

Diversity, Distribution, and Population Structure of Environmental Viruses

Mya Breitbart

San Diego State University

Phage are the most abundant biological entities on the planet



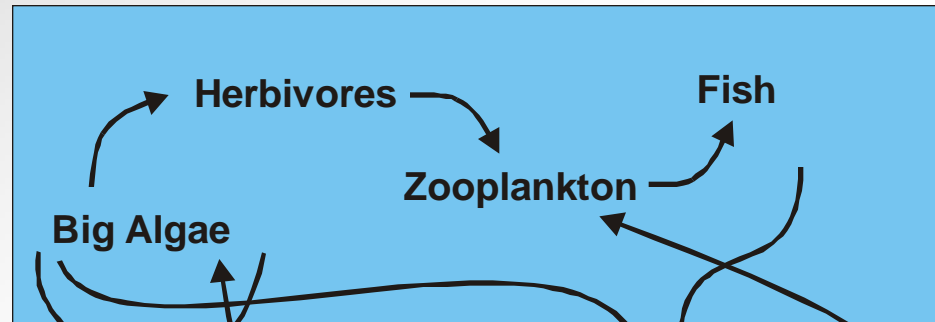
In 1 ml of seawater:

Viruses/Phage	10,000,000
Heterotrophic Bacteria	1,000,000
Photosynthetic Bacteria	100,000
Protozoa	4,000
Algae	3,000
Zooplankton	<<1
Great White Sharks	10^{-19}

Approximately 10^{31} phage in the world

Predators control microbial abundance

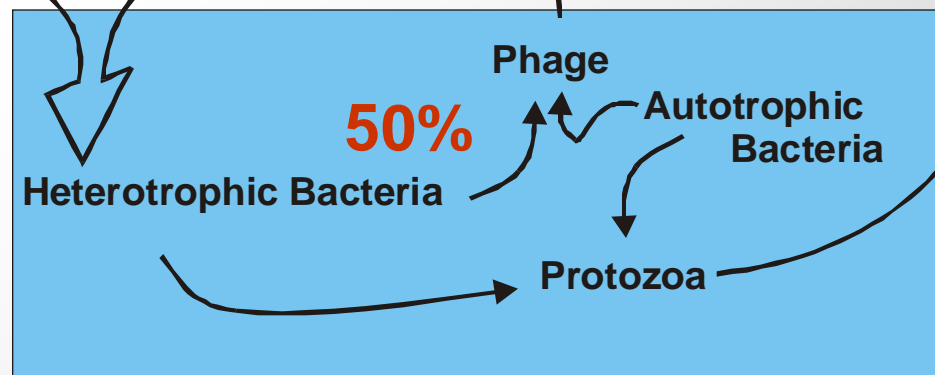
**Classical
Marine
Food Web**



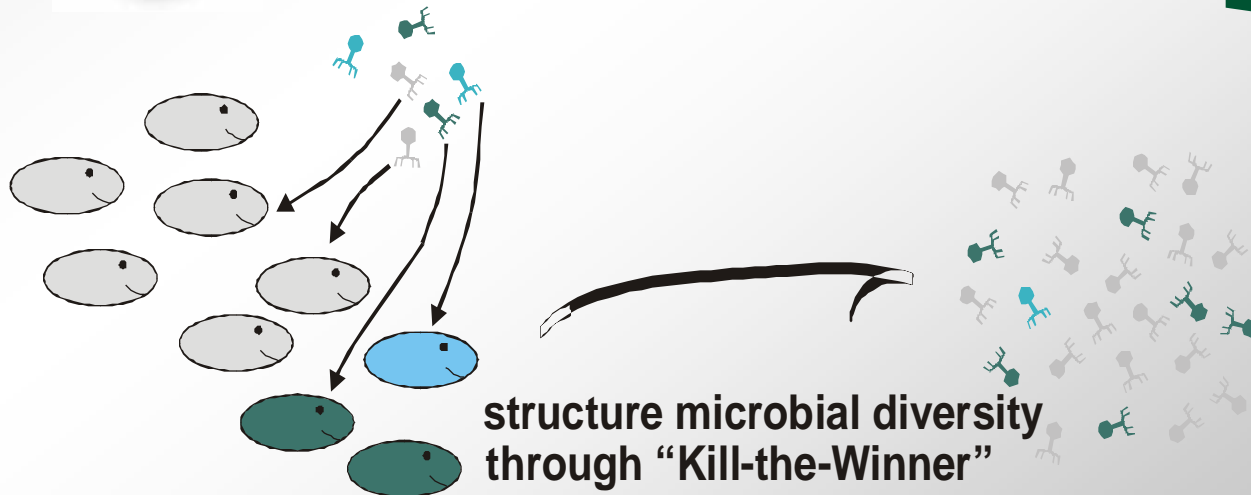
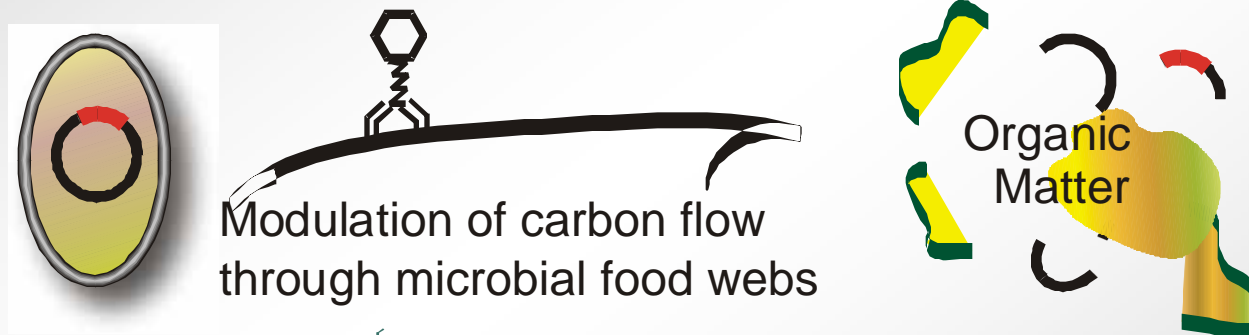
>50%



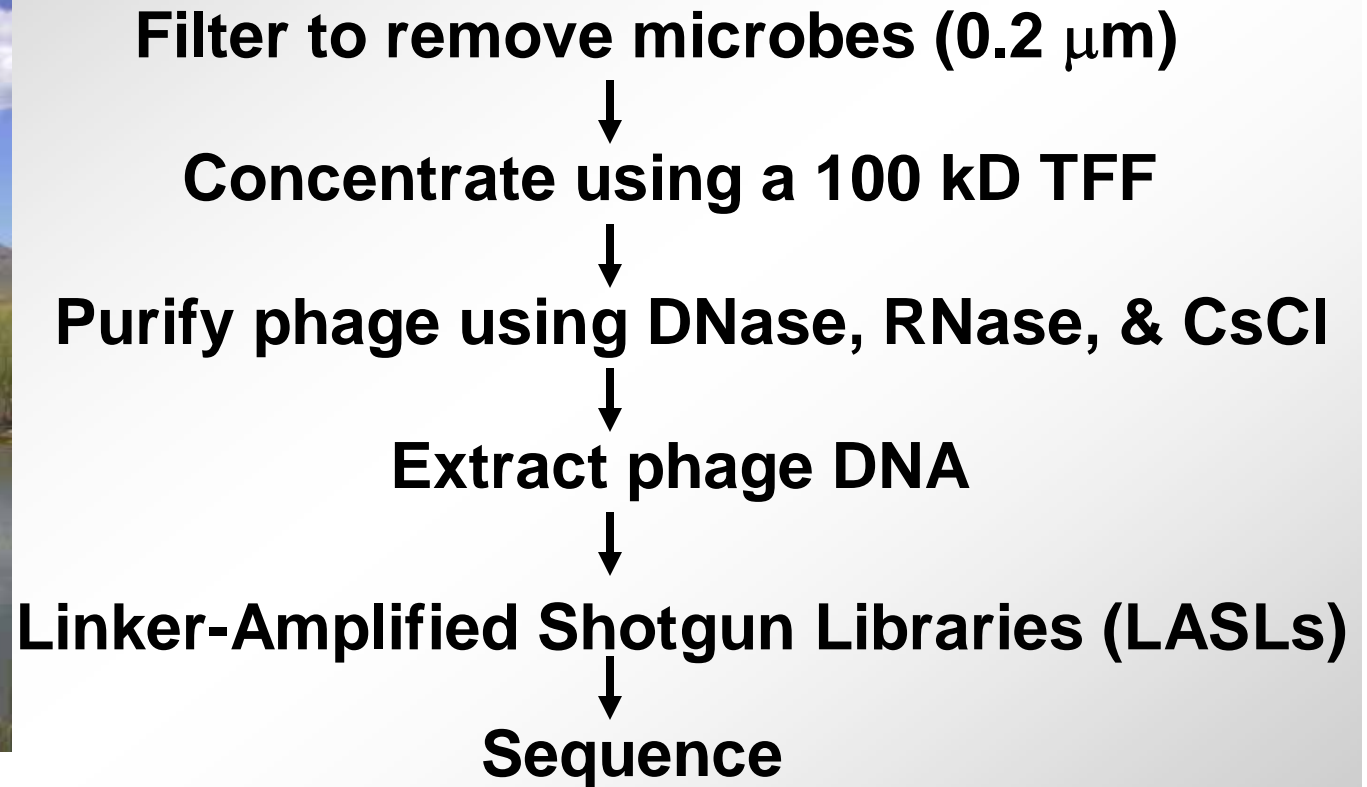
**Marine
Microbial
Food Web**



Ecological roles of phage

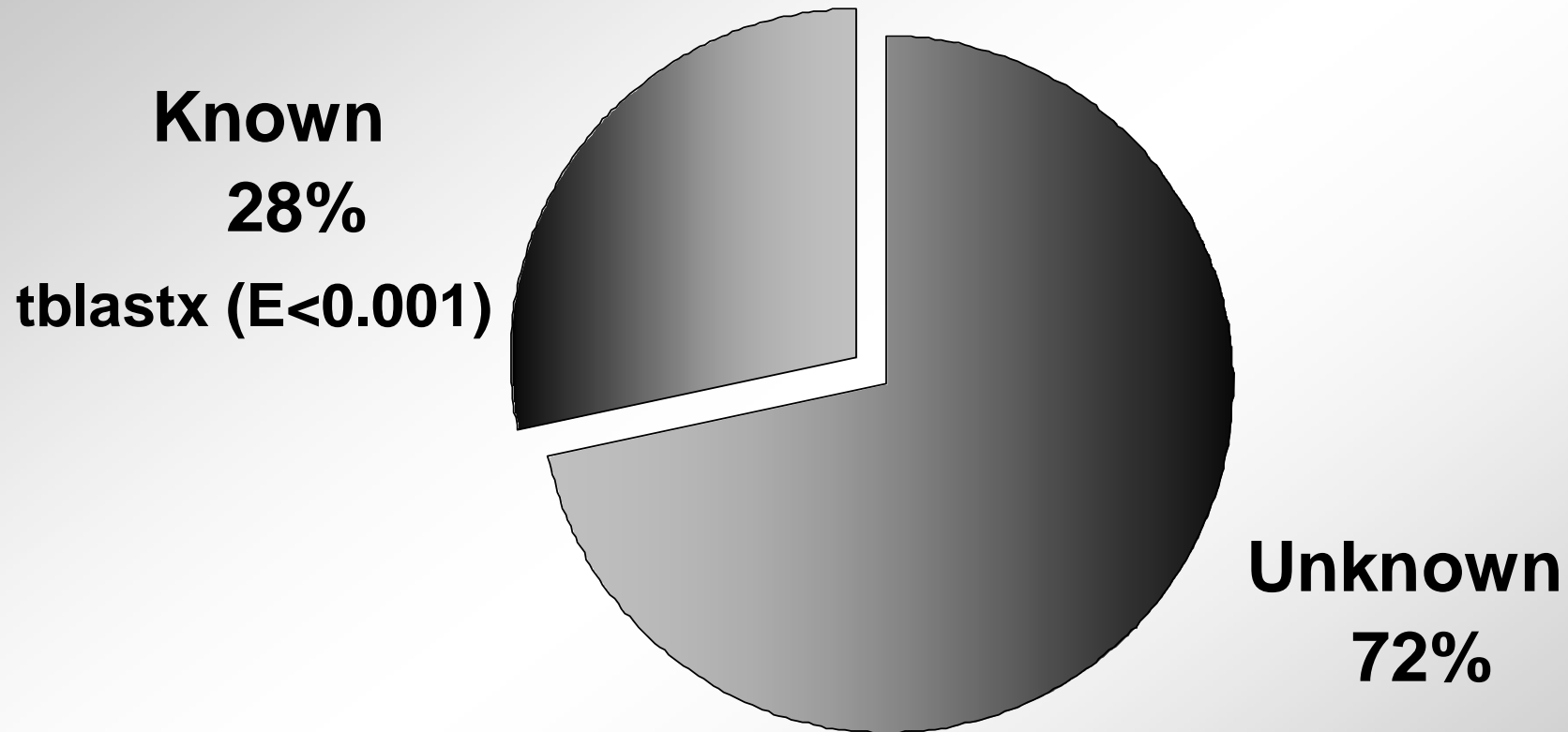


Shotgun libraries of uncultured phage communities



- * Breitbart et al., (2004) Diversity and population structure of a nearshore marine sediment viral community. Proc Royal Society B. 271. 565-574.
- * Breitbart et al., (2003) Metagenomic analyses of an uncultured viral community from human feces. J Bacteriology. 85. 6220-6223.
- * Breitbart et al., (2002) Genomic analysis of uncultured marine viral communities. PNAS. 99. 14250-14255.

