



UNIVERSIDAD DE SONORA



POSGRADO EN BIOCIENCIAS

MICROBIOLOGÍA MOLECULAR (2442)

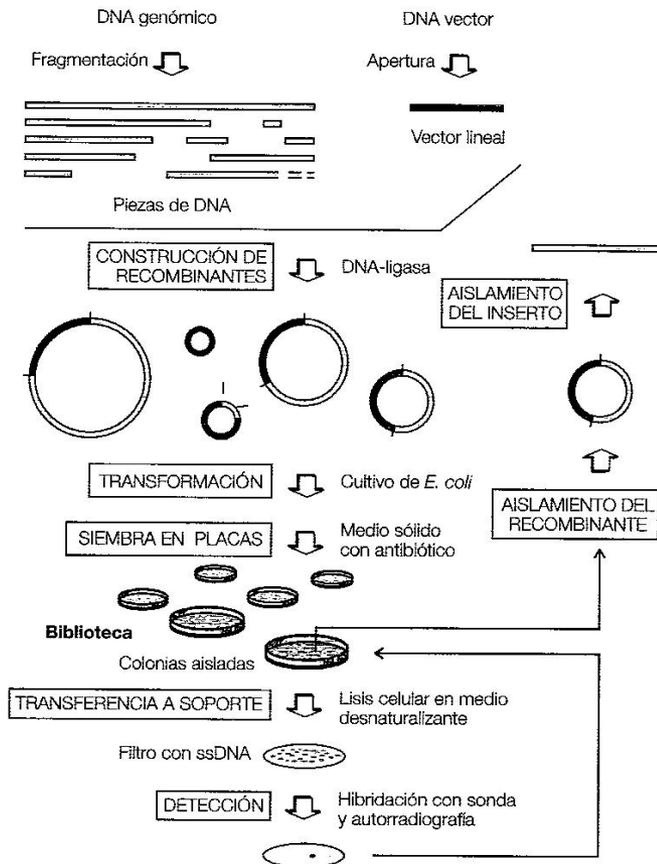
UNIDAD 6. BIBLIOTECAS GÉNICAS

Elaboró: Dra. Kadiya del C. Calderón Alvarado

OBJETIVOS DE LA CONSTRUCCIÓN DE UNA BIBLIOTECA GÉNICA

- ▶ AISLAMIENTO DE SECUENCIAS Y GENES.
- ▶ CONSERVACIÓN DEL DNA DE UNA ESTIRPE, UNA ESPECIE O UN INDIVIDUO
- ▶ ESTUDIO DE GRANDES REGIONES GENÓMICAS O DE GENOMAS COMPLETOS.

TIPOS DE BIBLIOTECAS O BANCOS GÉNICOS



► **BIBLIOTECAS GENÓMICAS O GENOTECAS:** fragmentos clonados, el genoma completo de una especie.

► **GENOTECAS PARCIALES, MINIGENOTECAS O BANCOS GÉNICOS**

► **BIBLIOTECAS DE ADN o ADNc:** construidos a partir de la población de mensajeros aislados de un tipo celular..

► **BIBLIOTECAS DE CLONACIÓN:** Construidas con vectores de clonación simples.

► **BIBLIOTECAS DE EXPRESIÓN:** aportan las señales activas en el hospedador que permiten la transcripción y traducción in vivo de sus pasajeros.

Esquema de la construcción de una genoteca. Perera, J et al., (2002) Ingeniería Genética Ed. Síntesis. ISBN: 84-7738-966-7

ADNc= ADN complementario

Las cualidades más importantes de una biblioteca genómica son:

▶ **REPRESENTATIVIDAD.**

▶ **COMPLEJIDAD O TAMAÑO:** número mínimo de clones independientes que debe contener para que la totalidad del ADN de partida esté representado

▶ **LONGITUD O TAMAÑO MEDIO DE LOS INSERTOS**

▶ **FACILIDAD DE PREPARACIÓN Y ANÁLISIS SENCILLO.**

DIFERENTES MÉTODOS DE FRAGMENTACIÓN DEL DNA GENÓMICO PARA LA CONSTRUCCIÓN DE UNA GENOTECA

- La digestión total o parcial con **endonucleasas de restricción.**

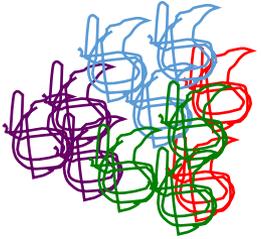
Generalidades

Los procedimientos básicos incluyen una serie de pasos:

1. Los fragmentos de ADN se generan utilizando enzimas de restricción
2. Los fragmentos producidos mediante la digestión con enzimas de restricción se unen a otras moléculas de ADN que sirven de vectores
3. La molécula de ADN recombinante, formada por un vector que lleva un segmento de ADN insertado, se transfiere a una célula huésped.
4. Al replicarse las células huésped, descendientes heredan el ADN recombinante, creándose una población de células idénticas, que llevan todas la secuencia clonada
5. Los segmentos de ADN clonados pueden recuperarse de las células huésped, purificarse y analizarse.

