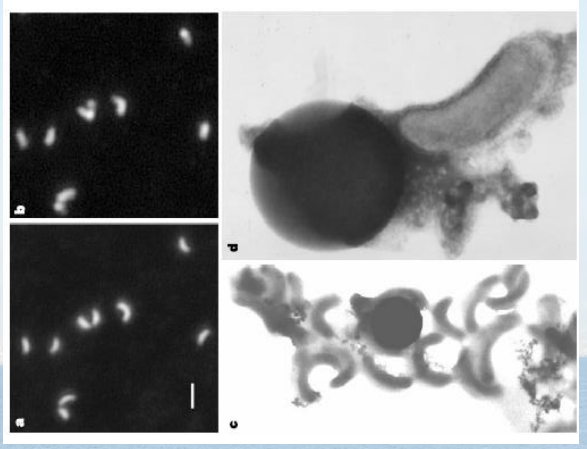
SAR11 clade dominates ocean surface bacterioplankton communities

Robert M. Morris*, Michael S. Rappé*, Stephanie A. Connon*, Kevin L. Vergin*, William A. Siebold*, Craig A. Carlson† & Stephen J. Giovannoni*

NATURE | VOL 418 | 8 AUGUST 2002

Cultivation of the ubiquitous SAR11 marine bacterioplankton clade

Michael S. Rappé, Stephanie A. Connon, Kevin L. Vergin & Stephen J. Giovannoni

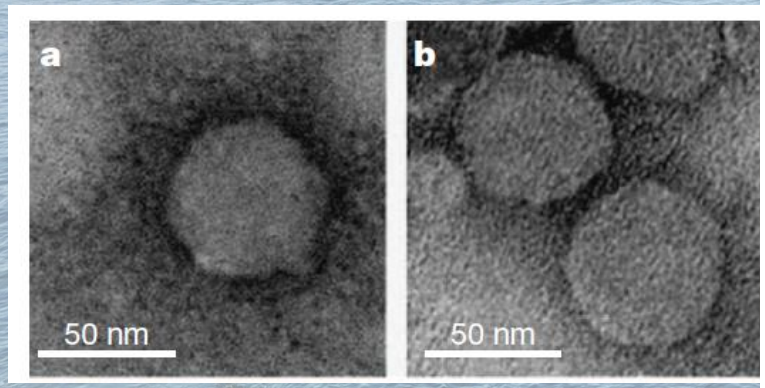


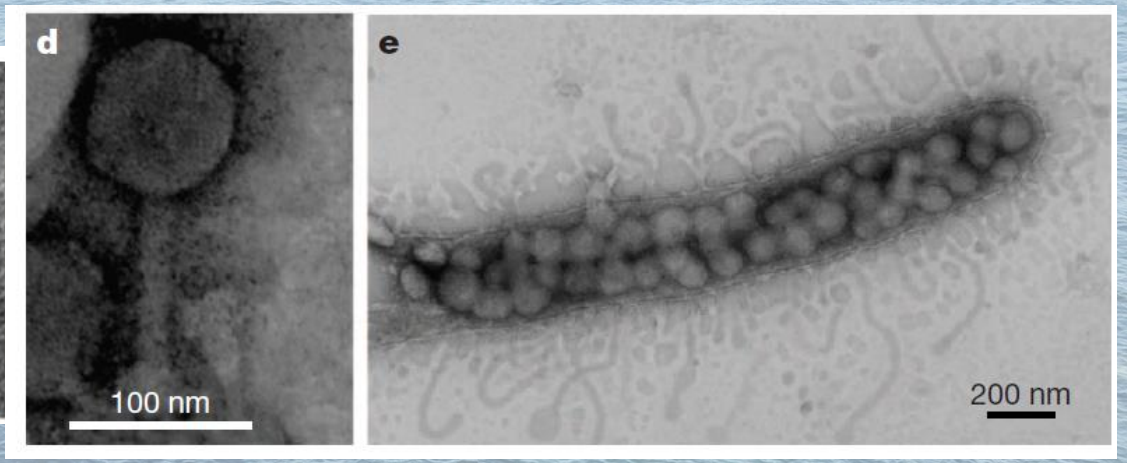
11 years

21 FEBRUARY 2013 | VOL 494 | NATURE | 359

Abundant SAR11 viruses in the ocean

Yanlin Zhao[1]*, Ben Temperton[1]*, J. Cameron Thrash[1], Michael S. Schwalbach[2], Kevin L. Vergin[1], Zachary C. Landry[1], Mark Ellisman[3], Tom Deerinck[3], Matthew B. Sullivan[4] & Stephen J. Giovannoni[1]





Limnol. Oceanogr., 24(5), 1979, 928–935
© 1979, by the American Society of Limnology and Oceanography, Inc.

Chroococcoid cyanobacteria in the sea: A ubiquitous and diverse phototrophic biomass[1]

10 years

Nature **334**, 340-343 (28 July 1988) | doi:10.1038/334340a0; Accepted 10 June 1988

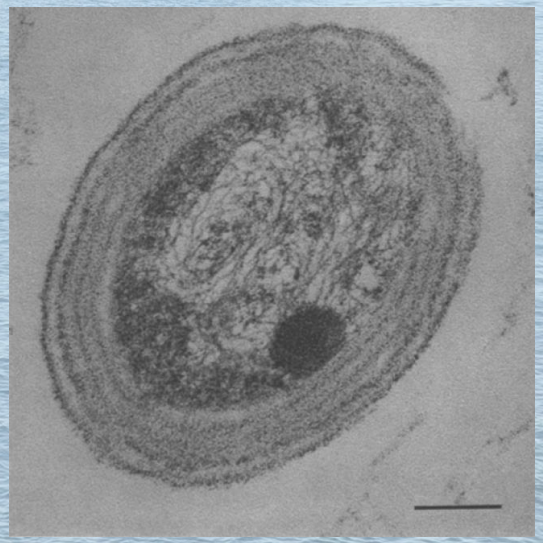
A novel free-living prochlorophyte abundant in the oceanic euphotic zone

Sallie W. Chisholm, Robert J. Olson*, Erik R. Zettler*, Ralf Goericke†, John B. Waterbury* & Nicholas A. Welschmeyer†

Arch Microbiol (1992) 157: 297–300

Prochlorococcus marinus **nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and *b***

Sallie W. Chisholm[1], Sheila L. Frankel[1], Ralf Goericke[2], Robert J. Olson[2], Brian Palenik[1,2,3], John B. Waterbury[2], Lisa West-Johnsrud[1], and Erik R. Zettler[2]



MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Mar. 1999, p. 106–127

Prochlorococcus, a Marine Photosynthetic Prokaryote of Global Significance

F. PARTENSKY,[1]* W. R. HESS,[2] AND D. VAULOT[1]

Station Biologique, CNRS, INSU et Université Pierre et Marie Curie, F-29680 Roscoff, France,[1] and Humboldt-Universität Berlin, Institut für Biologie/Genetik, D-10115 Berlin, Germany[2]

NATURE | VOL 424 | 14 AUGUST 2003

Bacterial photosynthesis genes in a virus

A bacteriophage may protect itself and its host against a deadly effect of bright sunlight.

NATURE | Vol 461 | 10 September 2009

Photosystem I gene cassettes are present in marine virus genomes

Itai Sharon[1,2]*, Ariella Alperovitch[1]*, Forest Rohwer[4,5], Matthew Haynes[4], Fabian Glaser[3], Nof Atamna-Ismaeel[1], Ron Y. Pinter[2], Frédéric Partensky[6], Eugene V. Koonin[7], Yuri I. Wolf[7], Nathan Nelson[8] & Oded Béjà[1]

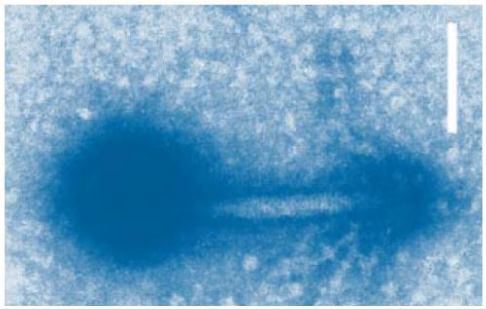
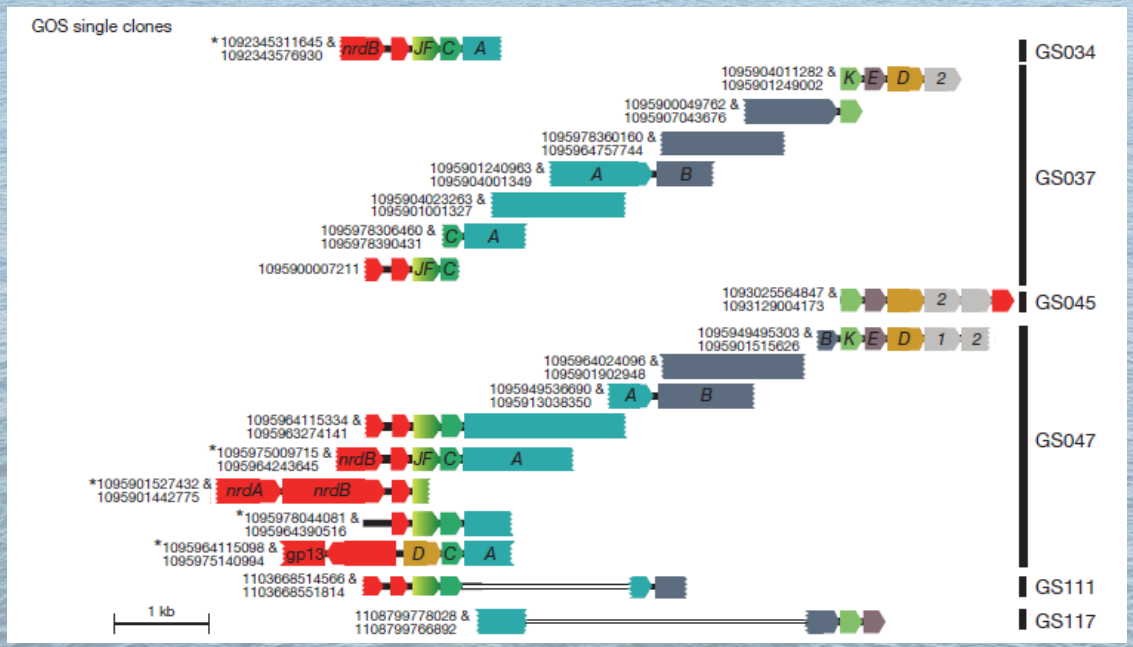


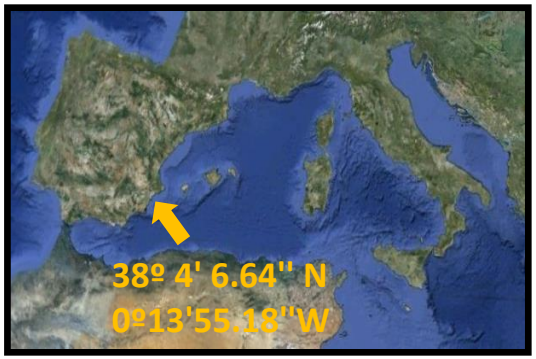
Figure 1 The bacteriophage S-PM2 (here artificially coloured blue), which infects marine cyanobacteria. Scale bar, 100 nm.