Most uncultured phage sequences are "Unknowns"

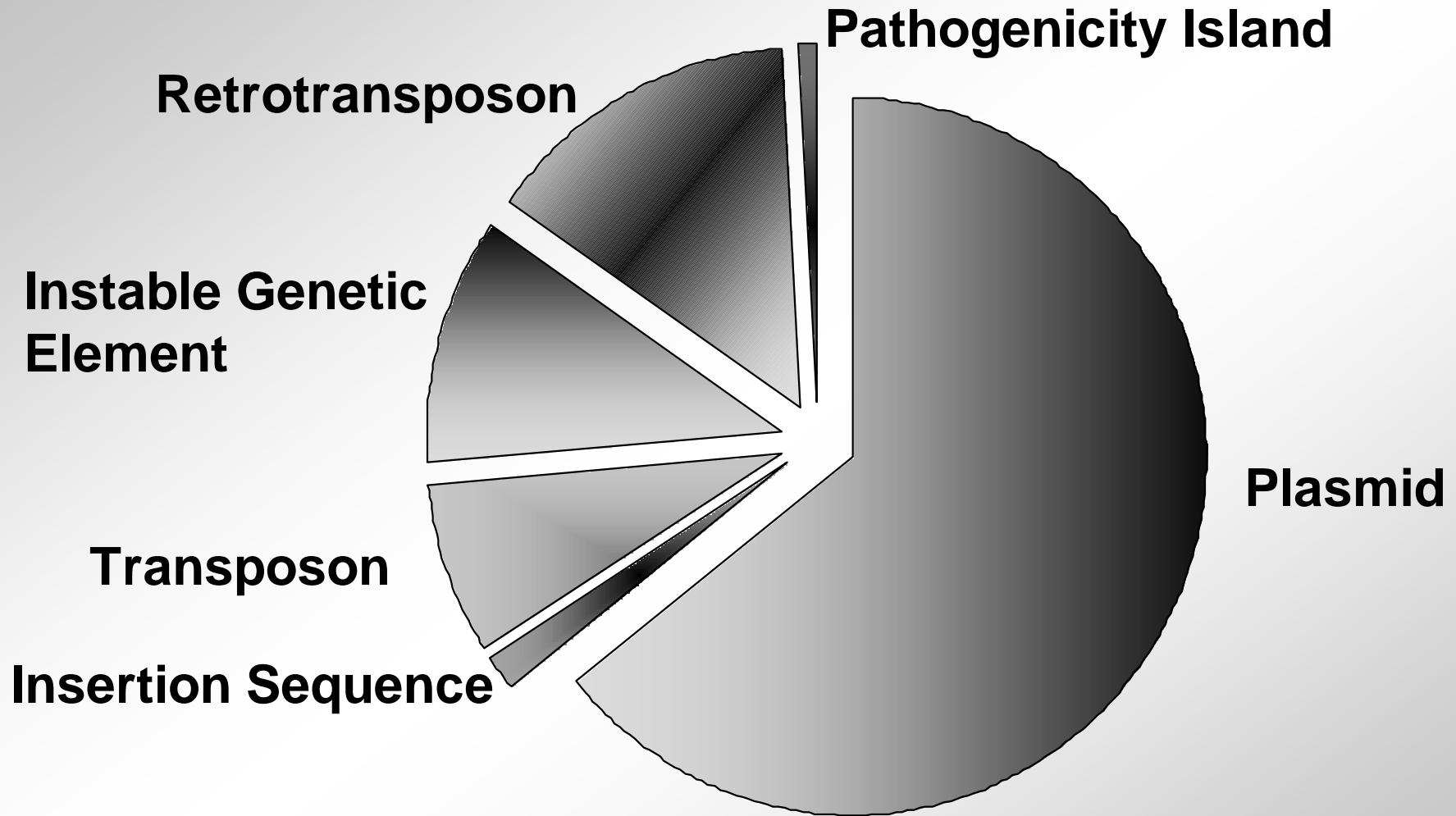


<0.0002% of the global phage metagenome has been sampled

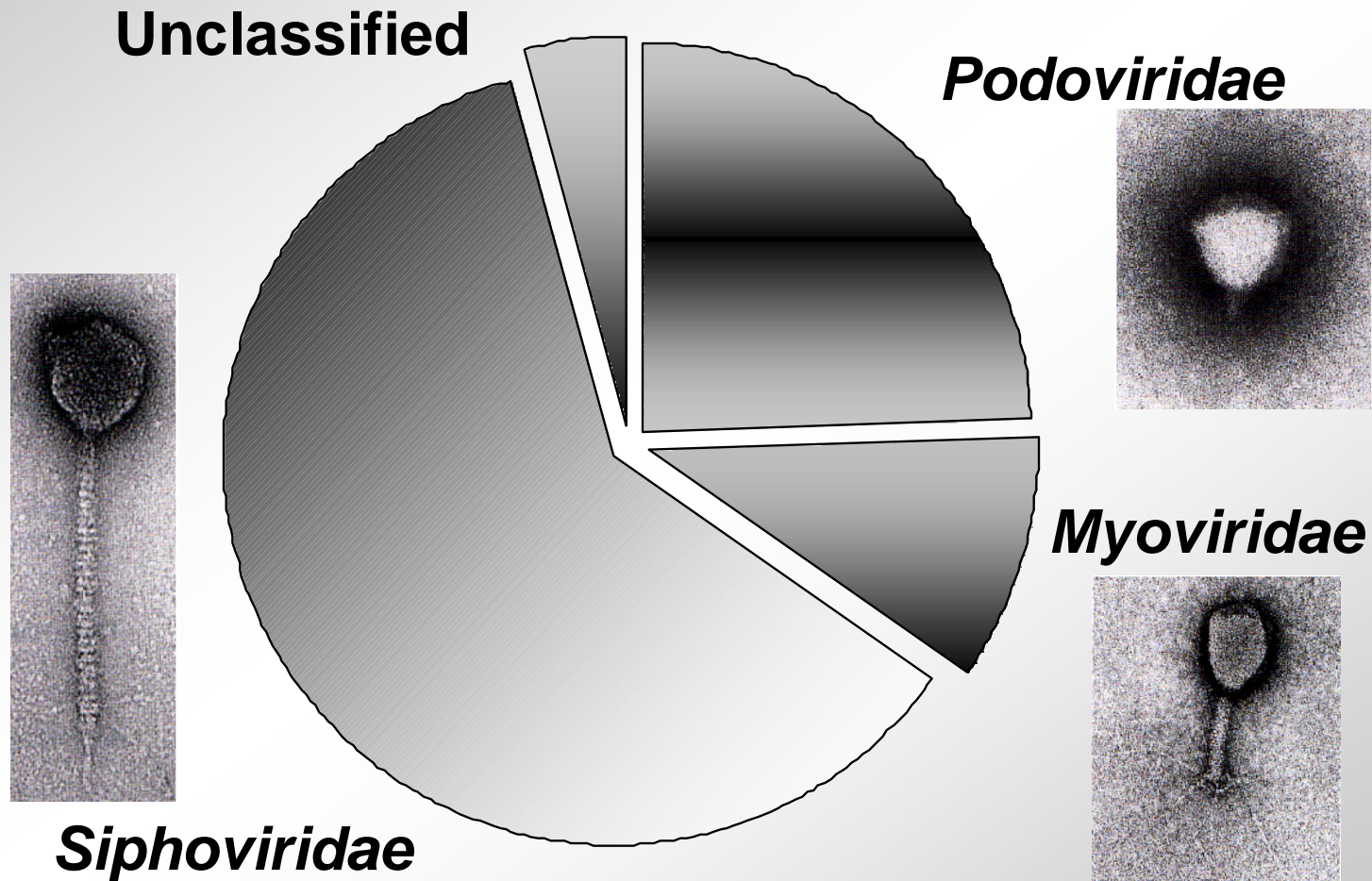
Rohwer (2003) Global phage diversity. Cell. 113.

> preview for Pedulla et al. (2003) Origins of highly mosaic mycobacteriophage genomes