Biblioteca de genes

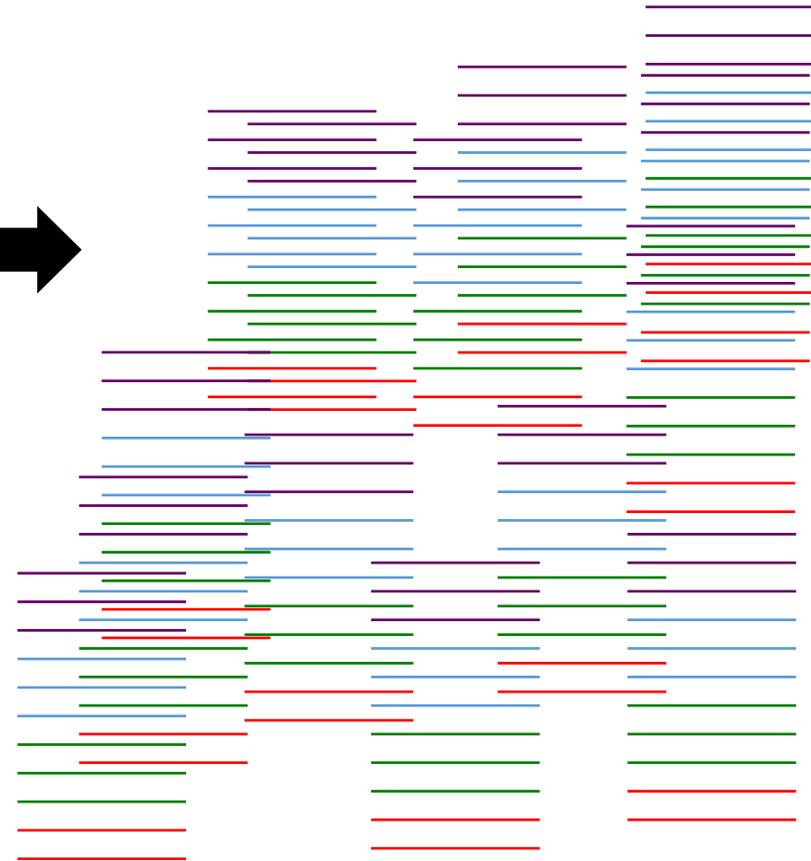
1er paso



ADN Ambiental



PCR con primers del gen
16S ARN

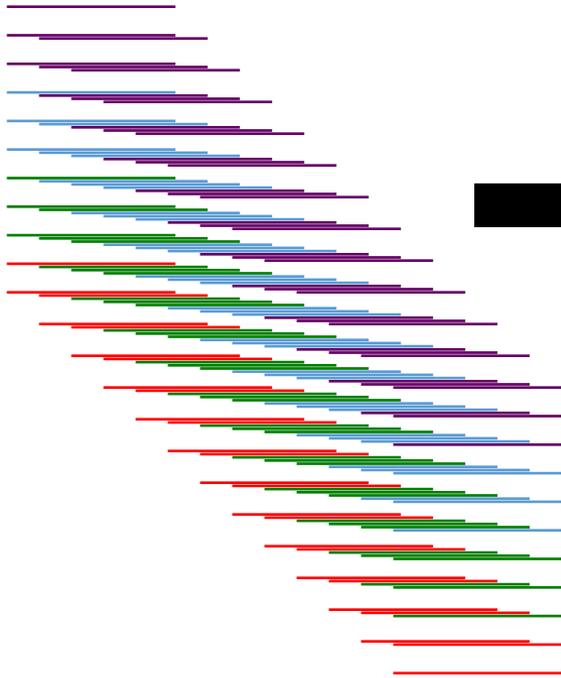


Si se tiene el mismo número de copias del gen 16S rARN por célula: misma proporción a ser amplificada

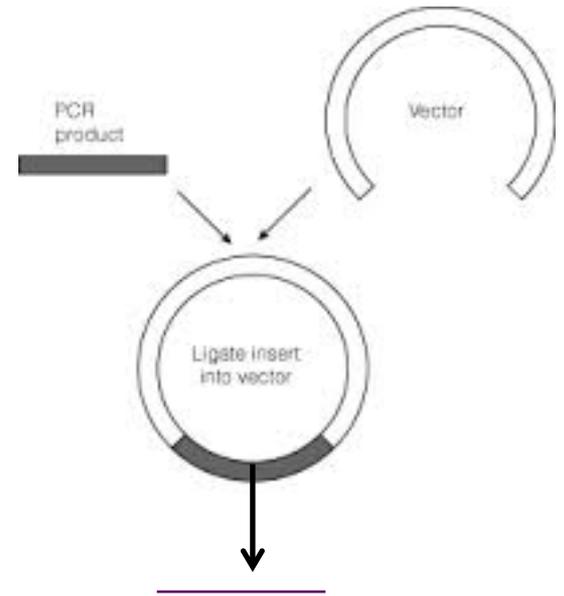
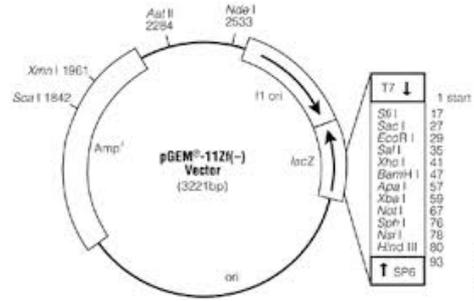
http://download.springer.com/static/pdf/683/chp%253A10.1007%252F978-3-540-77587-4_310.pdf?auth66=1413290296_1bc4f750f4125cb54198fce82daed4c5&ext=.pdf

Biblioteca de genes

2º paso

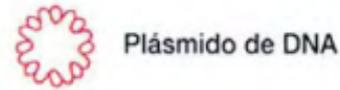


Ligación



Transformación del ADN

Transformación con un plásmido

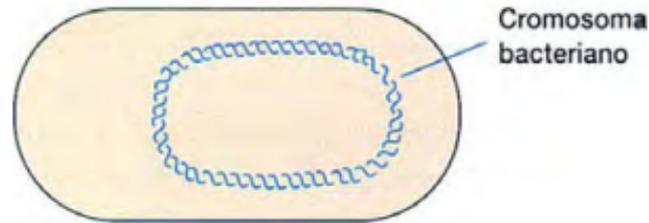


La transfo

laboratorio

Método I
en las cé

nte



Captación
del plásmido

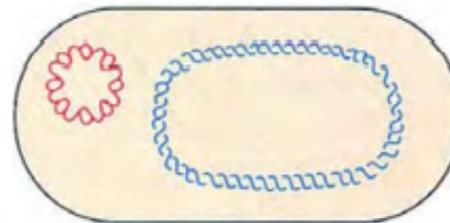
le

Tratamiento
aumenta la |

Integración
recombinación
no reciproca



Tran



Transformación estable

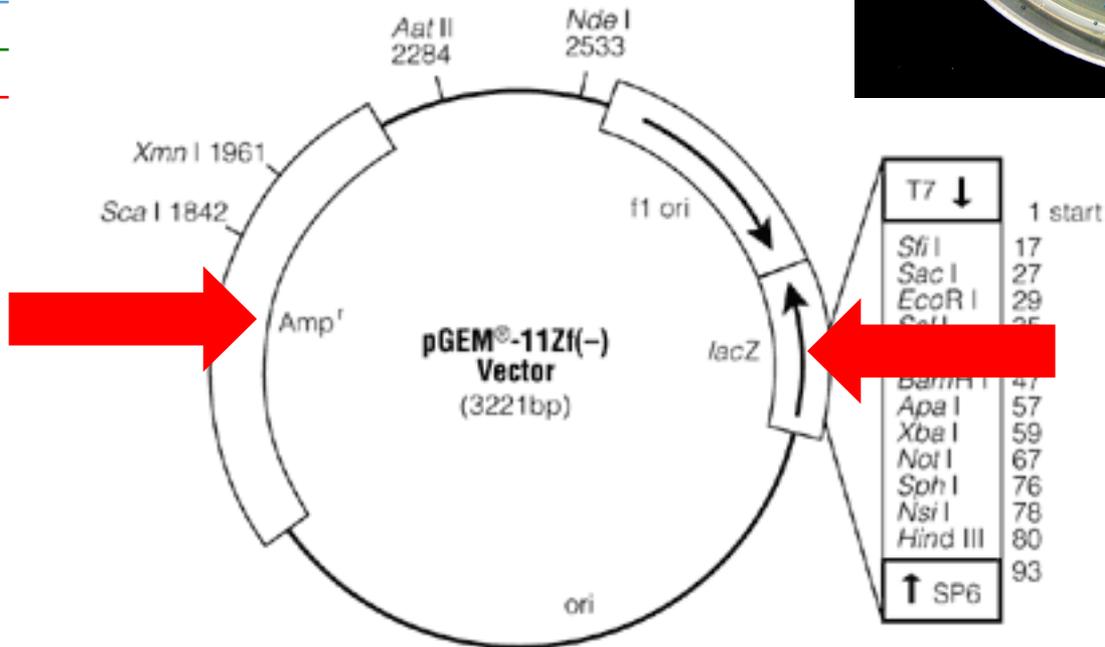
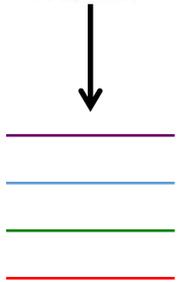
Transformación artificial con plásmidos.