The ISME Journal (2010), 1–13

Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing

Rohit Ghai[1], Ana-Belén Martin-Cuadrado[1], Aitor Gonzaga Molto[1], Inmaculada García Heredia[1], Raúl Cabrera[2], Javier Martin[3], Miguel Verdú[3], Philippe Deschamps[4], David Moreira[4], Purificación López-García[4], Alex Mira[2] and Francisco Rodriguez-Valera[1]



SAMPLE

200L of seawater

20 µm

5µm → Eukaryote

0.22 µm → Prokaryote*

3-5 µg of DNA

→ **METAGENOME**

- **Direct sequencing (454, Illumina)**
- **Fosmid library (batch sequencing of fosmids by 454, Illumina)**

Tangential flow filtration

TFF (30kD) → Viruses

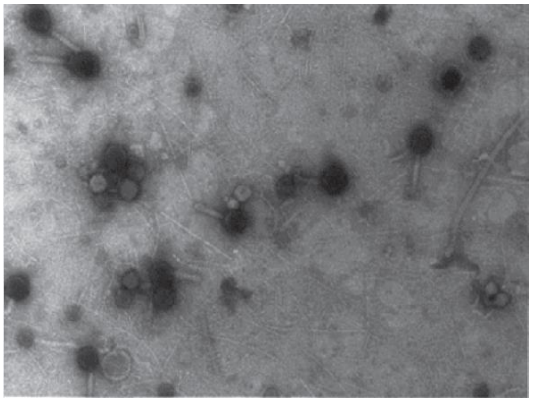
1 µg of DNA

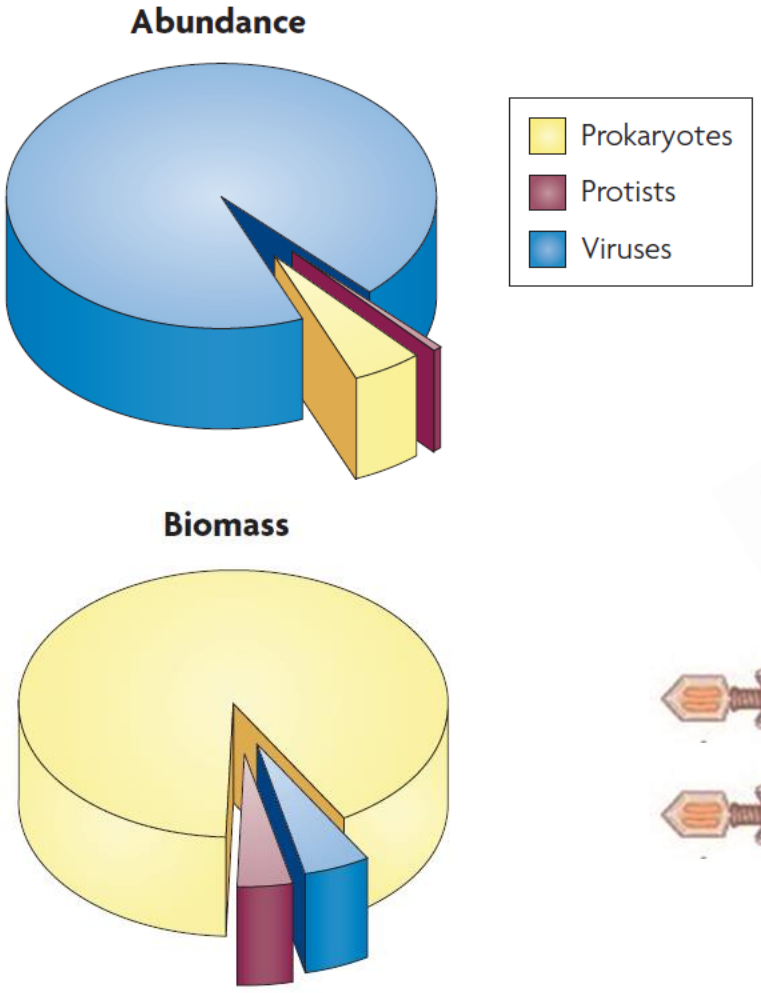
→ **METAVIROME**

NATURE REVIEWS | MICROBIOLOGY VOLUME 5 | OCTOBER 2007 | 801

Marine viruses — major players in the global ecosystem

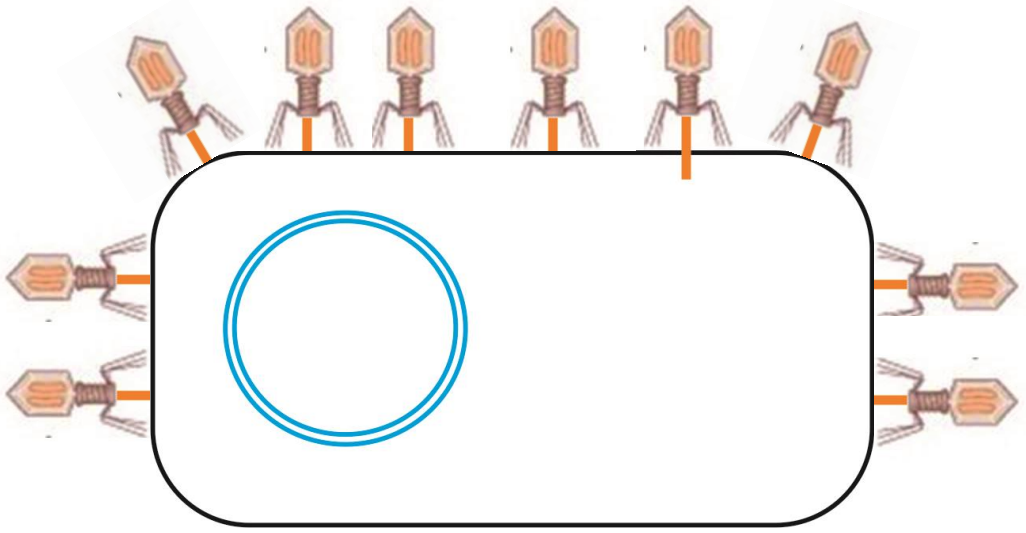
Curtis A. Suttle





phage genome: 50 kb

phage genome: 500 kb = 0.5 Mb



bacterial genome: 5 Mb

Huge Sampling effort

Virology 434 (2012) 181–186

Ocean viruses: Rigorously evaluating the metagenomic sample-to-sequence pipeline

Melissa B. Duhaime [a], Matthew B. Sullivan [b,*]

In a study designed to quantitatively evaluate biases of TFF and $FeCl_3$-precipitation concentration methods, DNA was extracted from 1080 L of viruses purified from the viral-fraction (particles < 0.2 µm) of a large-scale ocean analog, the Biosphere2 Ocean, and used to generate triplicate metagenomes from each concentration treatment. Analyses of these data showed that TFF

PLOS ONE April 2013 | Volume 8 | Issue 4 | e60604

Assembly of a Marine Viral Metagenome after Physical Fractionation

Jennifer R. Brum[¤], Alexander I. Culley, Grieg F. Steward*

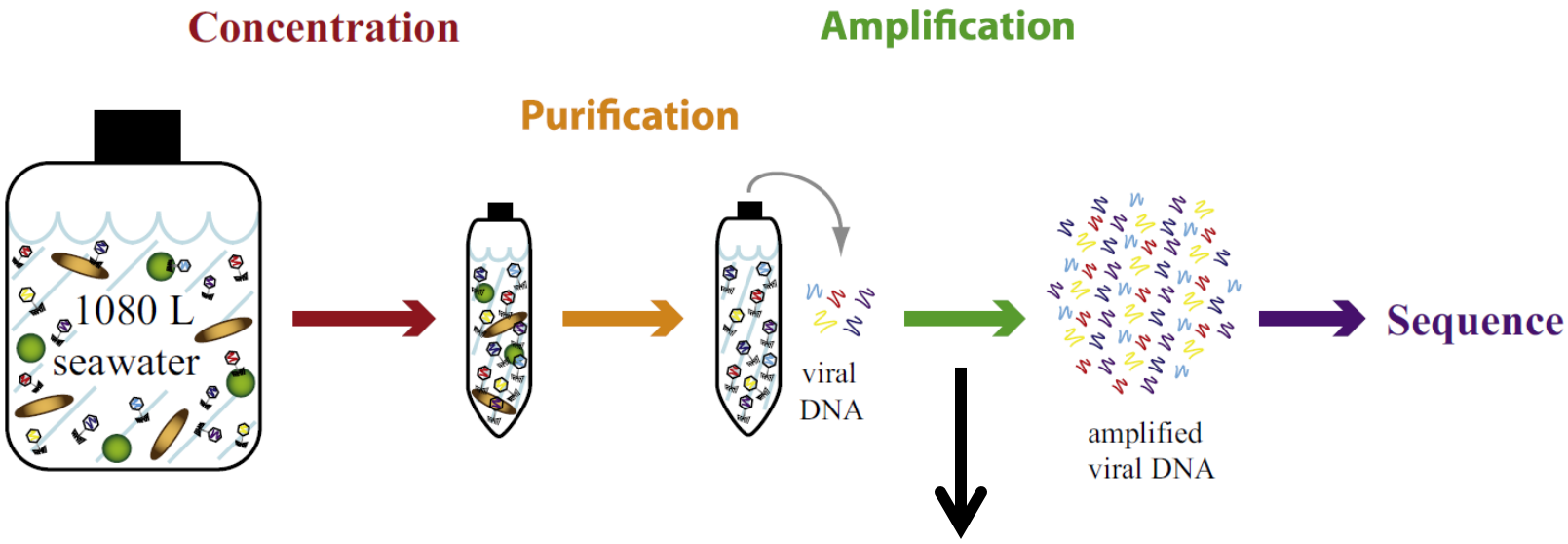
Sample Collection

A viral concentrate was collected on October 17, 2006 from a depth of 3 m approximately 25 m off the southeast shore of Coconut Island (Moku O Lo'e) located in Kāne'ohe Bay, Oahu, HI. Approximately 1800 l of water was filtered through 0.2 µm pore-size cartridge filters with polyethersulfone membranes (Polycap, Whatman). Viruses in the filtrate were concentrated with a tangential flow filtration cassette with 100 kDa nominal molecular weight cut-off (NMWCO) regenerated cellulose membrane (Pellicon 2, Millipore). The concentrate was stored at 4°C