Movement of DNA via mobile elements

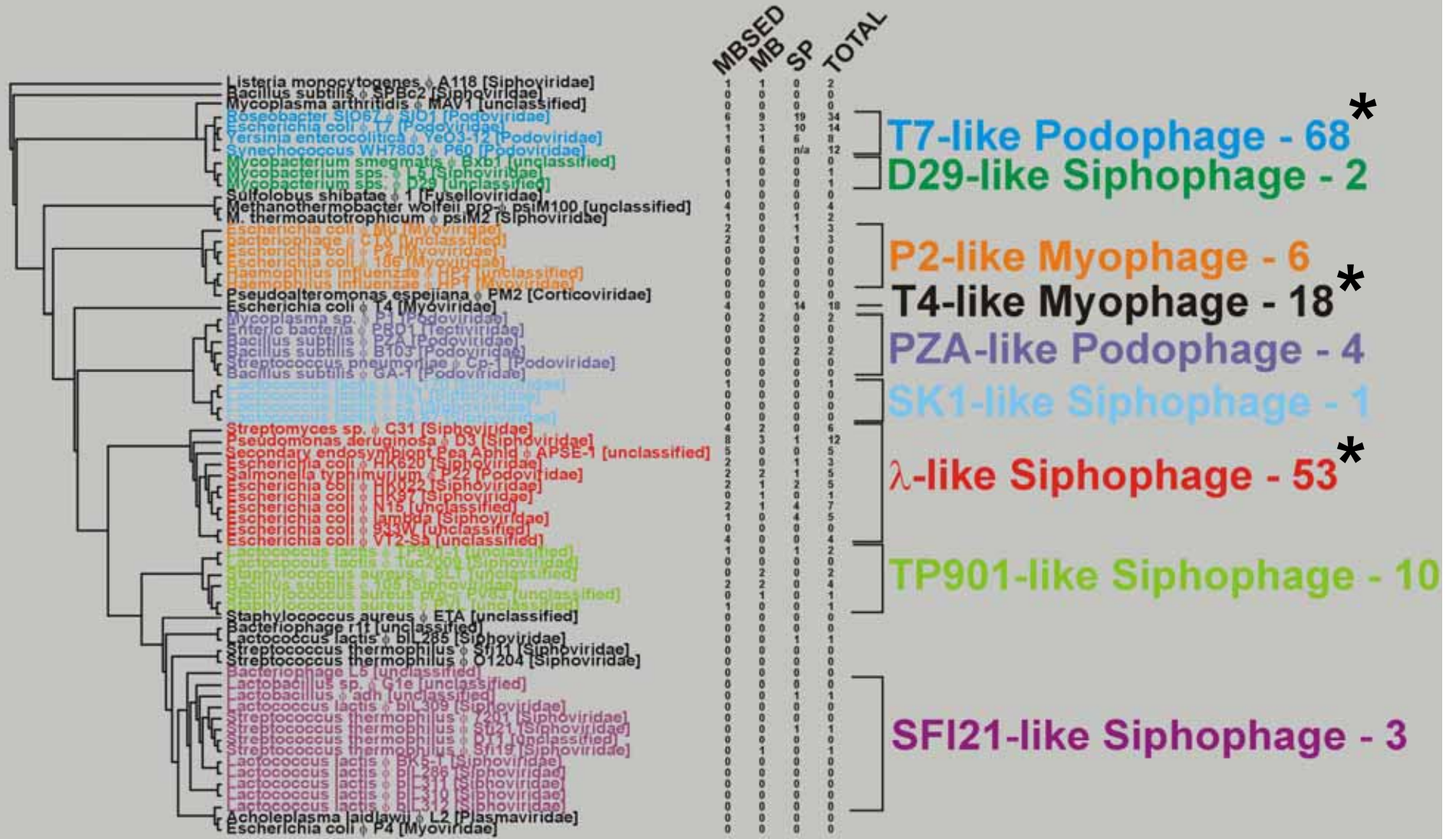


Siphophage are the most common type of phage



Lytic versus Temperate

Marine phage share a common origin

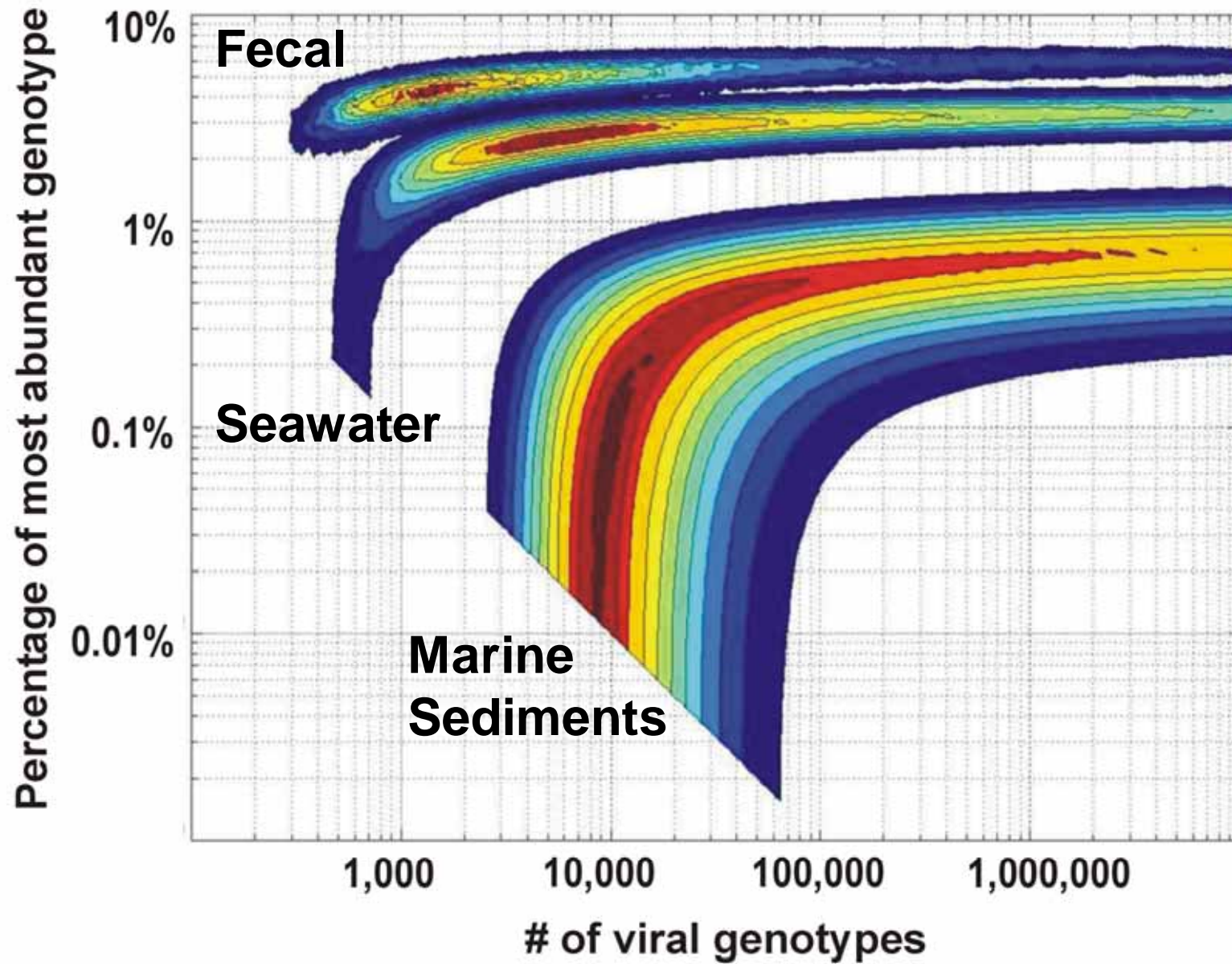


Total Phage With a Hit to the Tree

65 38 71

Phage “signatures” of different environments

Phage communities are extremely diverse



Lots of rare phage genotypes

Phage can move between environments

- **Identical T7-like DNA polymerase genes found in freshwater, seawater, estuarine, sediment, terrestrial, extreme, and metazoan-associated environments**

Breitbart et al., (2004) Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microb Lett* 236 (2). 249-256.

- **Phage-encoded exotoxin genes found throughout the environment**

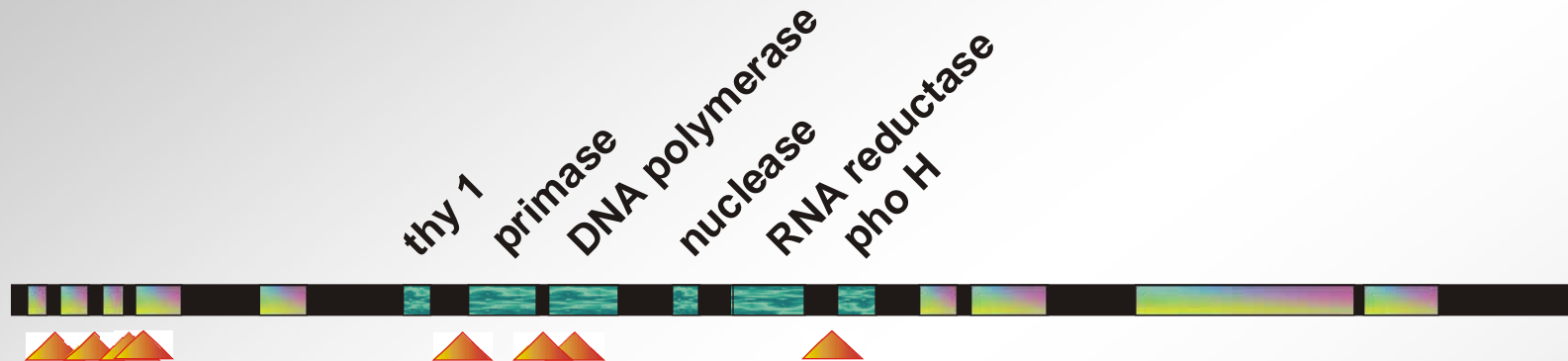
Casas et al., unpublished data.

- **Phage successfully propagate on hosts from another environment**

Sano et al., (2004) Movement of viruses between biomes. *AEM*. 70 (10).

Is everything everywhere?

Roseophage SIO1's transcription and replication are linked to phosphate availability



Pho-boxes spread throughout the genome

- transcription response elements that are activated by phosphate starvation

PhoH-like protein, RNA reductase, nuclease

- proteins involved in phosphate metabolism
- Roseophage SIO1 uses host dNTPs

Phage can carry genes that are important for the host's metabolism

Conclusions

Most genetic novelty and diversity is in the phage fraction

- Shotgun libraries of uncultured communities

Distinct phage groups dominate in different environments

- Environmental signatures

Phage are moving between biomes

- constrains total phage diversity



Math Guys - SDSU

Peter Salamon

Joe Mahaffy

James Nulton

Ben Felts

Beltran Rodriguez-Brito

David Bangor

Florent Angly

Rohwer Lab - SDSU

Forest Rohwer

Linda Wegley

Veronica Casas

Emiko Sano

Jon Miyake

Rob Edwards - UT Memphis

Anca Segall - SDSU

SDSU-MCF

David Mead 

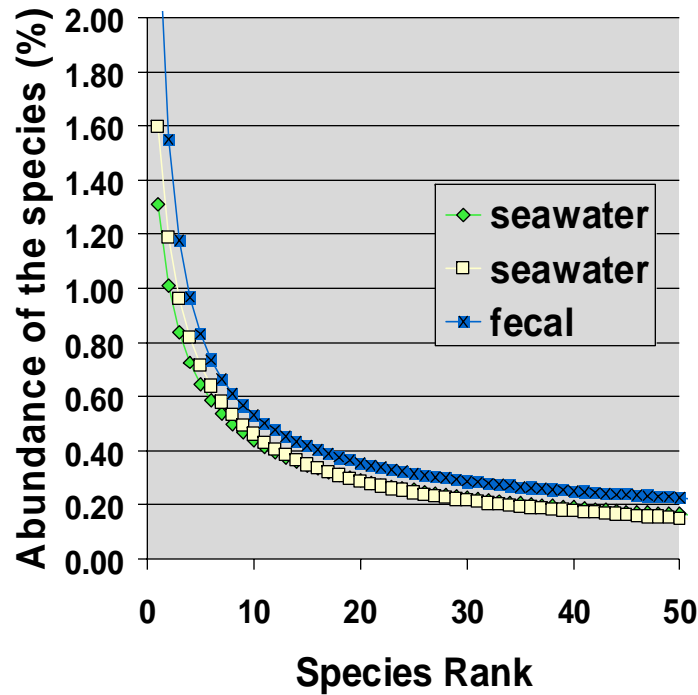
Funding

**NSF - Biotic Surveys and Inventories
- Biological Oceanography**

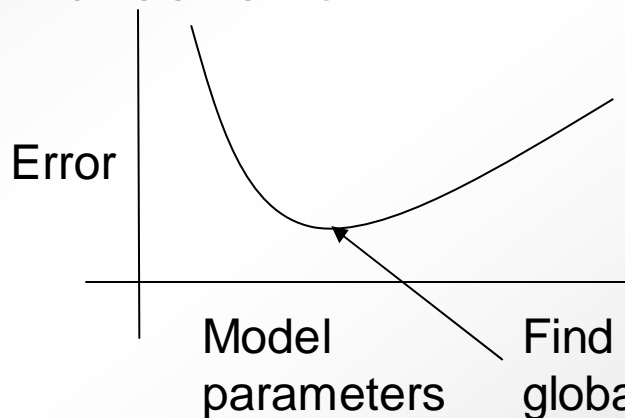
EPA – STAR Fellowship







Continue this procedure until we obtain the smallest error



Determine the actual contig spectrum of the sample



Predict a contig spectrum using a species abundance model

Compute the error between the actual and predicted

Adjust the parameters in the species abundance model to minimize errors



Rank abundance curve follows a power law

Mission Bay seawater sample

Model	% most abundant	Error	Number of species	Parameters of the model
Power Law	2.65	2.11	7420	B=0.73062
Exponential Law	0.42	16.2	7500	B=0.0042156
Logarithmic	3.02	2.81	1560	B=1.8054
Broken Stick	0.78	14.6	960	No additional parameters
Niche Preemption	0.476	38.1	8200	K=0.0047582
Lognormal	2.77	2.31	43110	Sigma=2.2237