Biblioteca de genes

3^{er} paso



¿plaquear... cómo?



0290V002_5A

“Rastreo”

Otros vectores, como pUC18



El terciopelo se

La placa original se presiona con terciopelo estéril

Producción de la réplica durante la noche.

Se identifica la colonia correspondiente de la placa original, y se utiliza para nuevos análisis.

Producen colonias azules que contienen una sustancia que interrumpe la capacidad de formar colonias azules, esta capacidad.

Las células que crecen en esta placa solo las células que contienen plásmidos formarán colonias.

Se inserta en el vector plasmídico

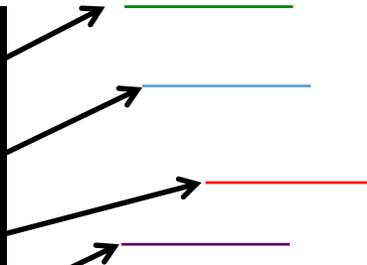
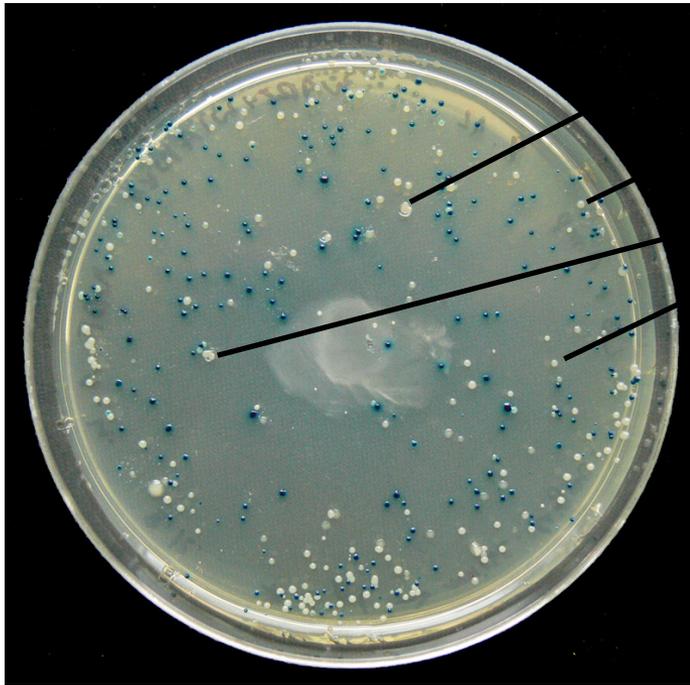
Se restringe el sitio de inserción de la capacidad de formación múltiple interrumpe

Klug and Cummings. Conceptos de Genética

Detailed description: The image shows a petri dish with a dense lawn of small, blue bacterial colonies on a dark agar surface. The colonies are scattered across the entire surface of the dish. The text is overlaid on the image in various positions, providing context for the experiment. The text describes the process of identifying a specific colony from a larger plate for further analysis. It mentions the use of a sterile cloth to press the original plate and the production of blue colonies during replication. It also notes that the colonies contain a substance that interrupts the ability to form blue colonies, and that only cells containing plasmids will form colonies. The text is partially obscured by a blue box.

Biblioteca de genes

4º paso



Seleccionar colonias (individualmente)

¿Cuántas clonas (colonias) debemos seleccionar?

El número de clones necesarios para contener todas las secuencias de un genoma depende del tamaño medio de los insertos clonados y del tamaño del genoma a clonar.

Supongamos que deseamos preparar una biblioteca del genoma humano utilizando como vector al fago lambda. El genoma humano tiene $3,0 \times 10^6$ kb; si el tamaño medio de los insertos clonados en el vector es de 17 kb, y queremos tener una probabilidad del 99% ($P = 0,99$) de que cualquier gen humano esté representado al menos en una copia, precisaremos una biblioteca de:

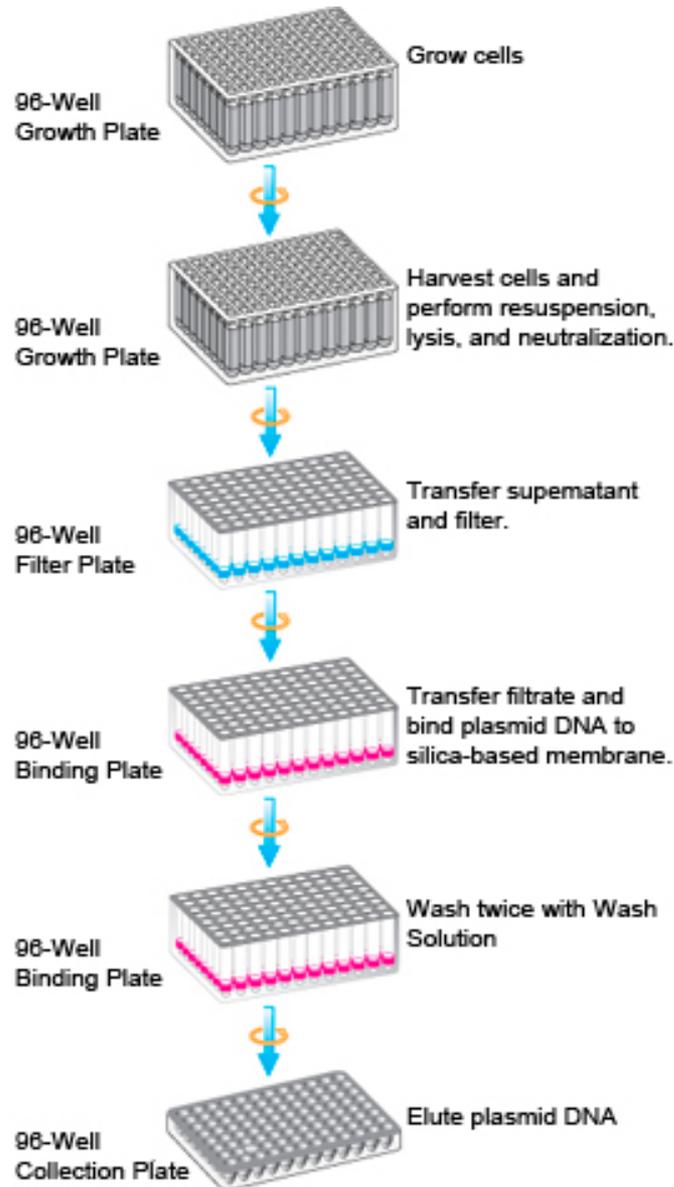
$$N = \frac{\ln(1-0,99)}{\ln\left(1 - \frac{17 \times 10^3}{3,0 \times 10^9}\right)} = 8,1 \times 10^5 \text{ clones}$$

y si hubiésemos seleccionado a pBR322 como vector, con un tamaño medio de insertos de 5 kb, la biblioteca necesaria contener:

$$N = \frac{\ln(1-0,99)}{\ln\left(1 - \frac{5 \times 10^3}{3,0 \times 10^9}\right)} = 2,8 \times 10^6 \text{ clones}$$