Virology 434 (2012) 181–186

Ocean viruses: Rigorously evaluating the metagenomic sample-to-sequence pipeline

Melissa B. Duhaime [a], Matthew B. Sullivan [b,*]



MDA: bias towards ssDNA
LA: apparently less biased

PLOS ONE April 2013 | Volume 8 | Issue 4 | e60604

Assembly of a Marine Viral Metagenome after Physical Fractionation

Jennifer R. Brum[¤], Alexander I. Culley, Grieg F. Steward*

Filtered 1800 L water

↓

Viral Fractionation

↓

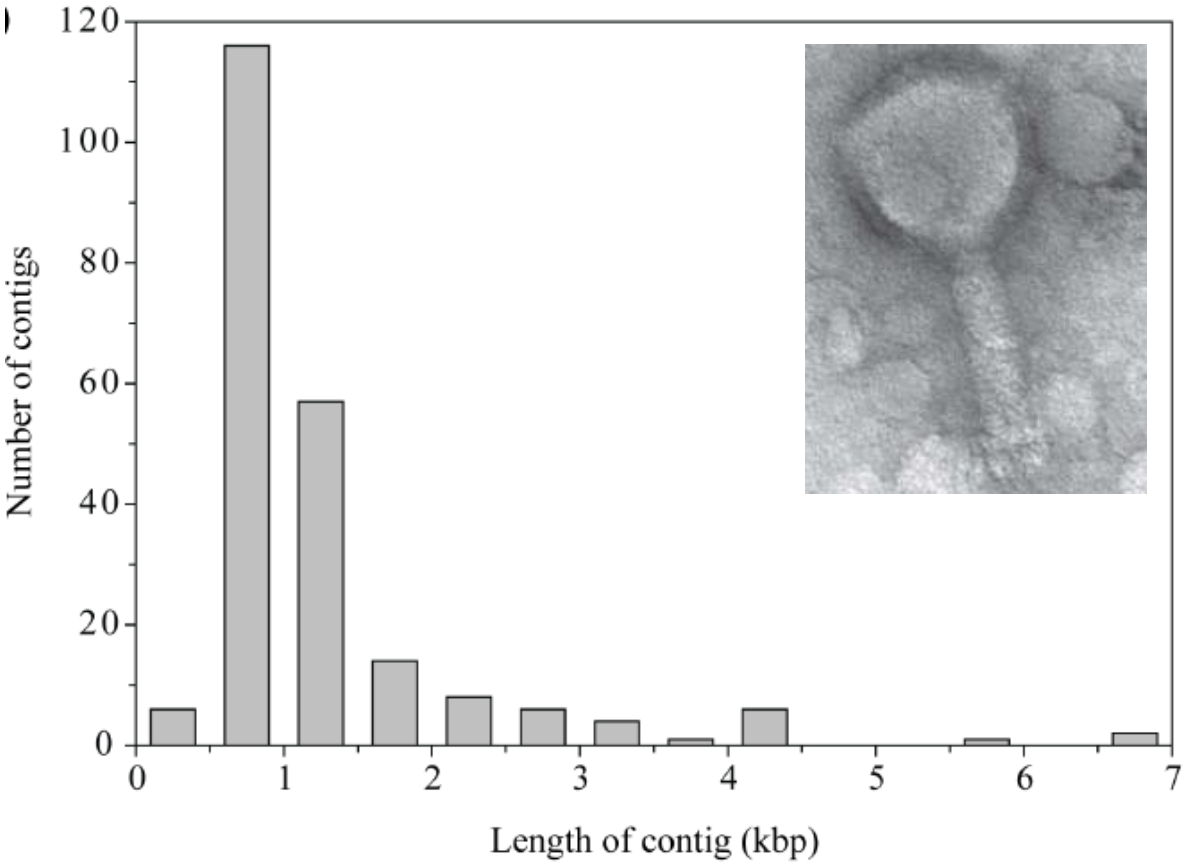
MDA

↓

Cloned libraty in plasmids

↓

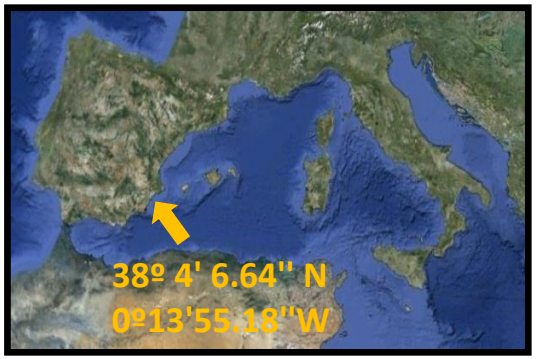
Sanger Sequenced



The ISME Journal (2010), 1–13

Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing

Rohit Ghai[1], Ana-Belén Martin-Cuadrado[1], Aitor Gonzaga Molto[1], Inmaculada García Heredia[1], Raúl Cabrera[2], Javier Martin[3], Miguel Verdú[3], Philippe Deschamps[4], David Moreira[4], Purificación López-García[4], Alex Mira[2] and Francisco Rodriguez-Valera[1]



SAMPLE

200L of seawater

20 µm

5µm → Eukaryote

0.22 µm → Prokaryote*

3-5 µg of DNA

Tangential flow filtration

TFF (30kD) → Viruses

MDA

1 µg of DNA

METAGENOME

- Direct sequencing (454, Illumina)
- Fosmid library (batch sequencing of fosmids by 454, Illumina)

METAVIROME

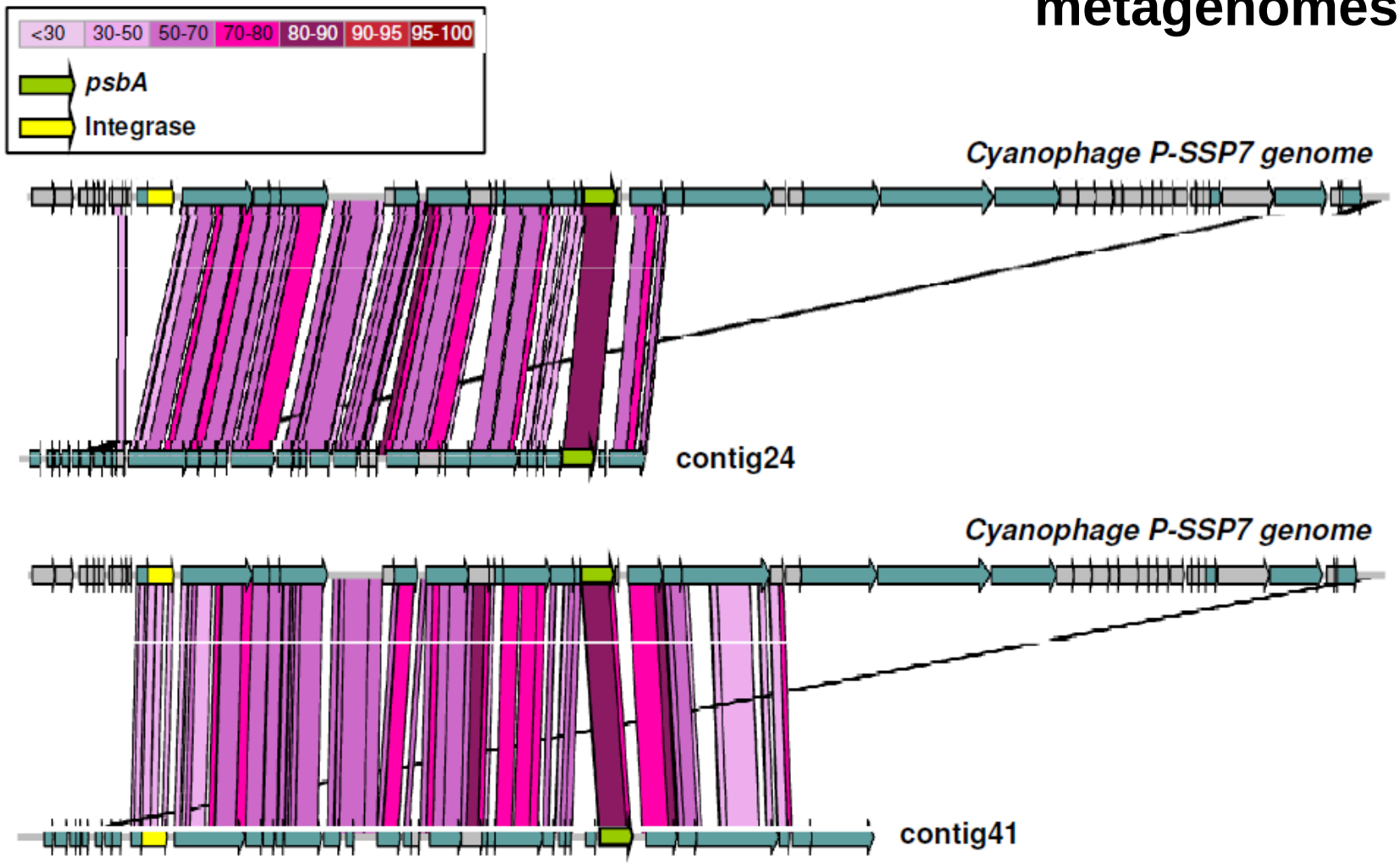
1/3 lane Illumina

The ISME Journal (2010), 1–13

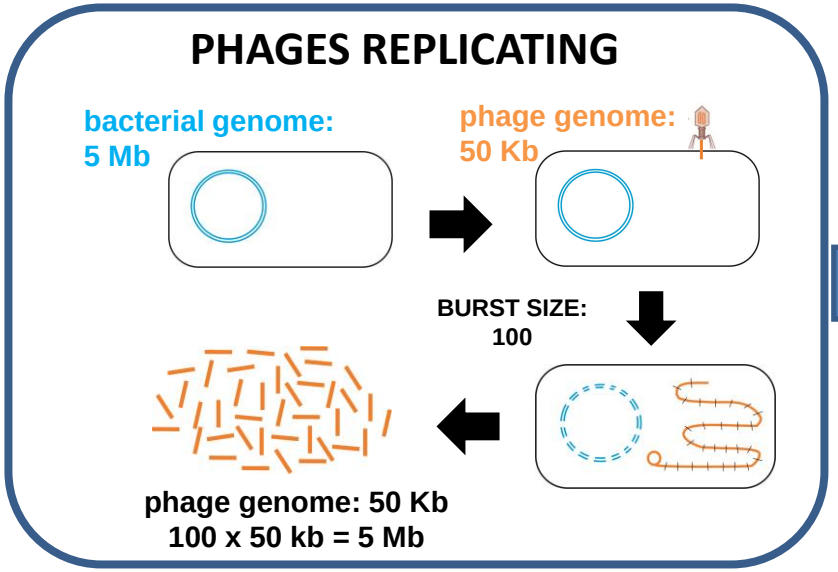
Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing

Rohit Ghai[1], Ana-Belén Martin-Cuadrado[1], Aitor Gonzaga Molto[1], Inmaculada García Heredia[1], Raúl Cabrera[2], Javier Martin[3], Miguel Verdú[3], Philippe Deschamps[4], David Moreira[4], Purificación López-García[4], Alex Mira[2] and Francisco Rodriguez-Valera[1]

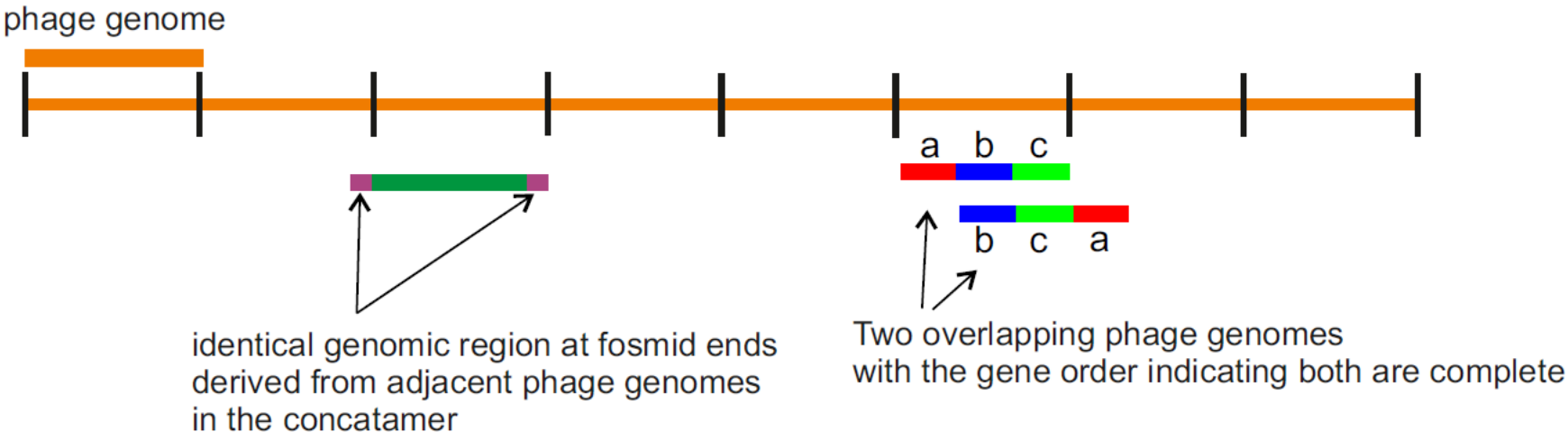
10-20% VIRAL CONTIGS in fosmids from cellular metagenomes



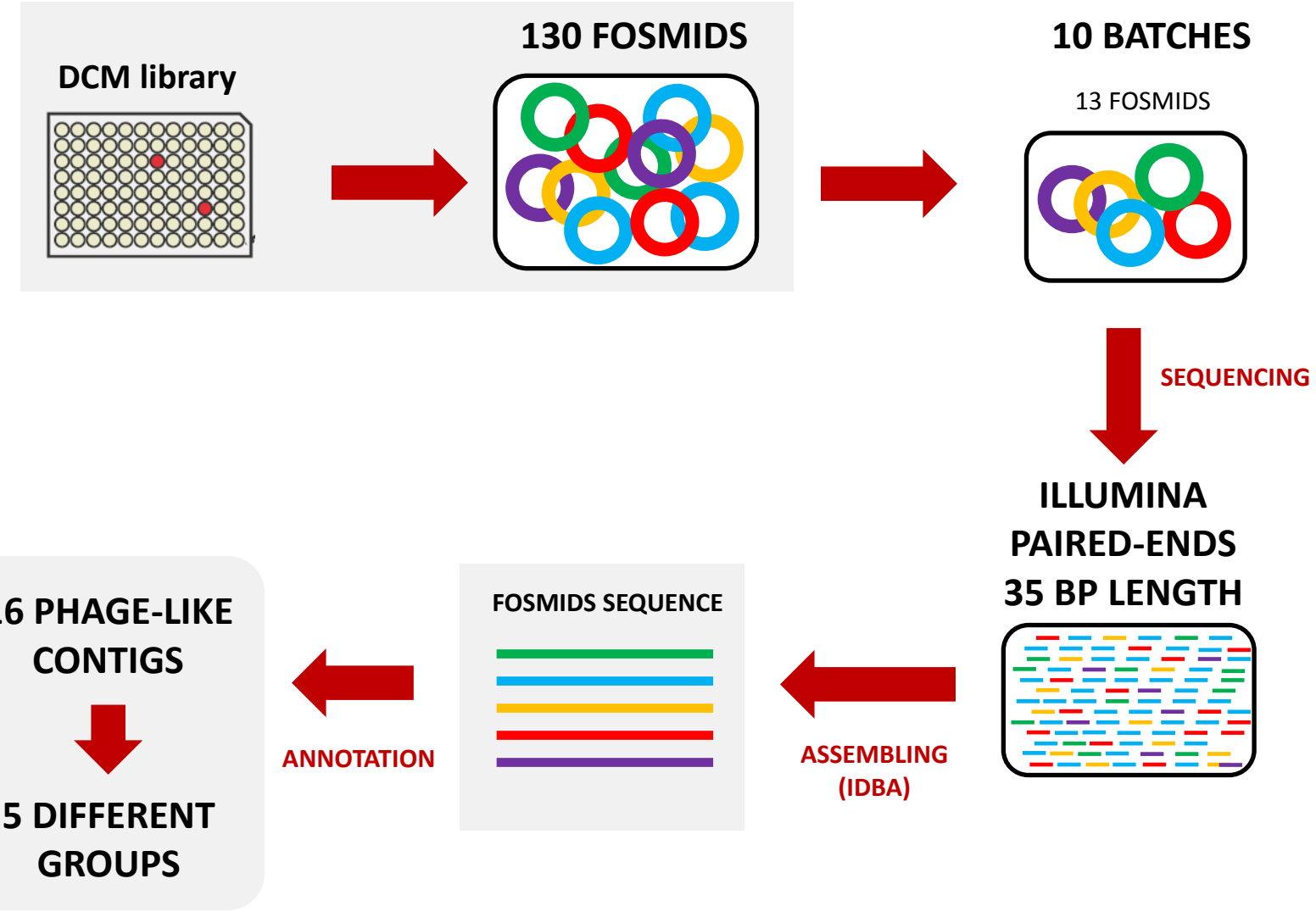
Phages from metagenomes



Natural amplification of phage DNA and the bias in cloning



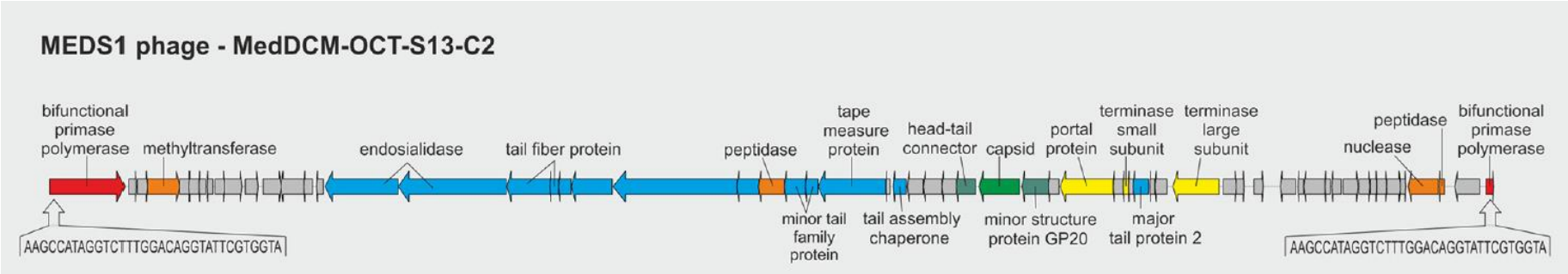
Using cellular metagenomes to retrieve viral genomes

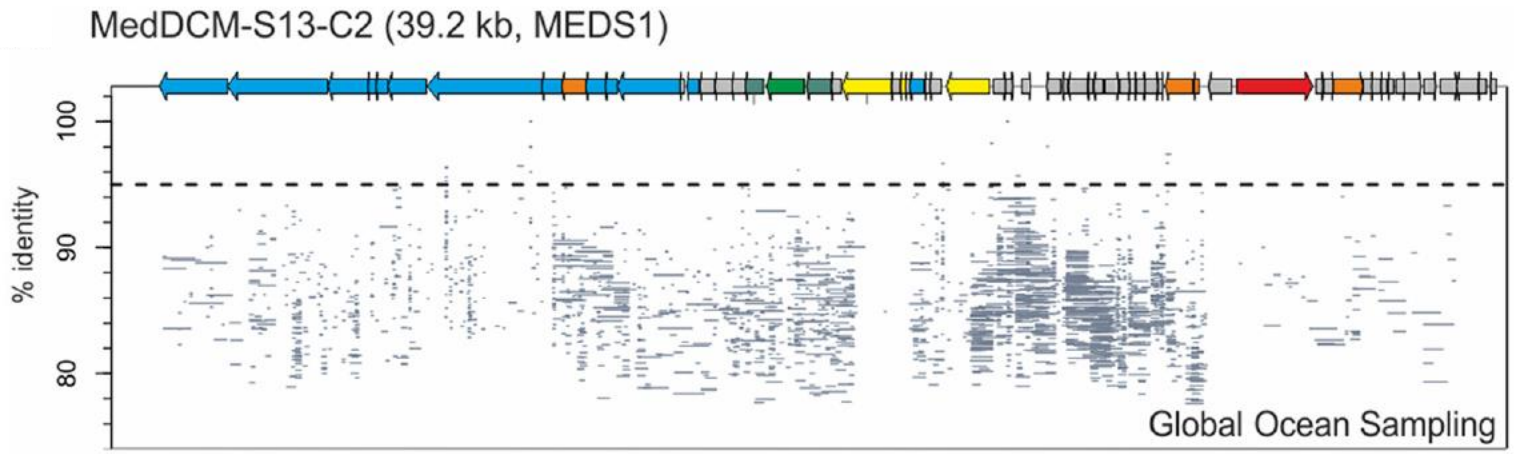


January 2013 Volume 79 Number 2
Applied and Environmental Microbiology p. 688–695