Power Law

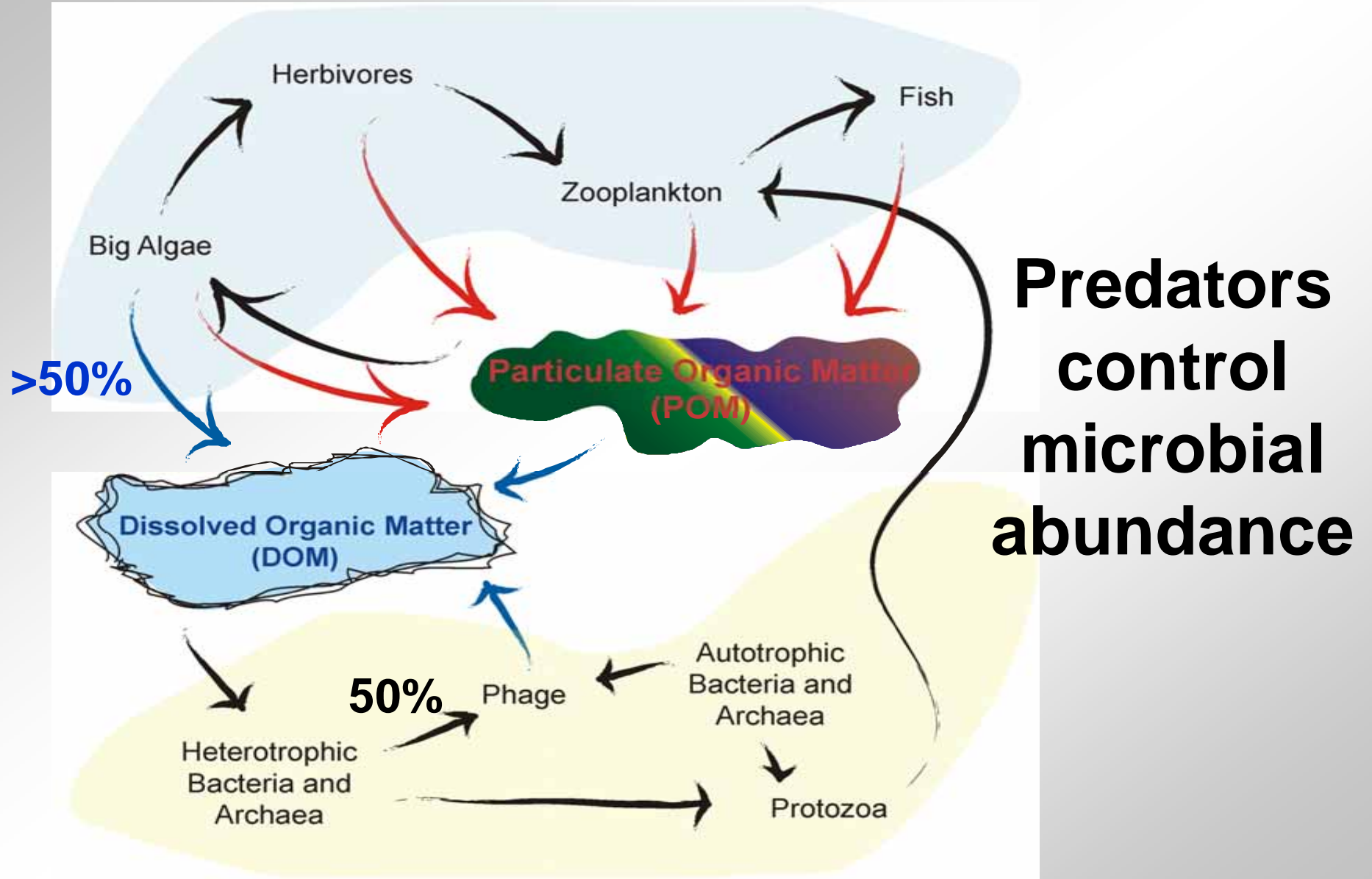
$$n_i = ai^{-b}$$

n_i = the number of individuals belonging to species i

a = the % abundance of the most abundant species

b = related to the evenness of species in the community

Classical Marine Food Web

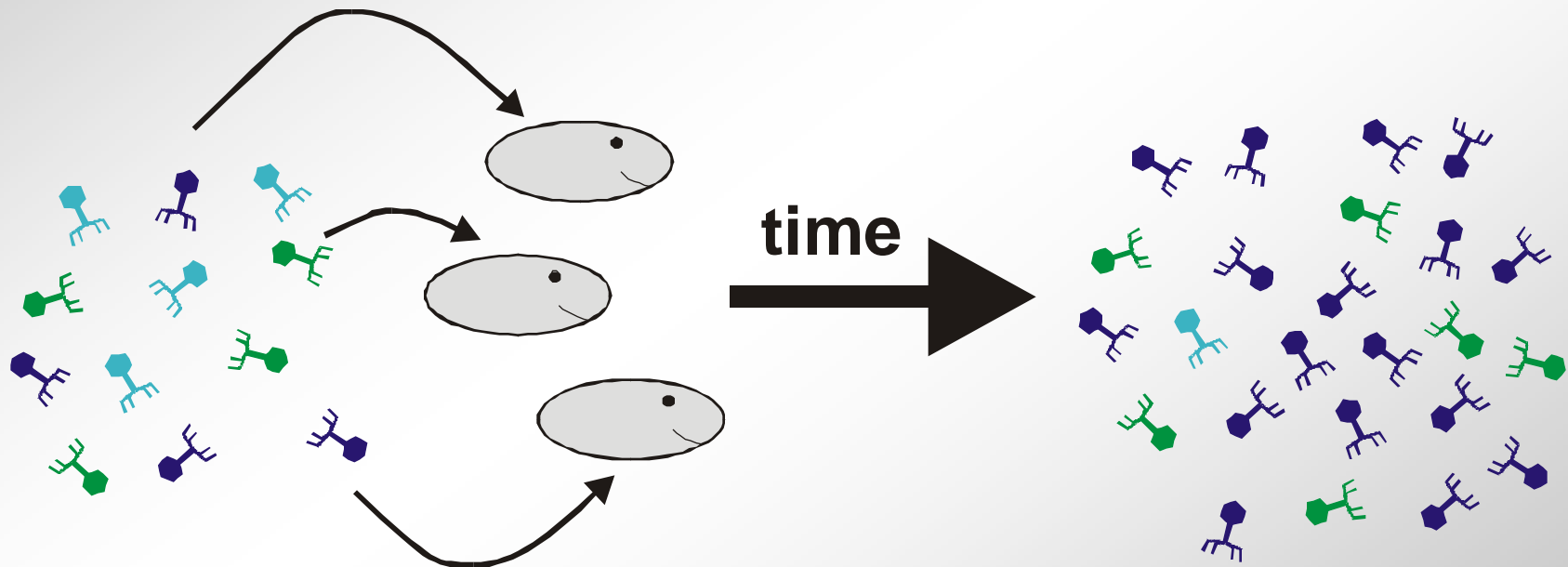


Marine Microbial Food Web

Why power law?

Arises from a series of connected, exponential events

Example – Phage competing for same host



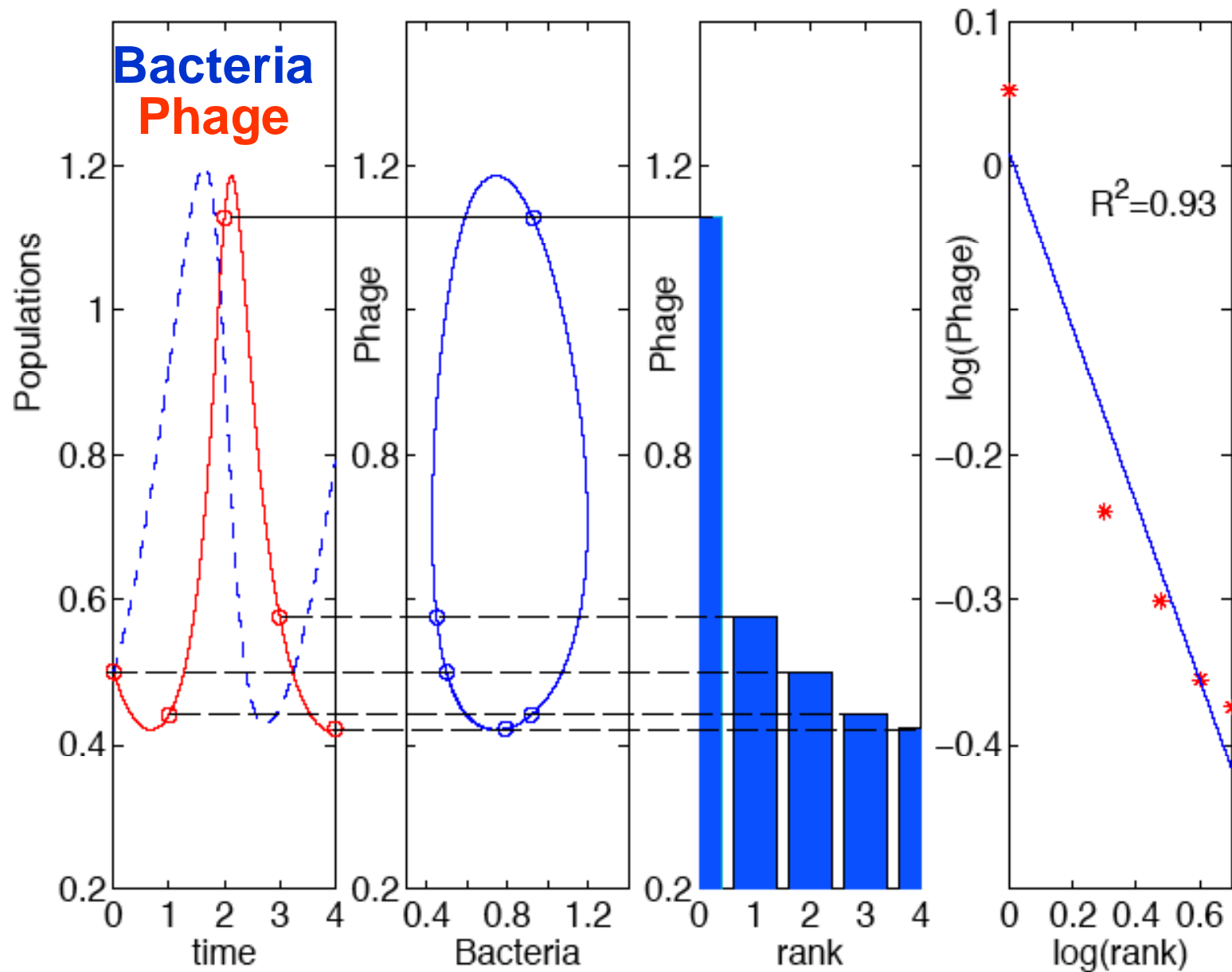
The power law can also be obtained using a single phage-host pair with Lotka-Volterra dynamics

Distinct phage groups dominate in different environments

		% in Marine	% in Fecal
Podophage	T7-like	41	5
	PZA-like	2	11
Myophage	T4-like	11	0
	P2-like	4	0
Siphophage	λ-like	32	11
	D29-like	1	5
	SK1-like	1	11
	TP901-like	6	21