Biblioteca de genes

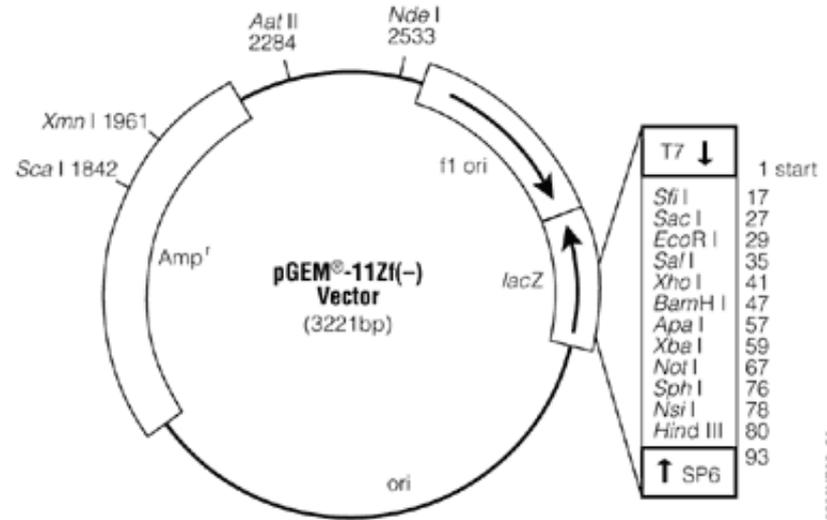
5º paso Miniprep



Biblioteca de genes

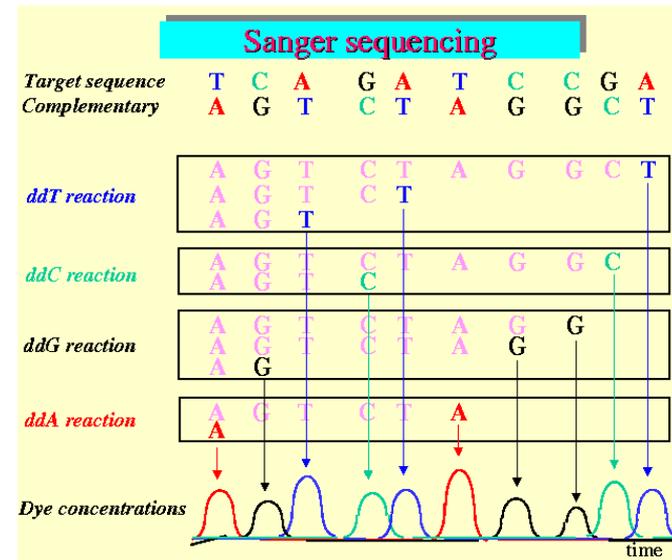
6º paso

PCR con plásmido
flanqueado en el vector



7º paso

Secuenciación Sanger

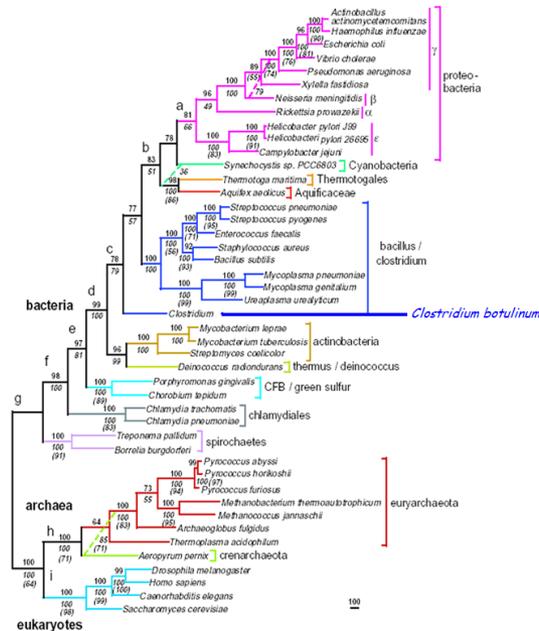


Biblioteca de genes

8o paso

>Sequence_3
CGGCGACGGGTGAGTAATGCCTAGGAATCTGCC TGGTAGTGGGGGATAACGTC CGGAAACGGGCGCTAATACCGCATACGTCCTGAGGGAGAAAAGTGGGGGATCTTCGGACCTCACGCTATCAGATGAGCCTAGGTCGGATTA
GCTAGTTGGTGGGGTAAAGGCCATACCAAGGCGACGATCCGTAAC TGGTCTGAGAGGATGATCAGTCACACTGGAAC TGGAGACAGCTCCAGCTCCACGGGAGGCAGCAGTGGGGGAATATGGACAATGGGGCAAAGCCTGAT
CCAGCCATGCCCGGTGTGTAAGAAGGCTTCGGATTGTAAAGCACTTTAAGTTGGGAGGAAAGCGAGTAAAGTTAATAACCTTGTCTGTTTTGACGTTTACCAACAGAAATAAGCACCCGGCTAACTTCGTGCCAGCAGCCGCGGTAAT
ACGAAGGGTGC AAGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTTCAGCAAGTTGGATGTGAAATCCCCGGGCTCAACCTGGGAAC TGCATCCAAACTACTGAGCTAGAGTACGGTAGAGGGTGGTGGAAAT
TCCTGTGTAGCGGTGAAATGCGTAGATATAGGAAGGAACACCAGTGGCGAAGGCGACCACCTGGACTGATAC TGCACACTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAACAGATGT
CGACTAGCCCGTTGGGATCCTTGAGATCTTAGTGGCGCAGCTAACCGGATAAGTCGACCCTGGGGAGTACGGCCCAAGGTTAAAAC TCAAATGAATTGACGGGGGCCCGACAAGCGGTGGAGCATGTGGTTTAAATTCGAAG
CAACGCGAAGAACC TTAAC TGGCC TTAGACATGCTGAGAACTTCCAGAGATGGAT TGGTGCC TTCGGGAACTCAGACACAGGTGCTGCATGGCTGTCGTGAGTGTGGGTTAAGTCCCGTAAACGAGCGCA
ACCCTTGTCC TTAGT TACCAGCACCTCGGGTGGGCACTCTAAGGAGACTGCCGGTGCACAAACCGGAGG AAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGGCCAGGGCTACACACGTGCTACAATGGTCGGTACA AAG
GTTGCCAAGCCGCGAGGTGGAGCTAATCCCATAAAACCGATCGTAGTCCGGATCGCAGTCTGCAACTCGACTCGGAAAGTTCGGAAATCG