Reconstruction of Novel Cyanobacterial Siphovirus Genomes from Mediterranean Metagenomic Fosmids

Carolina Megumi Mizuno, Francisco Rodriguez-Valera, Inmaculada Garcia-Heredia, Ana-Belen Martin-Cuadrado, Rohit Ghai



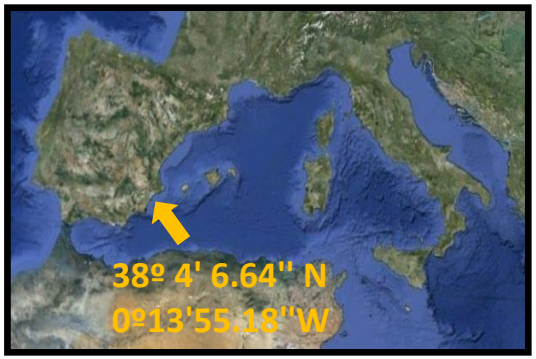


Possible to recover, complete, and naturally abundant phages

The ISME Journal (2010), 1–13

Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing

Rohit Ghai[1], Ana-Belén Martin-Cuadrado[1], Aitor Gonzaga Molto[1], Inmaculada García Heredia[1], Raúl Cabrera[2], Javier Martin[3], Miguel Verdú[3], Philippe Deschamps[4], David Moreira[4], Purificación López-García[4], Alex Mira[2] and Francisco Rodriguez-Valera[1]



SAMPLE

200L of seawater

20 µm

5µm → Eukaryote

0.22 µm → Prokaryote*

3-5 µg of DNA

Tangential flow filtration

TFF (30kD) → Viruses

MDA

1 µg of DNA

METAGENOME

- Direct sequencing (454, Illumina)
- Fosmid library (batch sequencing of fosmids by 454, Illumina)

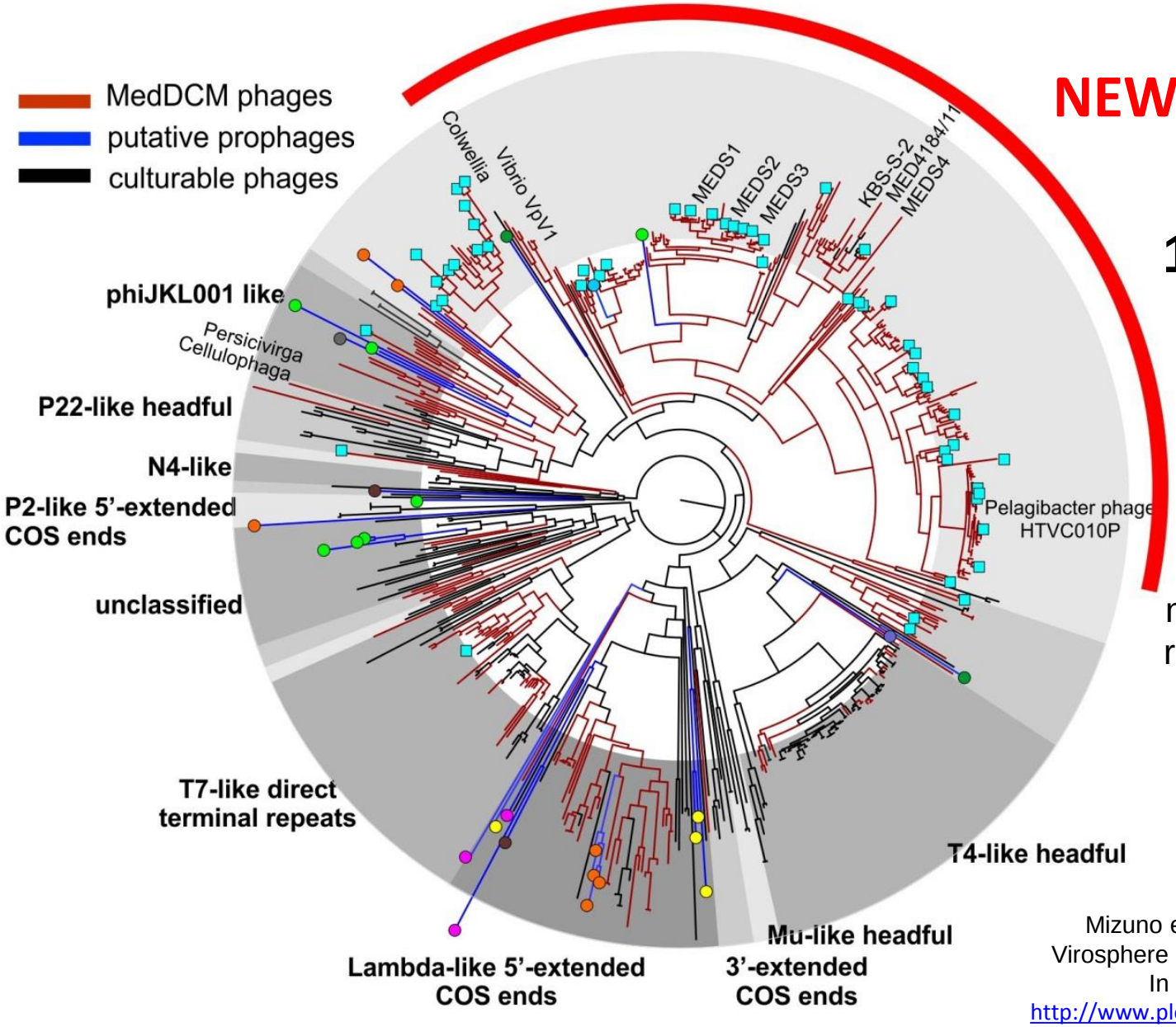
7000 FOSMIDS SEQUENCED

1 lane Illumina

METAVIROME

1/3 lane Illumina

New uncultured phages genomes from metagenomes

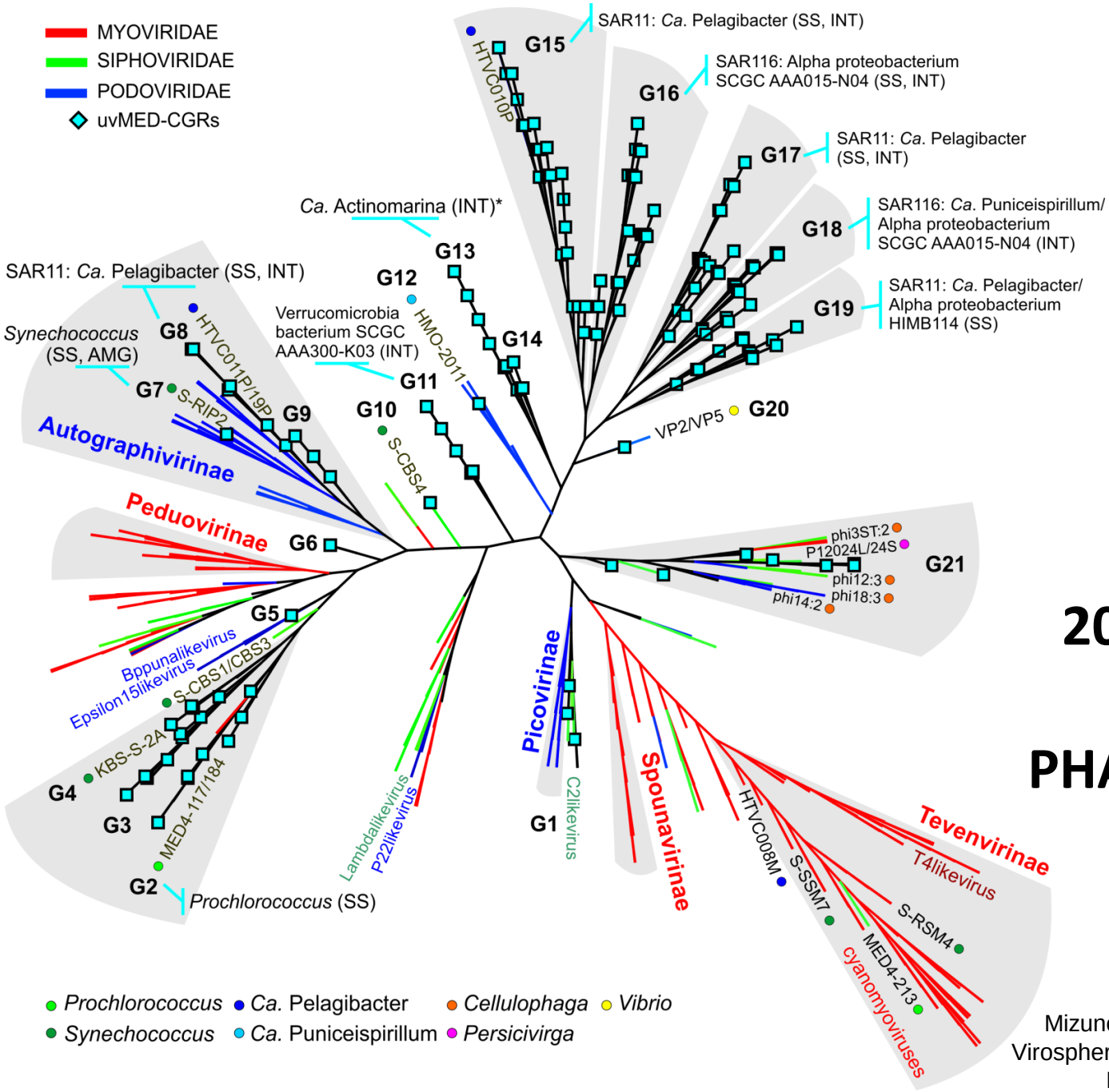


NEW TERMINASES

MORE THAN 1000 virus-like contigs!