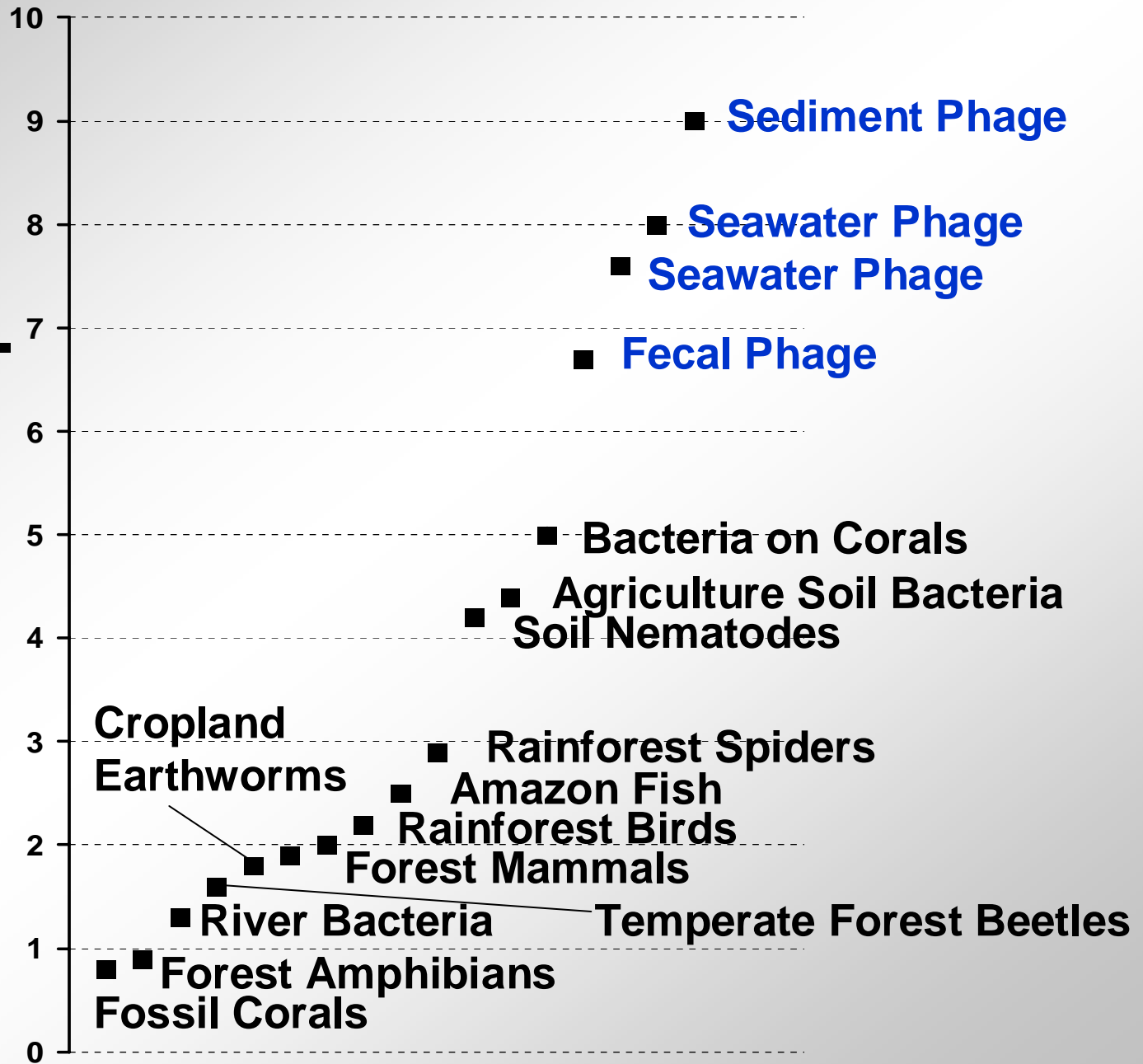
Common origin for marine phage

Modeling phage-host population structure



Phage diversity is the highest reported

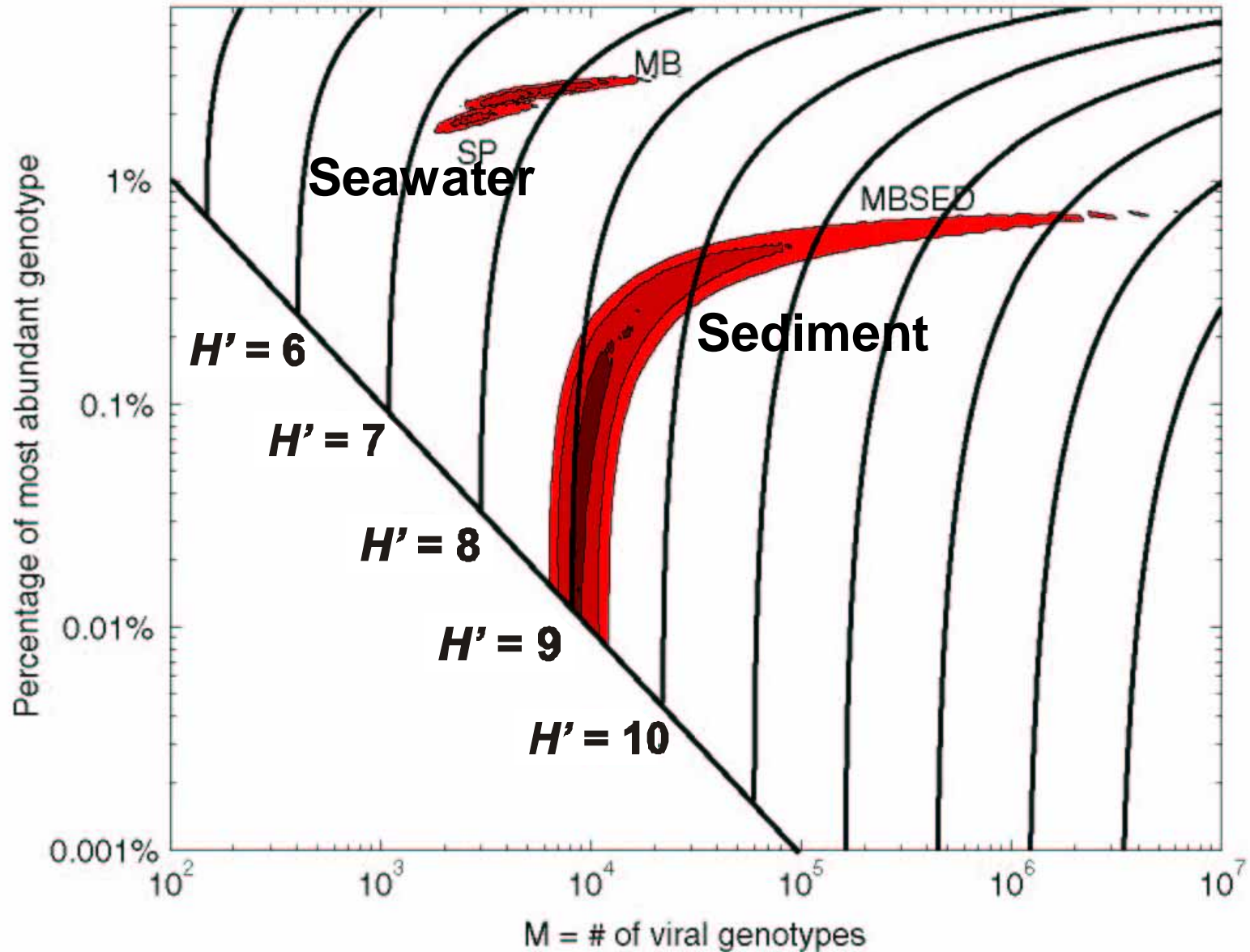
**Shannon-
Wiener
Index
(nats)**



Phage as Predators

- **More abundant than prey**
- **Smaller than prey**
- **More diverse than prey**
- **Can directly change genetic diversity of prey through horizontal gene transfer**
- **Power law distribution**