9o paso



VECTORES UTILIZADOS PARA LA CONSTRUCCIÓN DE BIBLIOTECAS GENÓMICAS

Table 1 Properties of the vectors used in construction of genomic libraries

Pequeño 
Tamaños
grandes 

Vector	Cloning capacity (kb)	Applications
Plasmids	0.1–12	Single gene cloning; shotgun sequencing libraries
Bacteriophage lambda	10–20	Single gene cloning
Cosmids	35–45	Single gene cloning; genome mapping and sequencing
Bacteriophage P1	30–90	Genome mapping and sequencing
BACs	30–300	Genome mapping and sequencing
YACs	100–1000	Genome mapping and sequencing

Tomado de Lu, Q. (2005) Trends in Biotechnology. 23: 199-206

vectores derivados del fago λ o cósmidos

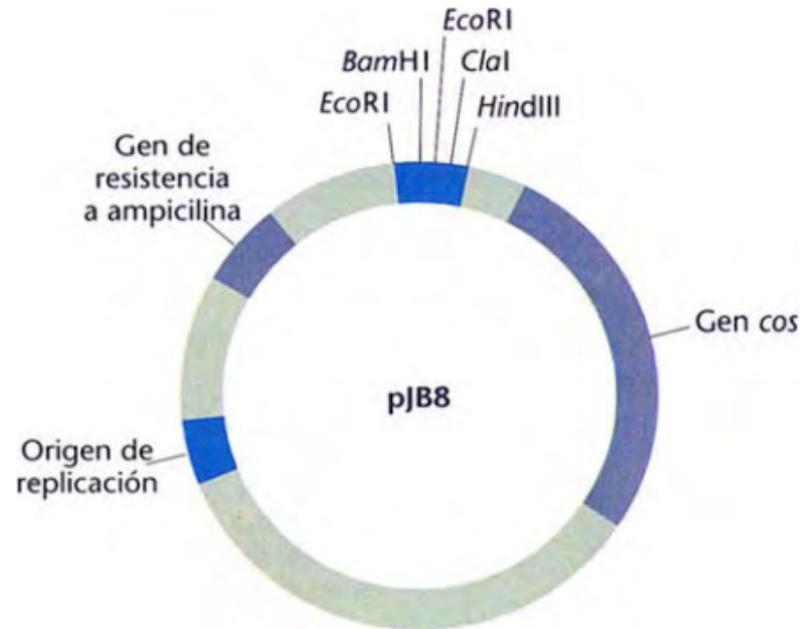
BACs (Bacterial Artificial Chromosomes)

YACs (Yeast Artificial Chromosomes)

Los cósmidos y los vectores transbordadores

Vectores híbridos contruidos utilizando partes del cromosoma del fago lambda y de ADN plasmídico

Se replica como un plásmido



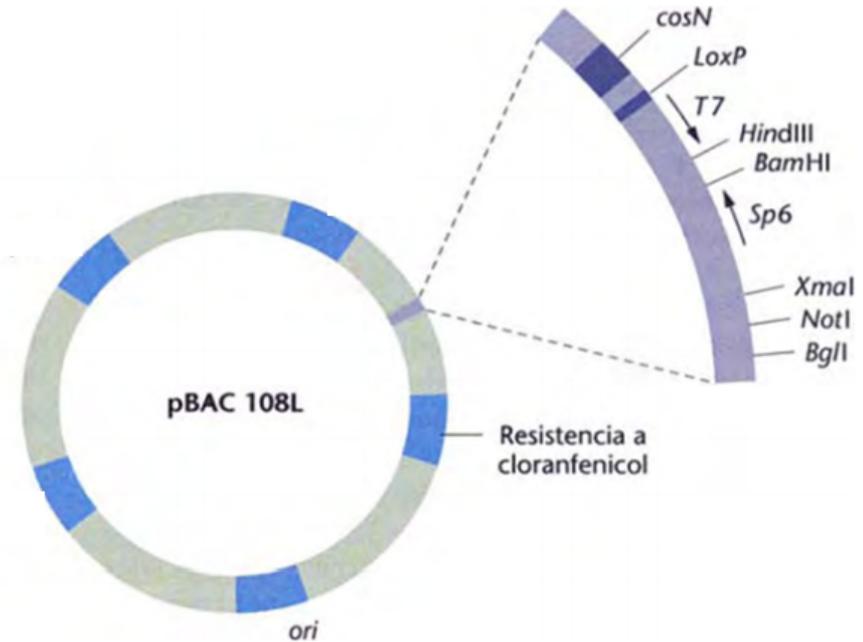
Pueden transportar insertos de ADN mucho más grandes casi 50kb

Cromosomas artificiales bacterianos (BACs)

Para analizar genomas eucarióticos complejos

(BAC) se basado en el factor F de bacterias, pueden transportar fragmentos del cromosoma bacteriano de hasta 1Mb,

Sitio de donación está flanqueado por regiones promotoras que pueden utilizarse para generar moléculas de ARN y para secuenciar el ADN del inserto clonado.



Resumen

- 1.- Pros y contras del plaque**
- 2.- Gen 16S rARN para propósitos taxonómicos**
- 3.- ¿Quién hace clones aún?**

Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads^{∇†}

Andrea K. Bartram,¹ Michael D. J. Lynch,² Jennifer C. Stearns,¹
Gabriel Moreno-Hagelsieb,² and Josh D. Neufeld^{1*}

Department of Biology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1,¹ and Department of Biology, Wilfrid Laurier University, Waterloo, Ontario, Canada N2L 3C5²

Received 25 November 2010/Accepted 24 March 2011

JOURNAL DETAILS:

IMPACT FACTOR: 3.668

5 YEAR IMPACT FACTOR: 4.359

EIGENFACTOR: 0.10906

#1 CITED JOURNAL IN MICROBIOLOGY

#1 CITED JOURNAL IN BIOTECHNOLOGY & APPLIED
MICROBIOLOGY



Introducción

Profundidad del muestreo

```
graph TD; A[Profundidad del muestreo] --> B[Desarrollaron un método para secuenciar y ensamblar millones de lecturas de extremos pares del gen 16S rARN]; B --> C[Región V3 (200 nt)]; C --> D[Illumina genome analyzer.]; E[Preparación de bibliotecas de secuencias por duplicado] --> B; F[Incorpora primes indexados para permitir la caracterización de múltiples comunidades microbianas en una sola celda de flujo] --> B; G[Económico] --> B;
```

Desarrollaron un método para secuenciar y ensamblar millones de lecturas de extremos pares del gen 16S rARN

Región V3 (200 nt)

Illumina genome analyzer.