Phage terminases recognize DNA for packaging and have a nuclease activity that is responsible for creating the ends of the virion chromosome

Mizuno et al., 2013. Expanding the Marine Virosphere Using Metagenomics. Plos Genetics. In press. December, 12th 2013
http://www.plosgenetics.org/doi/pgen.1003987



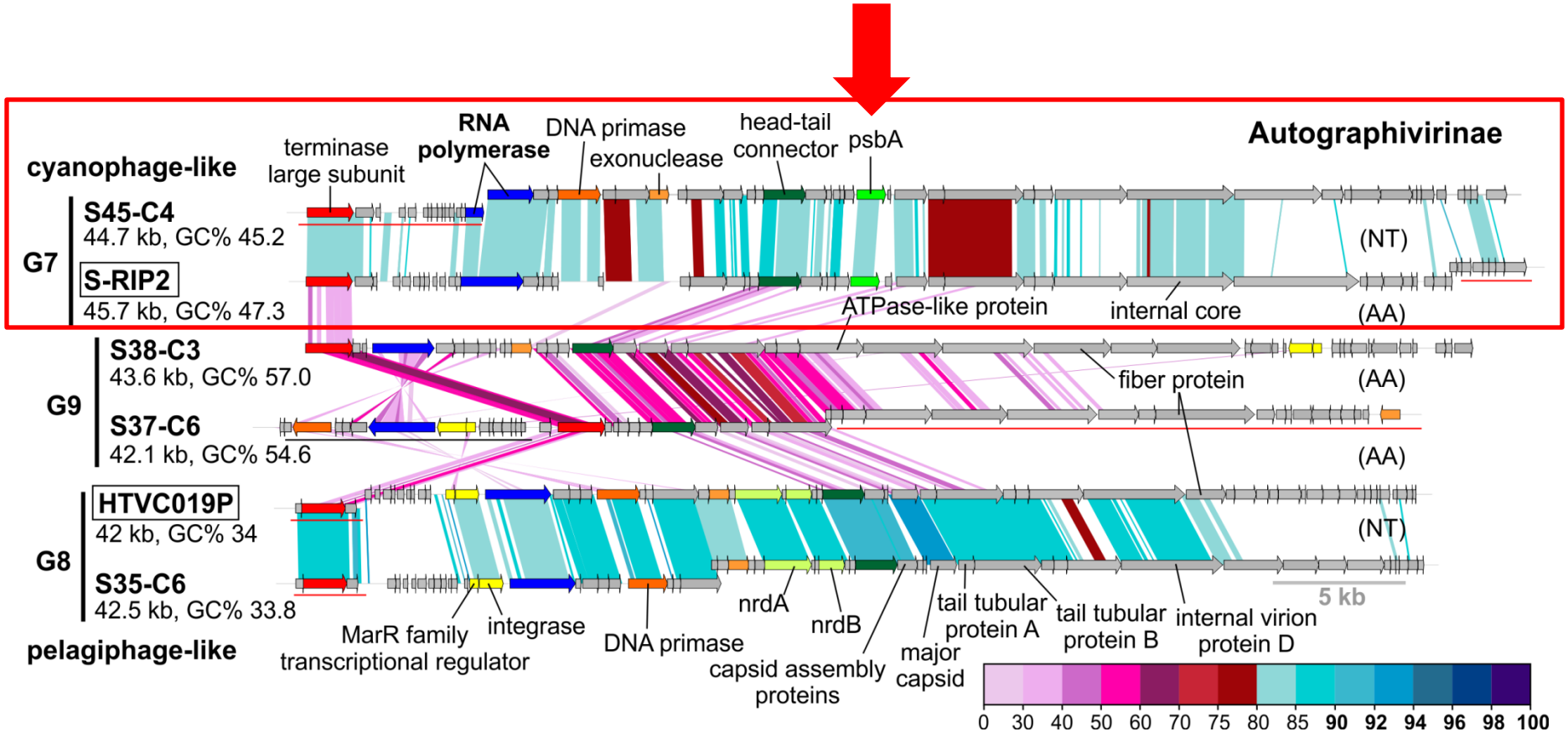
208 COMPLETE NOVEL PHAGE GENOMES

Mizuno et al., 2013. Expanding the Marine Virosphere Using Metagenomics. Plos Genetics. In press. December, 12th 2013
http://www.plosgenetics.org/doi/pgen.1003987

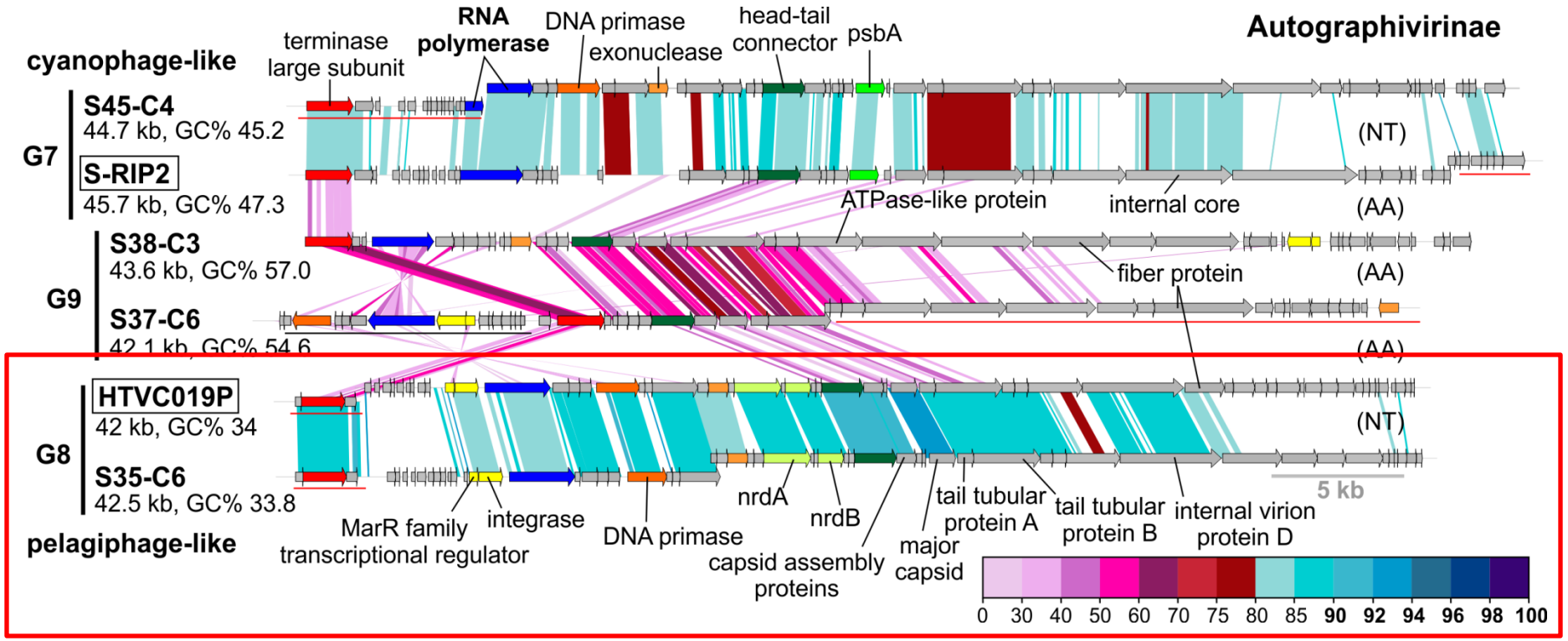
Host???

- Sequence similarity
- Characteristic metabolic genes (psbA for cyanophages)
- CRISPR (spacer matches a phage)
- Integrase/att site

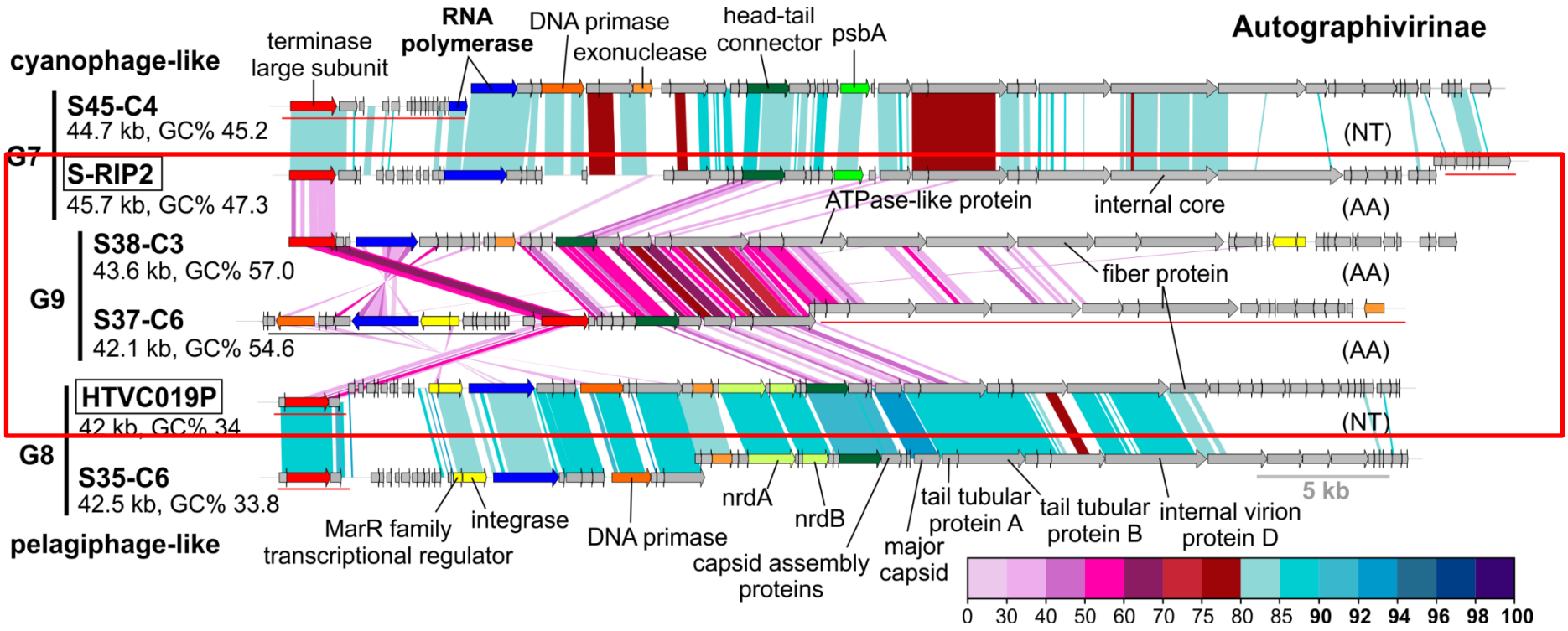
New cyanophages and pelagiphages



New cyanophages and pelagiphages



New cyanophages and pelagiphages



Host?