Phage community structure



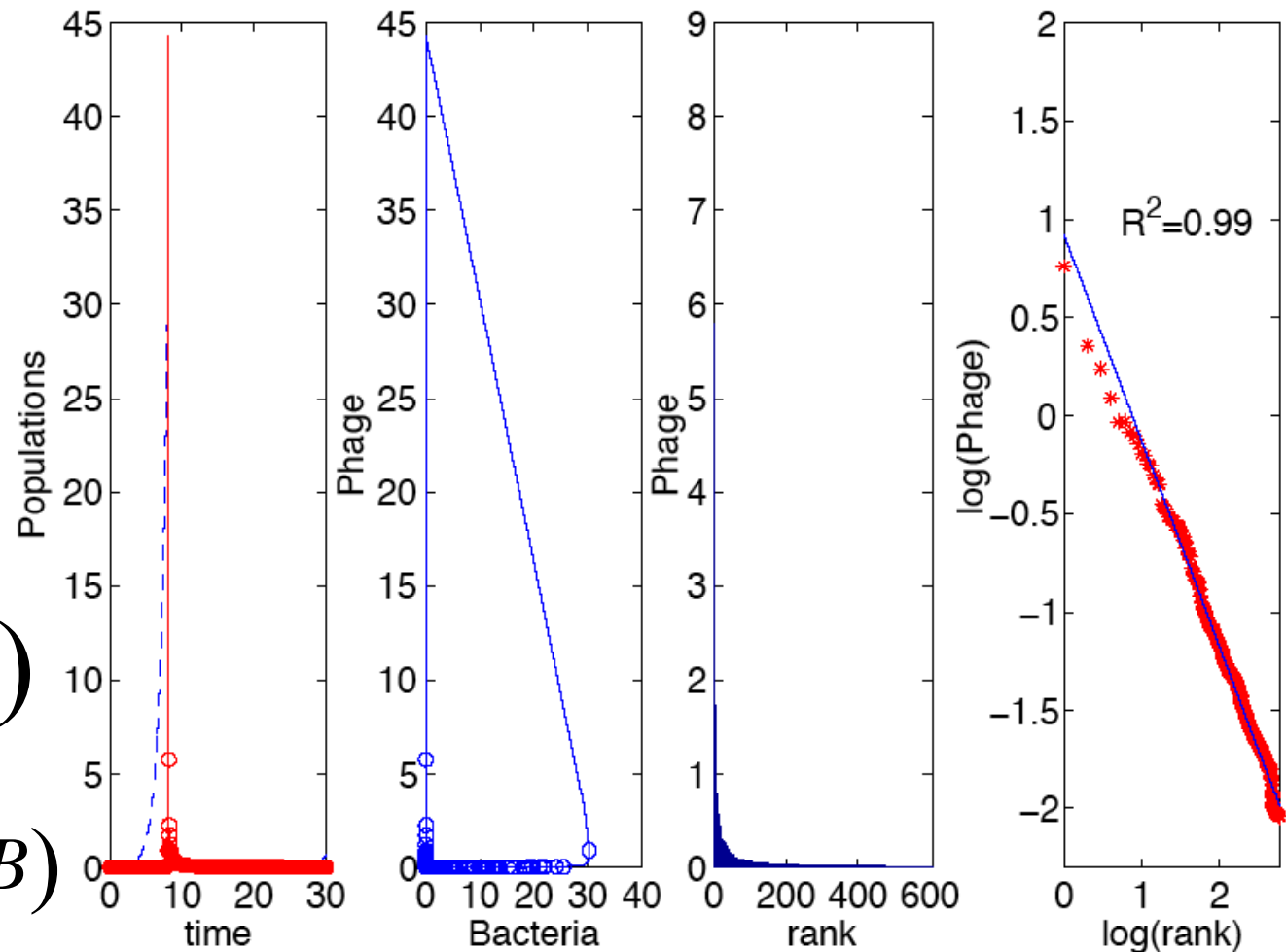
Phage diversity is the highest reported

Power-law for a single phage-host pair

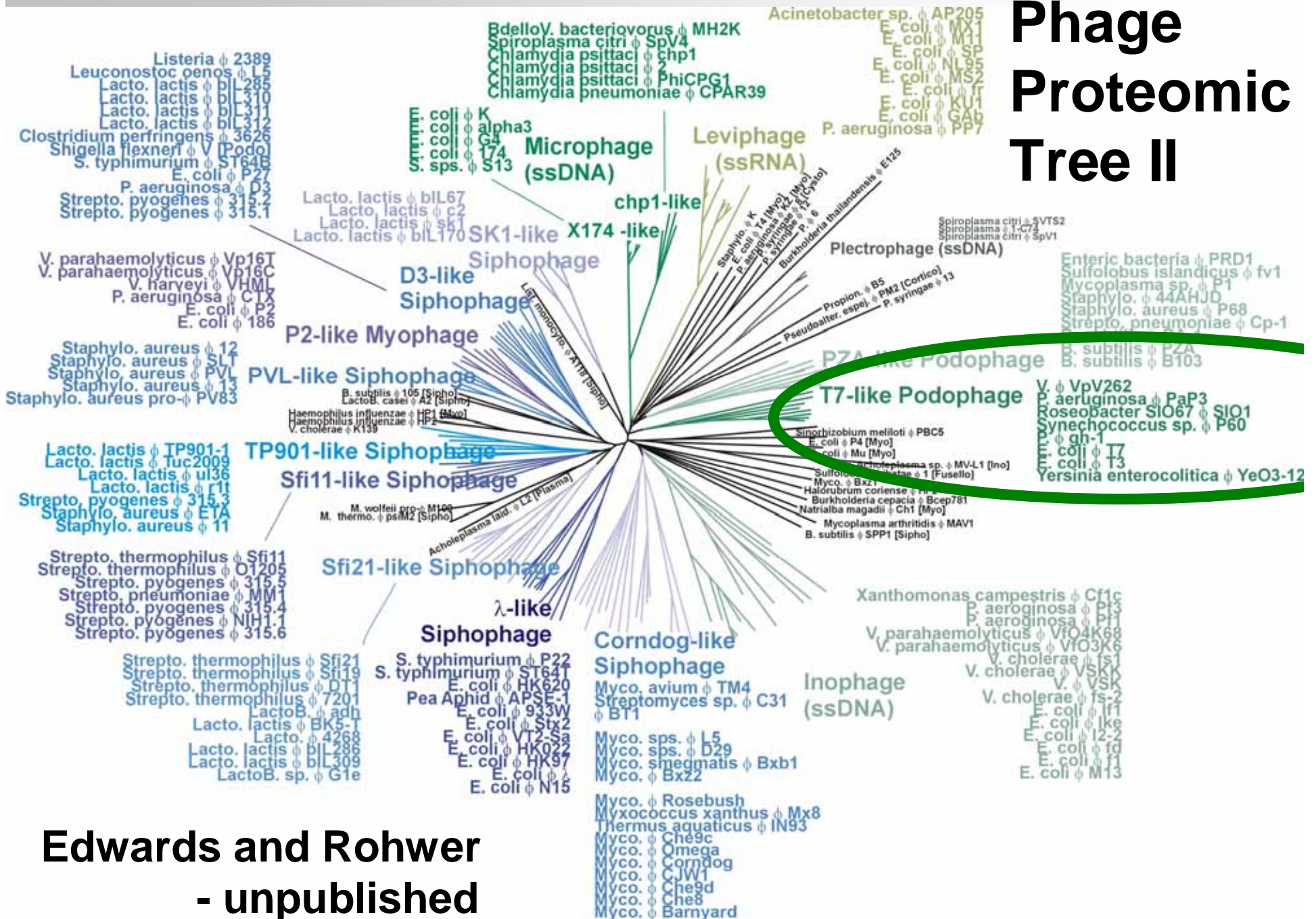
Modified Lotka-Volterra dynamics

$$\frac{dB}{dt} = B(\alpha - \beta P^2)$$

$$\frac{dP}{dt} = P^2(-\gamma + \beta B)$$



Phage Proteomic Tree II

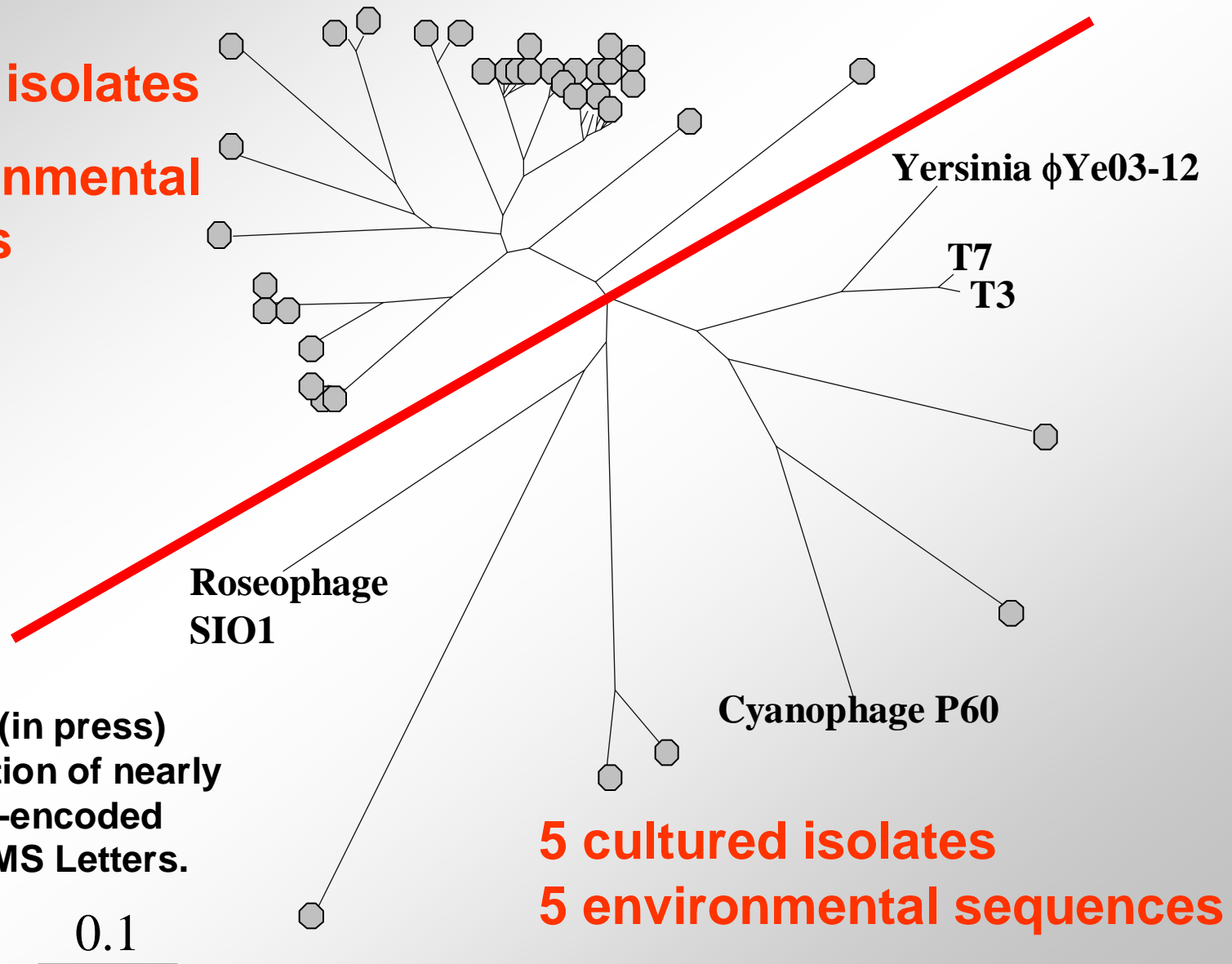


Edwards and Rohwer
- unpublished

Most of the environmental sequences belong to a novel phage group - The PUP Clade

0 cultured isolates

112 environmental sequences

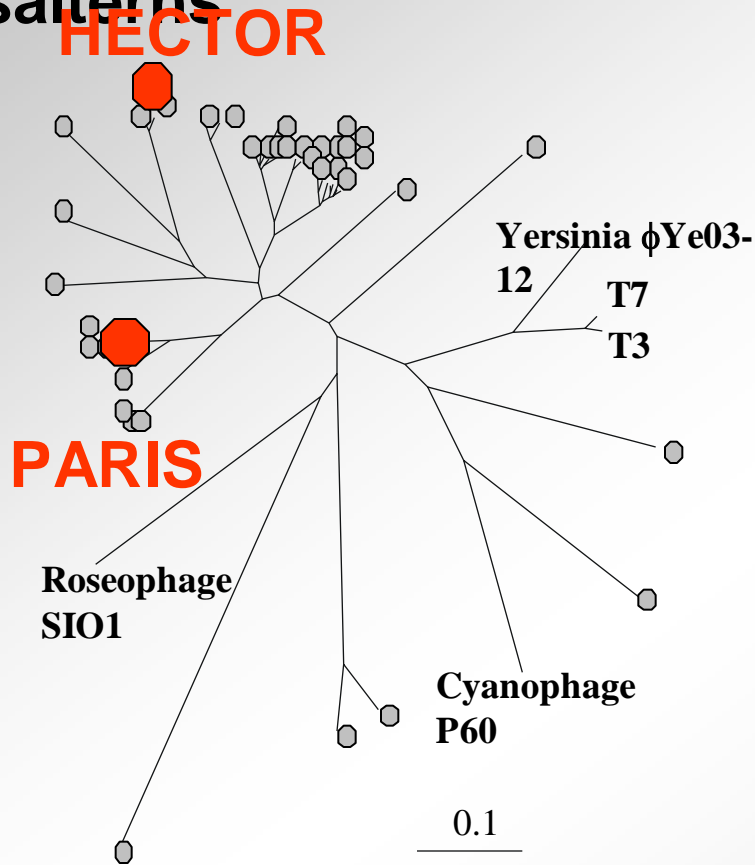


Breitbart et al., (in press)
Global distribution of nearly
identical phage-encoded
sequences. FEMS Letters.

0.1

5 cultured isolates
5 environmental sequences

Identical DNA polymerase sequences were found in phage communities from every major biome - aquifers, marine, soil, sediments, sea ice, corals, rumen, human, salterns



- **HECTOR and PARIS were present in ~1 out of every 10^6 phage particles sampled**
- **$\sim 10^{26}$ copies of HECTOR on the planet = 60 metric tons of this DNA sequence**

Empirical functional forms

- Power Law: $n_i = ai^{-b}$
- Logarithmic: $n_i = a(\log(i + 1))^{-b}$
- Exponential Law: $n_i = ae^{-ib}$
- Lognormal
 - n_i = the number of individuals belonging to species i
 - \mathcal{D} = the % of abundance of the most abundant species
 - b = related to the evenness of species in the community

Characteristics of phage populations (same/diff than euks?)

- **Very diverse**
- **everything is very rare**
- **power law distribution**
 - **Kill the winner with sharp peaks and long lulls**
- **things can move between environments (t7, emiko, toxins)**
 - **But still can tell environments apart (PTP)**
- **lysogeny / importance of gene transfer**
 - **But, still have the ability to identify groups**

Human diseases associated with plasmid- and phage-encoded exotoxins

	Anthrax	<i>Bacillus anthracis</i>	Plasmid
	Botulism	<i>Clostridium botulinum</i>	Phage and Plasmid
CTX	Cholera	<i>Vibrio cholerae</i>	Phage and Plasmid
DTX	Diphtheria	<i>Corynebacterium diphtheriae</i>	Phage
STX	Diarrheagenic <i>E. coli</i>	<i>Escherichia coli</i> (EHEC strains)	Phage
	Tetanus	<i>Clostridium tetani</i>	Plasmid
	Toxic shock	<i>Staphylococcus aureus</i>	Phage
SEA	Staph Food Poisoning	<i>Staphylococcus aureus</i>	Phage and Plasmid
	Scalded Skin Syndrome	<i>Staphylococcus aureus</i>	Phage and Plasmid
	Scarlet fever	<i>Streptococcus pyogenes</i>	Phage
PTX	Whooping cough	<i>Bordetella pertussis</i>	Phage (?)

Exotoxin-specific PCR positives from various sites in San Diego County

Soil

Location	Toxin				
	STX	SEA	DTX	PTX	CTX
SDSU: Book Store				+	
SDSU: Little Theatre			+		
SDSU: Scripps Cottage		+			
SDSU: Softball field		+	+		
SDSU: Olmeca Dorms		+			
SDSU: Physics			+	+	+
Singing Hills Golf Course				+	+
Flynn Springs		+	+	+	+
Crown Point Elementary field		+		+	
School				+	
Cuyamaca College	+	+	+		
Monte Vista HS	+			+	
Stelzer Park			+		
Santee Lakes Park				+	+
Lake Murray		+	+	+	+

Sediment

Location	Toxin				
	STX	SEA	DTX	PTX	CTX
San Vicente		+		+	+
Mission Bay 1				+	+
Lindo Lake				+	+
Reservoir				+	+
San Diego		+	+	+	+
Santee Lake #3				+	+
Fiesta Island				+	+
Pacific Beach				+	+
Solana Beach				+	
Ocean Beach				+	+
Encinitas				+	
Del Mar			+	+	+
Sea				+	+
Carlsbad				+	+
Otay Mesa	+		+		+
Imperial Beach				+	+
Chula Vista			+	+	+
Coronado				+	+
National City		+	+	+	
Scripps Ranch				+	+
Lake Murray				+	+
Lake Hodges		+	+	+	
La Jolla Shores				+	
T.J. Estuaries		+	+	+	+

Water

Location	Toxin	
	PTX	CTX
Lindo Lake	+	
Mission Beach	+	+
Lake Murray	+	+
Fiesta Island	+	+
San Diego River	+	+
Santee Lakes #3	+	+
Mission Bay #1	+	
Mission Bay #2	+	+
El Cap Reservoir	+	
Lake Jennings	+	+
Oceanside	+	
Imperial Beach	+	+
Chula Vista	+	
Pacific Beach	+	
Cardiff by the Sea	+	
Torrey Pines	+	
Ocean Beach	+	
Encinitas	+	+
San Vicente	+	
Coronado	+	
Scripps Ranch	+	+
Lake Hodges		+
La Jolla Shores	+	+
TJ Estuaries	+	+

~10% of samples are positives

RT- PCR for DTX (90 samples total)

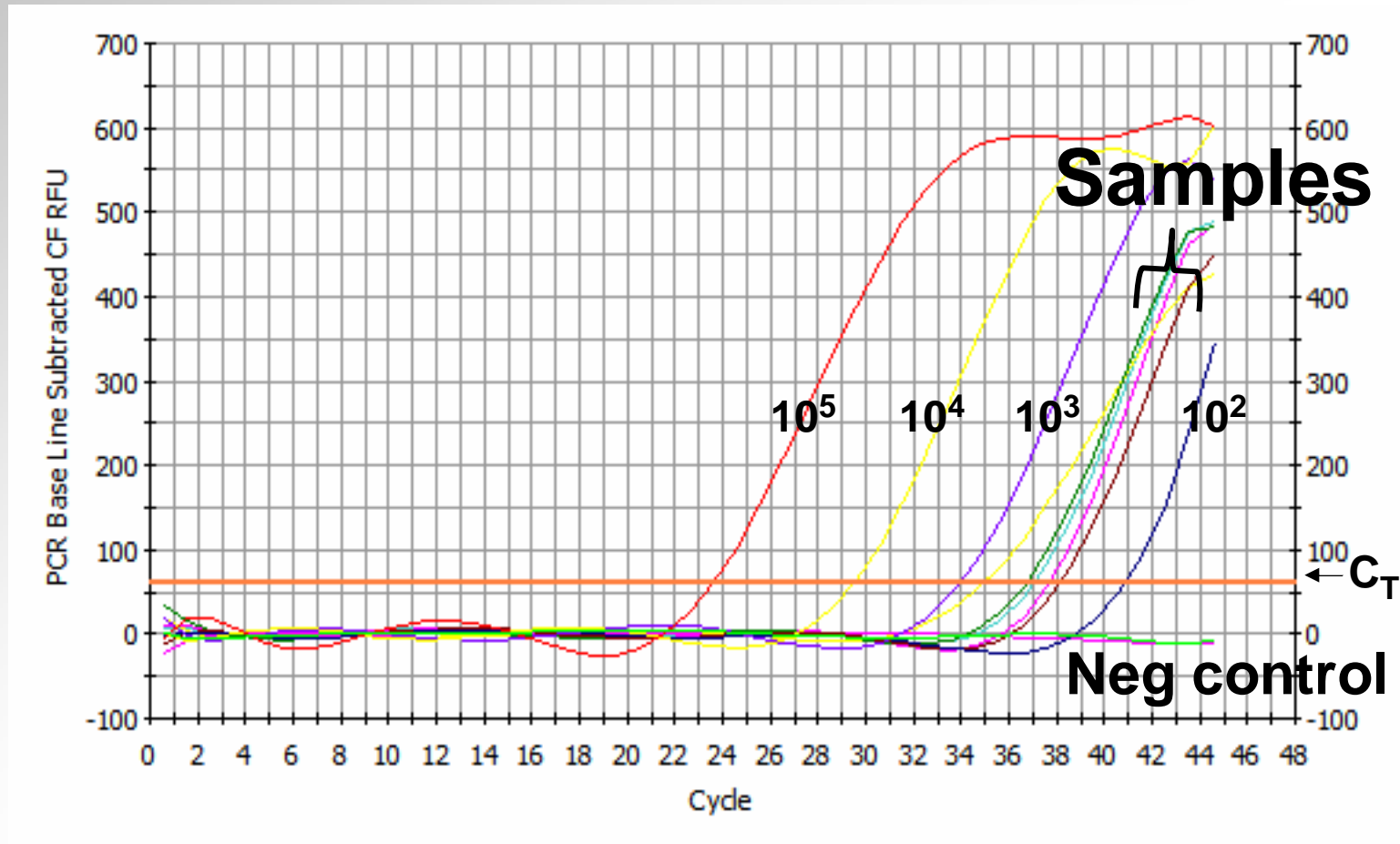
copies
of DTX

Sample type

Sample location

Sample type	Sample location	# copies of DTX
Soil	Valhalla HS (1 g)	54
	Old Town (1 g)	211
	Perkins Elementary (1 g)	83
	Cuyamaca College (1 g)	239
Sediment	Otay Mesa Reservoir (1 g)	181
	Imperial Beach (1 g)	625
Water	National City Marina (1 l)	279
	Cardiff (1 l)	123
	Torrey Pines (1 l)	961

SEA RT-PCR of phage fraction from soils



Also observed exotoxin genes in uncultured shotgun libraries

Exotoxin genes are in the free-phage fraction

Can phage move between biomes?