Preparación de bibliotecas de secuencias por duplicado

Incorpora primes indexados para permitir la caracterización de múltiples comunidades microbianas en una sola celda de flujo

Económico

Materiales y métodos

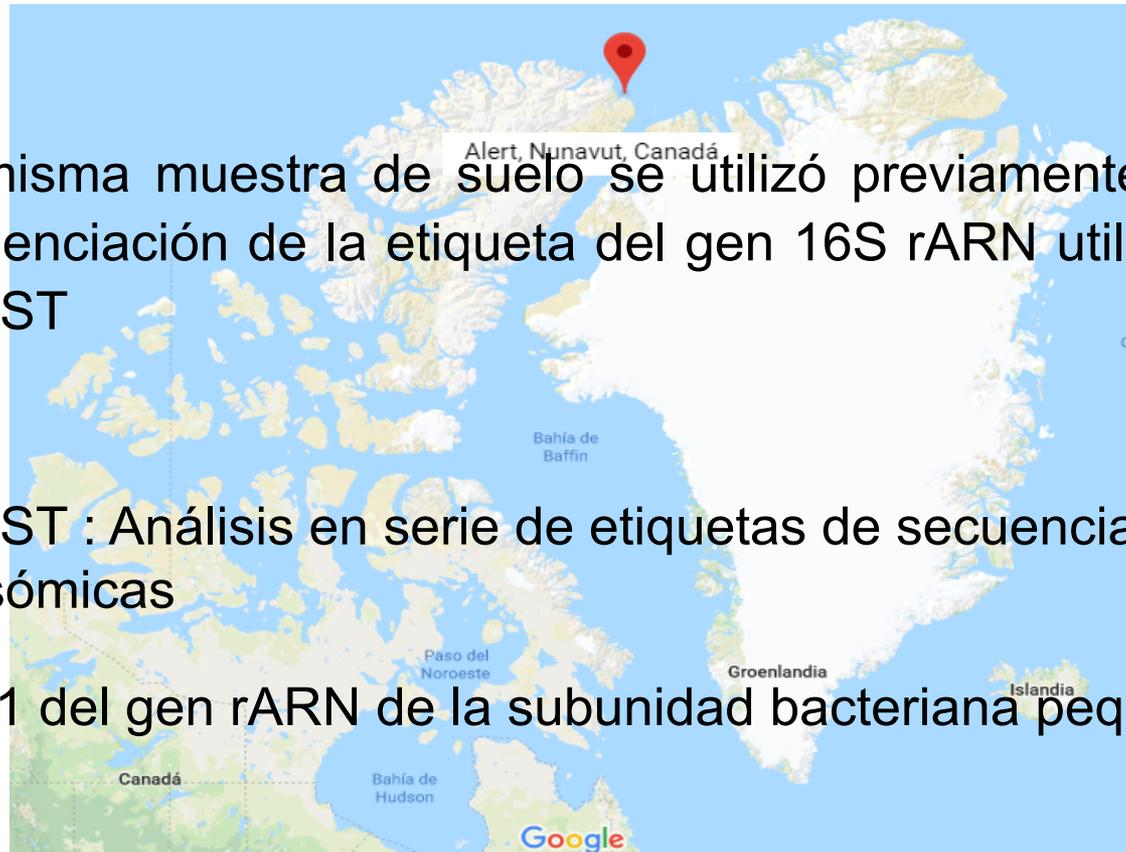
Toma de muestras y aislamiento de ADN

Una muestra compuesta de suelo de tundra ártica

La misma muestra de suelo se utilizó previamente para la secuenciación de la etiqueta del gen 16S rARN utilizando el SARST

SARST : Análisis en serie de etiquetas de secuencias ribosómicas

V1 del gen rARN de la subunidad bacteriana pequeña



"barcode"

Para una comunidad definida, se seleccionaron seis cepas bacterianas como controles:

- 1) *Escherichia coli* (ATCC 11303)
- 2) *Pseudomonas aeruginosa* (ATCC 10145)
- 3) *Bacillus subtilis* (ATCC 6633)
- 4) *Flexibacter canadensis* (ATCC 29591)
- 5) *Methylococcus capsulatus* Bath (ATCC 33009)
- 6) *Paracoccus denitrificans* (ATCC 17741)

Extracción de ADN genómico del suelo y de cultivos bacterianos



Fast Extraído por triplicado y agrupados)

Generación de la biblioteca Illumina:

La región V3 del gen 16S rARN se amplificó utilizando primers 341F y 518R

Forward Primers	Sequence (5' to 3')
V3_F	aatgatacggcgaccaccgagatct <u>acactctttccctacacgacgctcttccgatct</u> CCTACGGGAGGCAGCAG
V3_F modified ²	aatgatacggcgaccaccgagatct <u>acactctttccctacacgacgctcttccgatct</u> NNNNCCTACGGGAGGCAGCAG
Reverse Primers	
V3_1R	caagcagaagacggcatacagagat CGTGAT <u>gtgactggagttcagacgtgtgctcttccgatct</u> ATTACCGCGGCTGCTGG
V3_2R	caagcagaagacggcatacagagat ACATCG <u>gtgactggagttcagacgtgtgctcttccgatct</u> ATTACCGCGGCTGCTGG
V3_3R	caaqaqaagacqqcatacqaqat GCCTAA <u>gtgactggagttcagacgtgtgctcttccgatct</u> ATTACCGCGGCTGCTGG

caagcagaagacggcatacagagat**CGTGAT**gtgactggagttcagacgtgtgctcttccgatctATTACCGCGGCTGCTGG