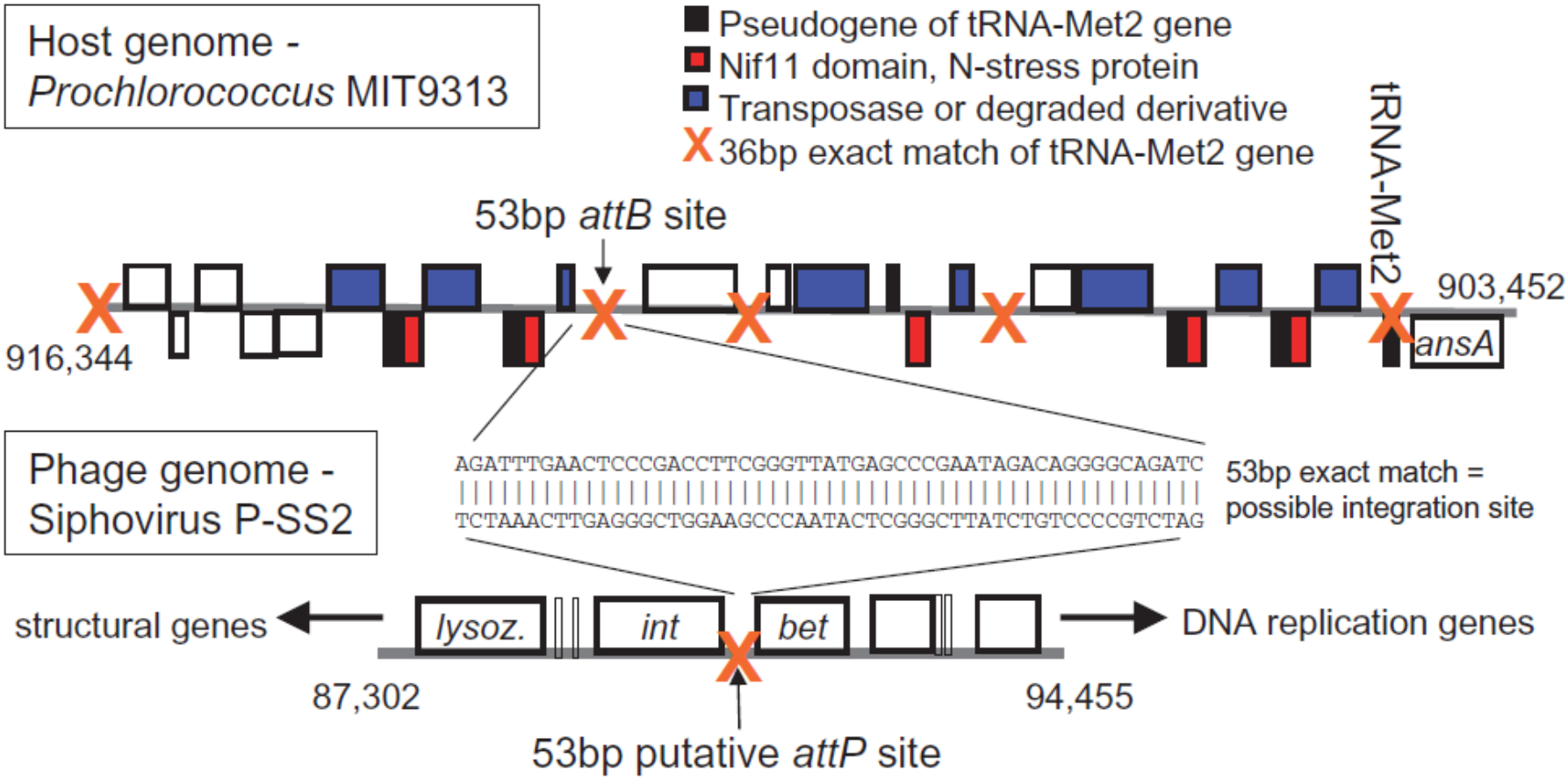
- Sequence similarity
- Characteristic metabolic genes (psbA for cyanophages)
- CRISPR (spacer matches a phage)
- Integrase/att site

Environmental Microbiology (2009) 11(11), 2935–2951

The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'

Matthew B. Sullivan,[1*†] Bryan Krastins,[2]
Jennifer L. Hughes,[3] Libusha Kelly,[1] Michael Chase,[2]
David Sarracino[2] and Sallie W. Chisholm[1*]

Integrase/att site



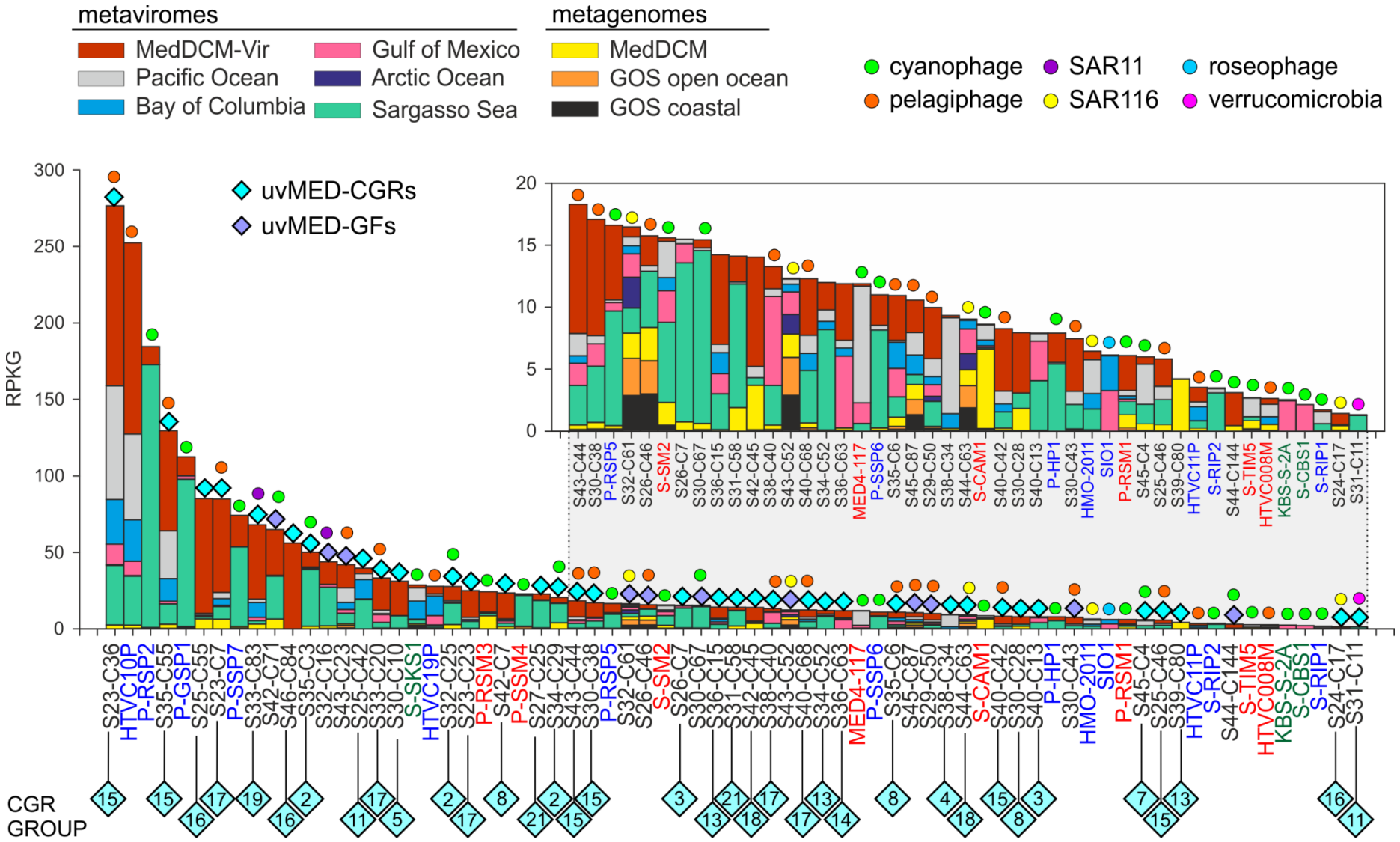
Unculture phage genomes of uncultured bacteria

CGR group identifier	No. of CGRs in group	Average GC%	Average length (bp)	Putative host indicated for members of this group	Evidence
G1	3	35.9	37028	-	-
G2	11	36.7	38586	Cyanobacteria	*Prochlorococcus* (SS)
G3	11	49.0	41598	-	-
G4	5	49.2	40325	-	-
G5	1	35.8	40289	-	-
G6	2	54.5	34532	-	-
G7	2	46.0	44075	Cyanobacteria	*Synechococcus* (SS, AMG)
G8	7	39.6	40595	SAR11 cluster	*Ca.* Pelagibacter (SS, INT)
G9	4	54.9	42841	-	-
G10	1	53.0	43957	-	-
G11	7	45.2	38903	Verrucomicrobia	Verrucomicrobia bacterium SCGC AAA300-K03 (INT)
G12	1	56.0	41524	-	-
G13*	10	35.4	35410	Actinobacteria	*Ca.* Actinomarina minuta (INT)
G14	5	34.4	35580	-	-
G15	32	35.1	36262	SAR11 cluster	*Ca.* Pelagibacter (SS, INT)
G16	29	35.2	35537	SAR116 cluster	Alpha proteobacterium SCGC AAA015-N04 (INT)
G17	25	34.5	36042	SAR11 cluster	*Ca.* Pelagibacter (SS, INT)
G18	22	44.2	35814	Alphaproteobacteria	*Ca.* Puniceispirillum/Alpha proteobacterium SCGC AAA015-N04 (INT)
G19	20	31.9	31299	SAR11 cluster	*Ca.* Pelagibacter/Alpha proteobacterium HIMB114 (SS)
G20	1	40.4	42335	-	-
G21	11	34.5	34994	-	-