Fraction Name	Microbial	VLP	DOM
Filter Size	0.45 μm	0.2 μm	100 kD
Filterate Contains	Microbes, VLPs and DOM	VLPs and DOM	DOM

innoculum

food

Microbial fraction



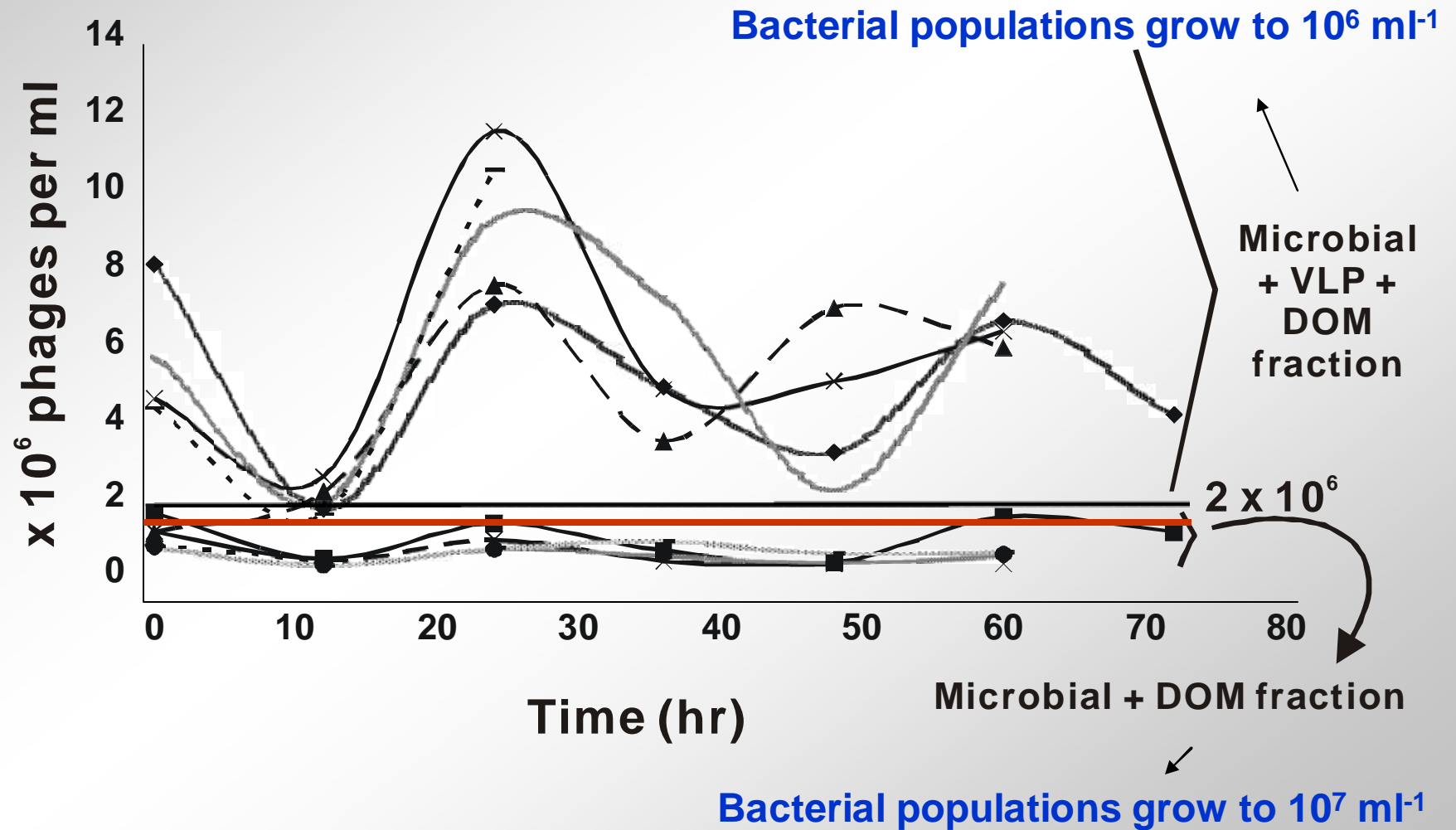
Count VLPs and microbes at different time points