Enlace de la celda de flujo

Index - barcode

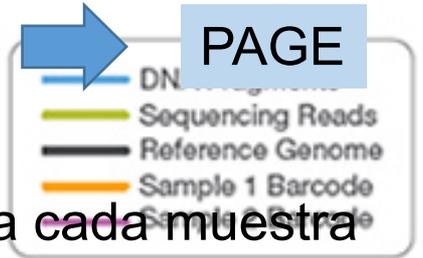
Illumina primers

Primers de la región V3

Total de primers: 25

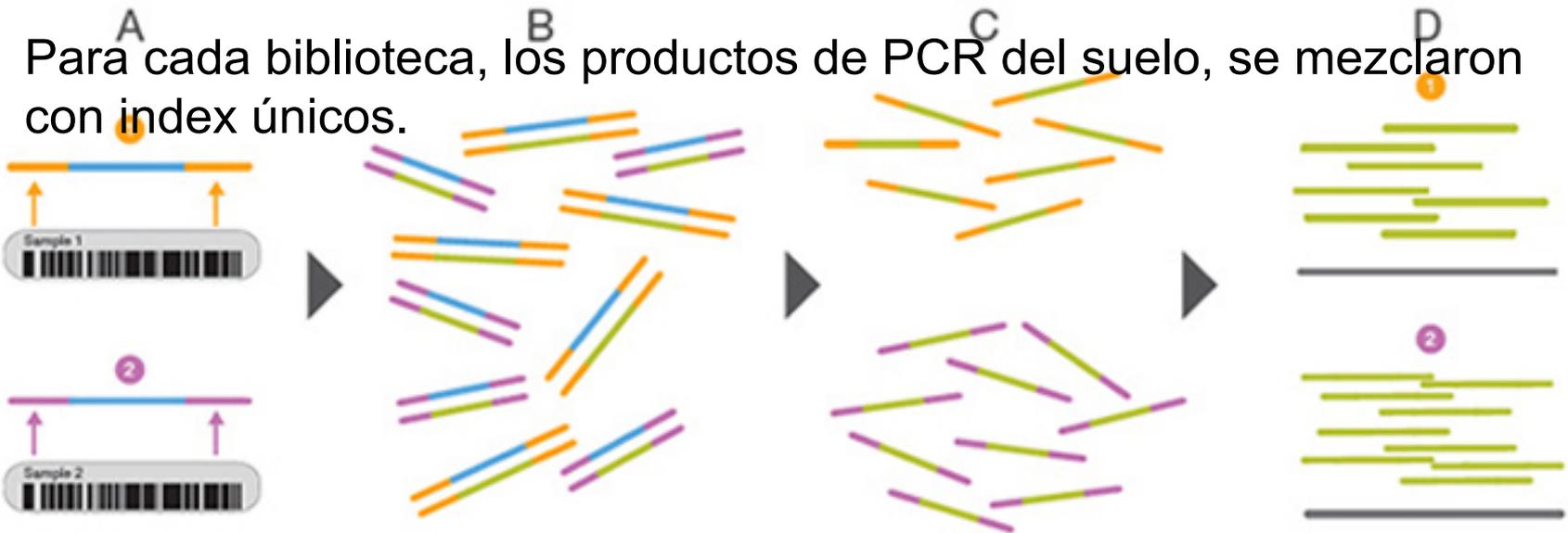
Generación de la biblioteca Illumina:

Primers personalizados se sintetizaron y purificaron
Index- barcode → permitiendo multiplexing



Se llevaron a cabo tres amplificaciones por PCR para cada muestra

Para cada biblioteca, los productos de PCR del suelo, se mezclaron con index únicos.



- Two representative DNA fragments from two unique samples, each attached to a specific barcode sequence that identifies the sample from which it originated.
- Libraries for each sample are pooled and sequenced in parallel. Each new read contains both the fragment sequence and its sample-identifying barcode.
- Barcode sequences are used to de-multiplex, or differentiate reads from each sample.
- Each set of reads is aligned to the reference sequence.

Generación de la biblioteca Illumina:

ADN no definido > ADN definido

La biblioteca fue amplificada → generación de clusters (Illumina, versión 4)

biblioteca resultante → secuenciación → reads pares de 125 bp Illumina
GAIIx

Bibliotecas de clones

ADN Genómico: suelo + cultivo puro → Plantillas: Primers 27f y 1492 r



Gen 16S rARN completo



Se realizaron amplificaciones por PCR



Los productos de reacción se clonaron en el vector TOPO

Selección de 95 clones positivos de cada biblioteca (ya sea suelo o biblioteca de cultivo puro)



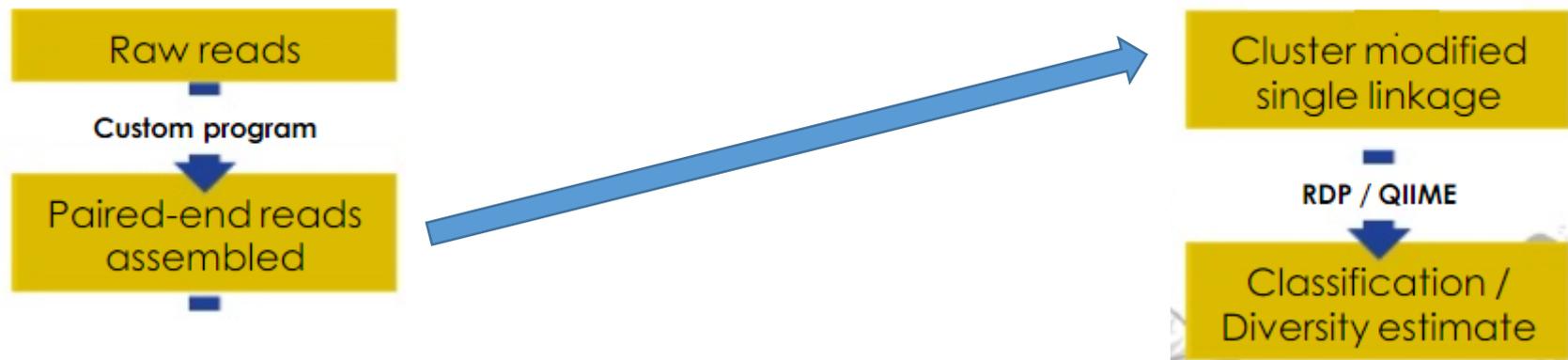
Secuenciado con Sanger technology

Filtrado de calidad inicial

Algoritmo personalizado → PANDAseq

Los reads se agruparon según el index → reads de extremos pares se alinearon → formado “contings” → descarte de secuencias ambiguas

Análisis bioinformático.