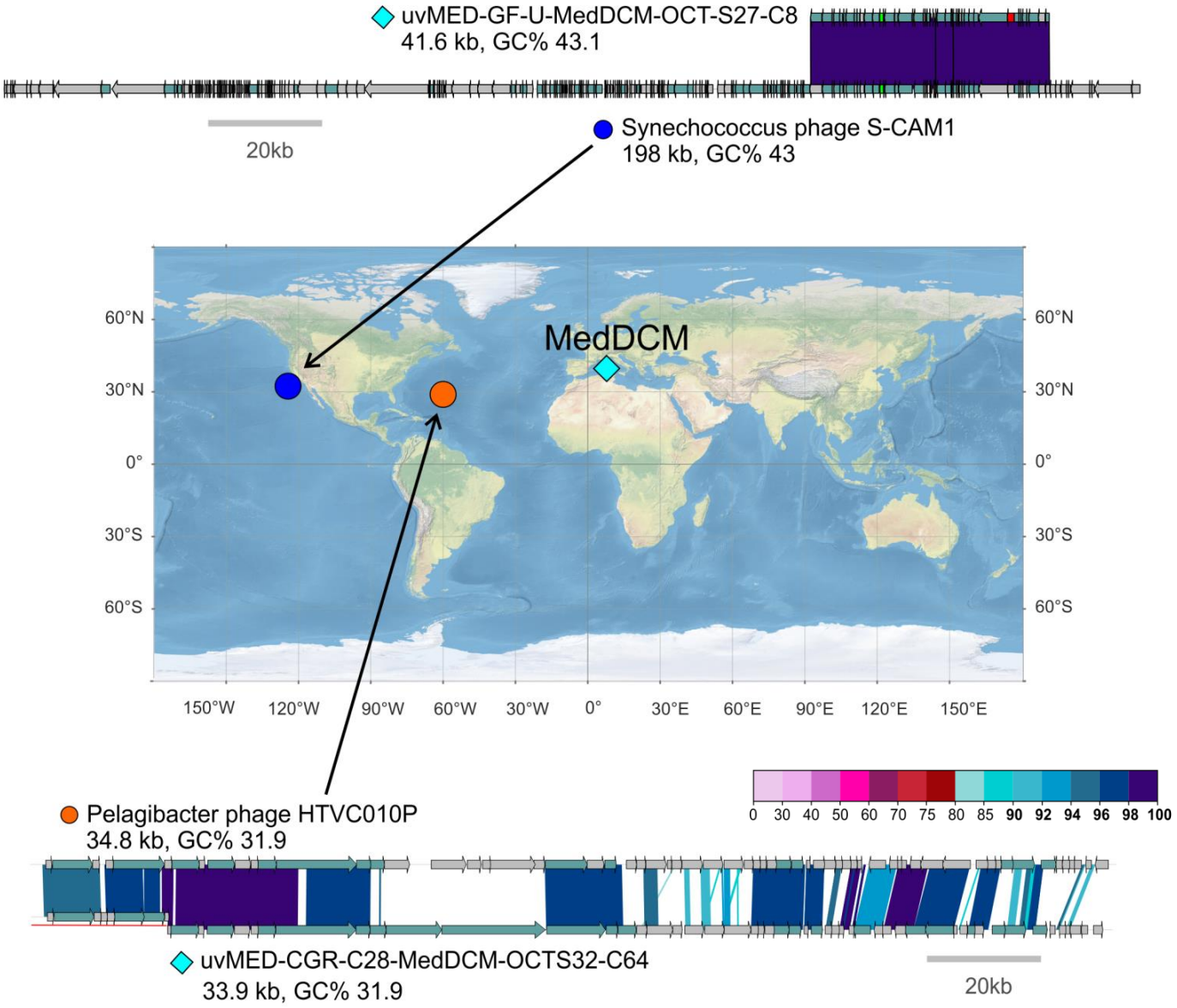
Unculture phage genomes of uncultured bacteria

CGR group identifier	No. of CGRs in group	Average GC%	Average length (bp)	Putative host indicated for members of this group	Evidence
G1	3	35.9	37028	-	-
G2	11	36.7	38586	Cyanobacteria	*Prochlorococcus* (SS)
G3	11	49.0	41598	-	-
G4	5	49.2	40325	-	-
G5	1	35.8	40289	-	-
G6	2	54.5	34532	-	-
G7	2	46.0	44075	Cyanobacteria	*Synechococcus* (SS, AMG)
G8	7	39.6	40595	SAR11 cluster	*Ca.* Pelagibacter (SS, INT)
G9	4	54.9	42841	-	-
G10	1	53.0	43957	-	-
G11	7	45.2	38903	Verrucomicrobia	Verrucomicrobia bacterium SCGC AAA300-K03 (INT)
G12	1	56.0	41524	-	-
G13*	10	35.4	35410	Actinobacteria	*Ca.* Actinomarina minuta (INT)
G14	5	34.4	35580	-	-
G15	32	35.1	36262	SAR11 cluster	*Ca.* Pelagibacter (SS, INT)
G16	29	35.2	35537	SAR116 cluster	Alpha proteobacterium SCGC AAA015-N04 (INT)
G17	25	34.5	36042	SAR11 cluster	*Ca.* Pelagibacter (SS, INT)
G18	22	44.2	35814	Alphaproteobacteria	*Ca.* Puniceispirillum/Alpha proteobacterium SCGC AAA015-N04 (INT)
G19	20	31.9	31299	SAR11 cluster	*Ca.* Pelagibacter/Alpha proteobacterium HIMB114 (SS)
G20	1	40.4	42335	-	-
G21	11	34.5	34994	-	-

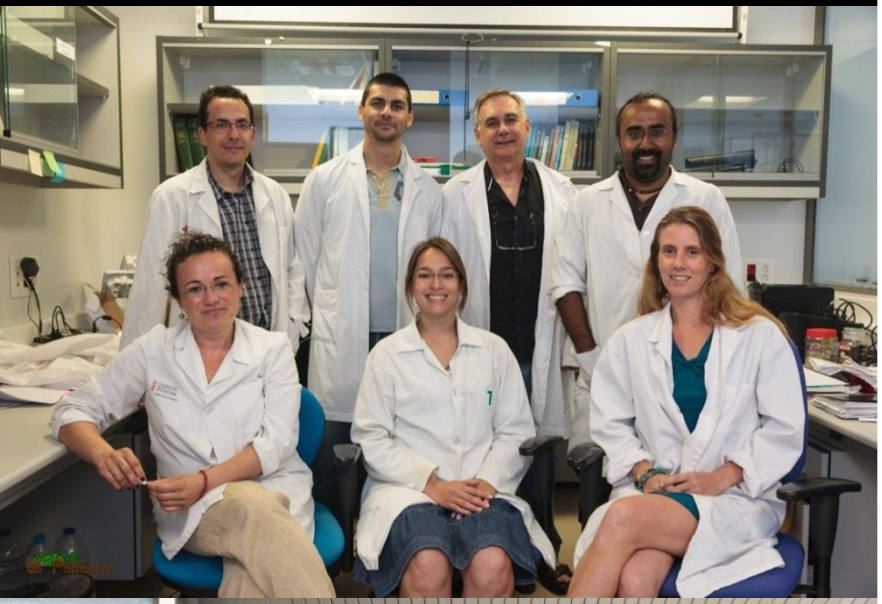
Abundant and widespread phages



The same phage everywhere!!!



Thank you!
Gracias
Obrigada!



Universidad Miguel Hernández

Evolutionary Genomics Group

Group leader:
Prof. Francisco Rodriguez Valera

Members of the group:
Dr. Rohit Ghai
Dr. Ana Belén Martin Cuadrado
Inmaculada García Heredia
Dr. Nikole Kimes
Aitor Molto Gonzaga
Mario Lopez Perez

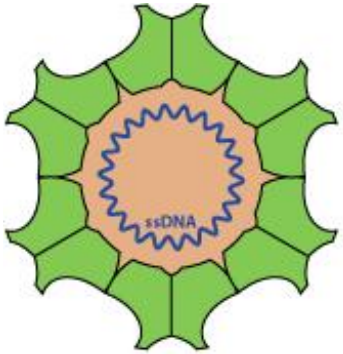
MICROGEN

MAGYK

MaCuMBA

ssDNA viruses

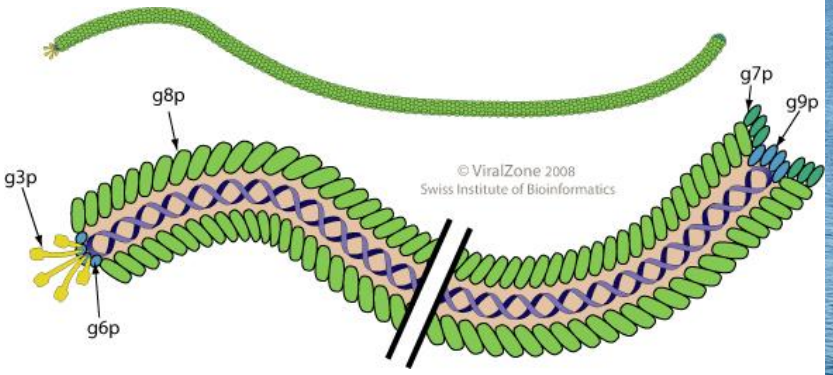
Circoviridae



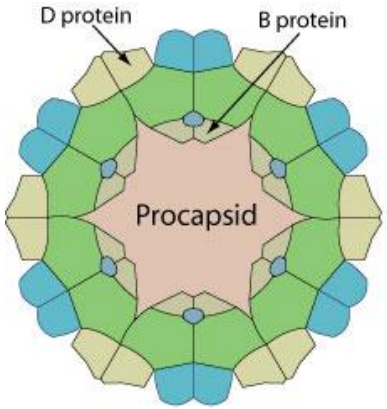
© ViralZone 2008
Swiss Institute of Bioinformatics

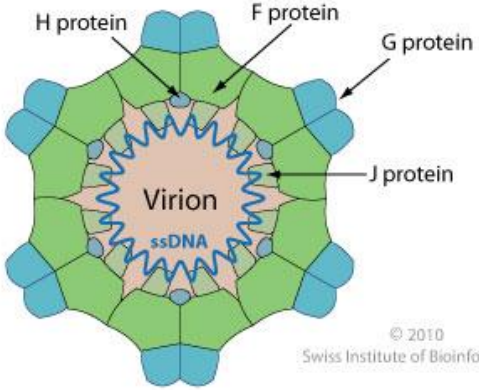
T=1

Inoviridae



Microviridae





© 2010
Swiss Institute of Bioinformatics

T=1

ViralZone