VLP = viral-like particle \approx phage

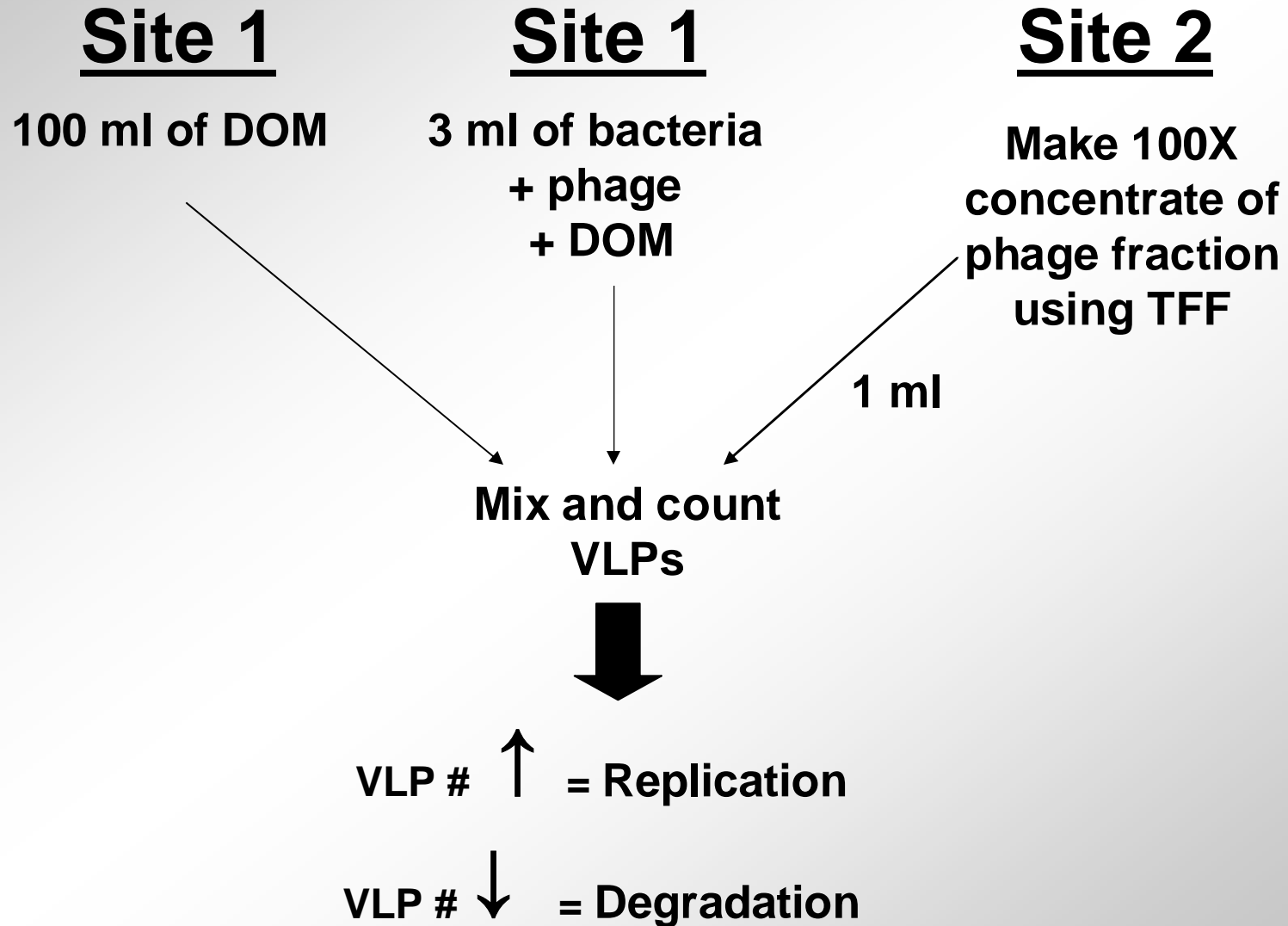


Phage cannot find hosts at low concentrations

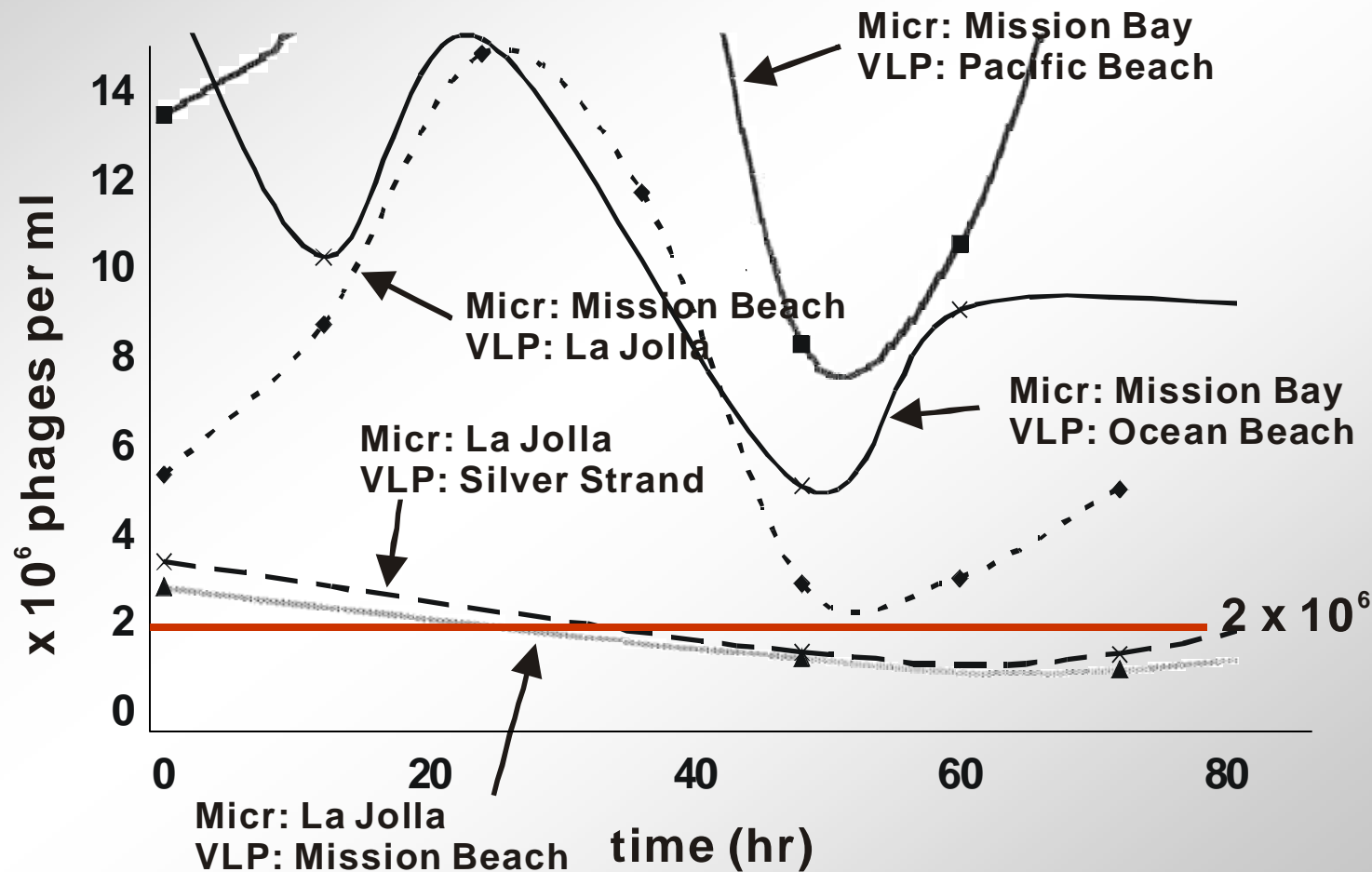
- communities go extinct -



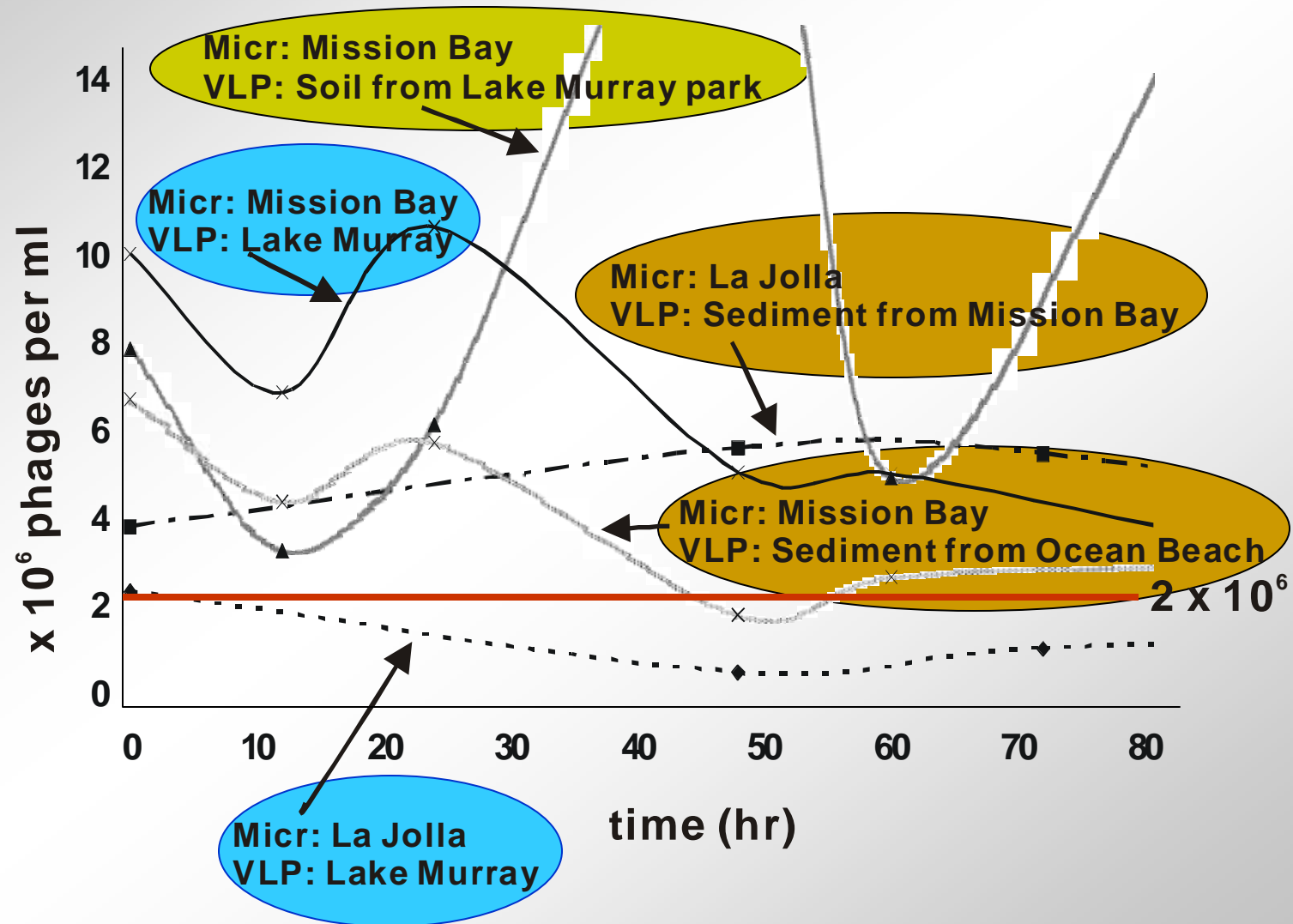
Mixing experiments



Marine phages can propagate on marine microbes from different locations

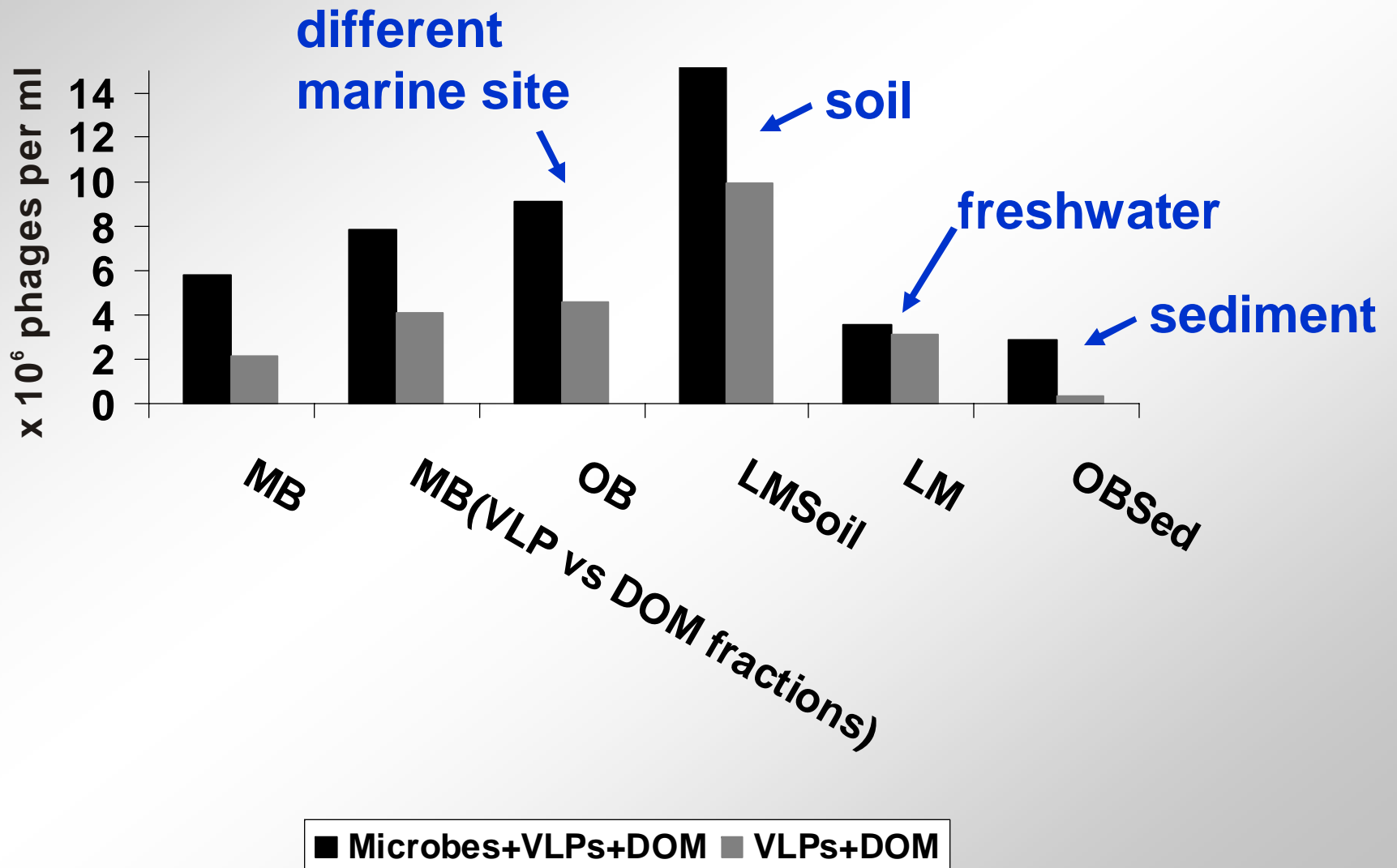


Freshwater, sediment, and soil phage communities can propagate on marine microbes



Growth versus Decay - endpoints

- Mission Bay microbial community



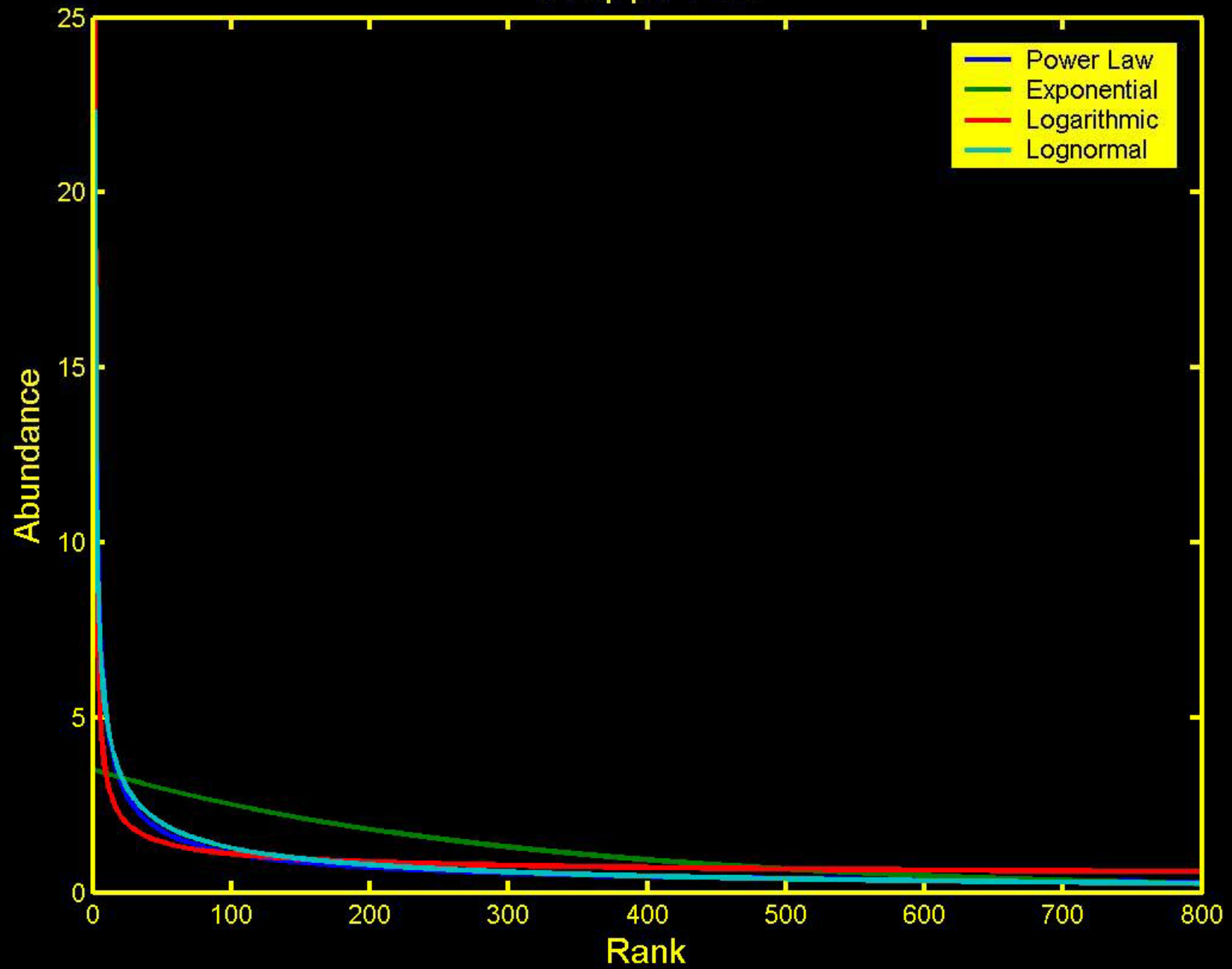
There are a lot of different phage types:

- Majority of phage ORFs are unknowns
- ~20% of phage genomes have no known close relative
- Very high richness and Shannon Index

There are not a lot of different phage types:

- Hector and Paris are present in ~1 out of 10^6 phage
- Exotoxin-encoding phage are common
- Mixing experiments
- Populations have power law distributions not lognormal

Scripps Pier



Scripps Pier water



Breitbart et al., (2002) Genomic analysis of uncultured marine viral communities. PNAS. 99:14250-14255.

Mission Bay water



Mission Bay sediment

Breitbart et al., (2004) Diversity and population structure of a nearshore marine sediment viral community. Proc Royal Society B. 271. 565-574.



Fecal

Breitbart et al., (2003) Metagenomic analyses of an uncultured viral community from human feces. J Bacteriology. 85 (20). 6220-6223.