Resultados

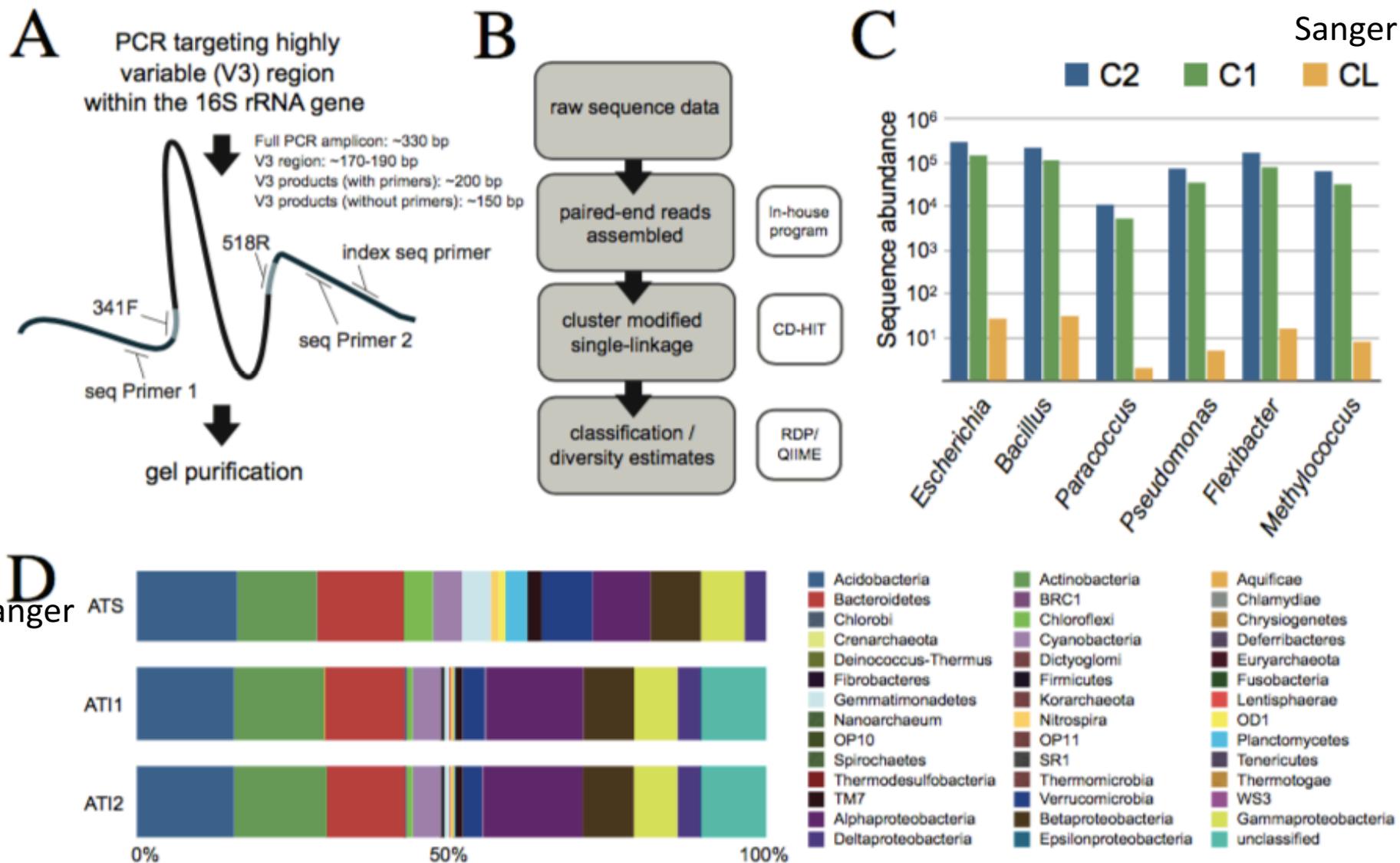


FIG. 1. Overview of the Illumina 16S rRNA gene sequencing method and generated library data. (A) The schema indicates a PCR (20 cycles) and gel purification of ~330-base PCR products, including the conserved 16S rRNA gene primer-binding region. (B) Informatics pipeline for generating clusters and taxonomic affiliations. (C) Resulting taxonomic affiliations for the replicate control libraries (C1 and C2) and the Sanger sequencing-based library (CL). (D) Taxonomic affiliations for the Alert tundra duplicate libraries (AT1 and AT2) and the Sanger sequencing-based library (ATS).

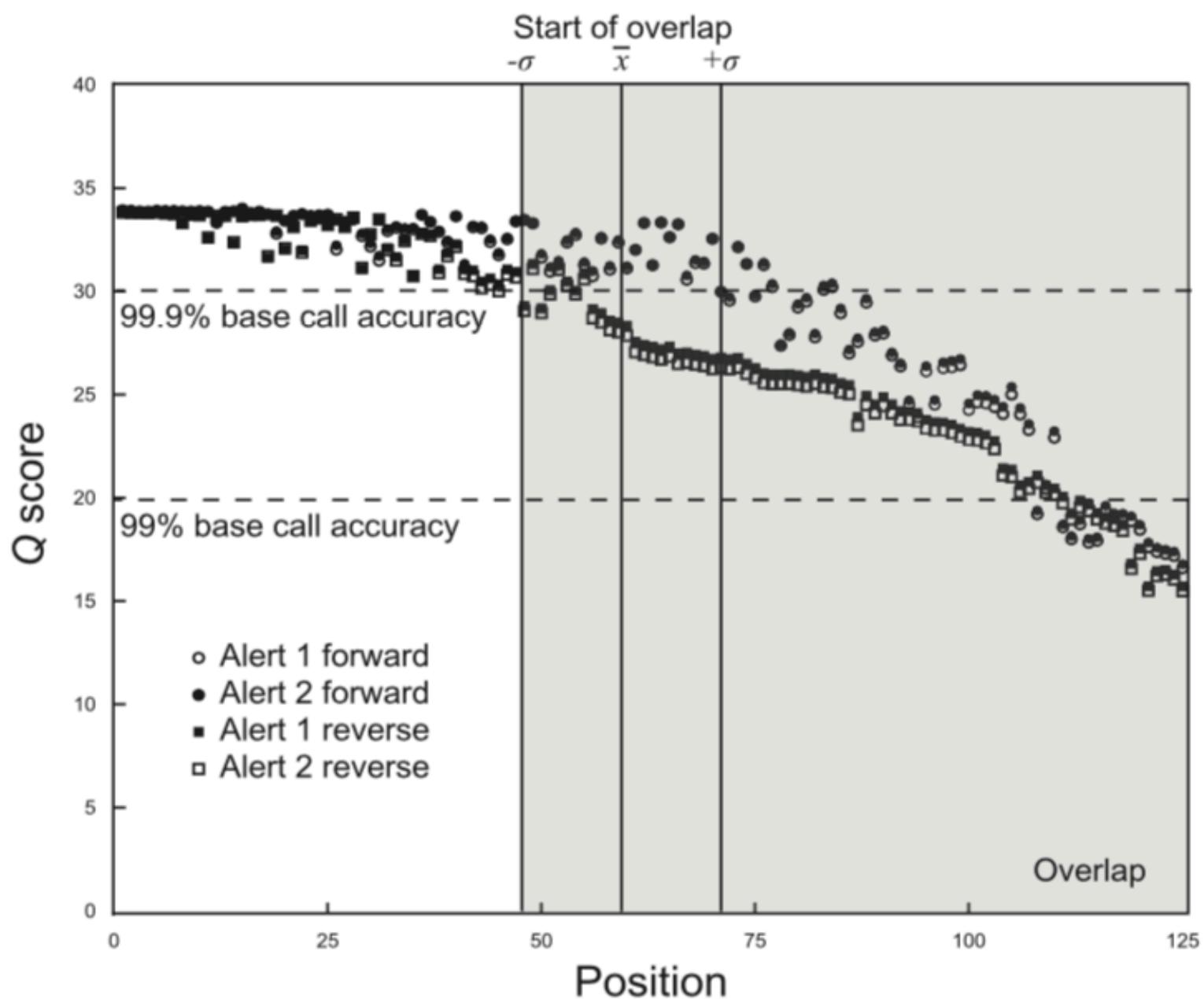


FIG. 2. Quality (Q) scores for all 125-base sequence reads. The Q score is an integer mapping of P , the probability that the corresponding base call is incorrect, with higher Q scores indicating lower error rates. The magnitude of sequence overlap for each assembled read was characterized, and the mean (\bar{x}) and standard deviation ($\pm\sigma$) were plotted relative to sequence length. The region of potential read overlap as presented does not explicitly calculate the additive Q score at each position, as the range of overlap varied due to the large range of V3 lengths.

TABLE 1. Counts of paired-end rRNA gene sequences obtained from the Illumina flow cell (preassembly) and following assembly (postassembly) for the replicate libraries included in this study

Library	No. of sequences		Remaining sequences (%)
	Preassembly	Postassembly	
AT1	7,570,249	4,073,963	53.8
AT2	4,371,453	2,396,331	54.8
C1	716,366	464,045	64.8
C2	1,350,602	842,585	62.4

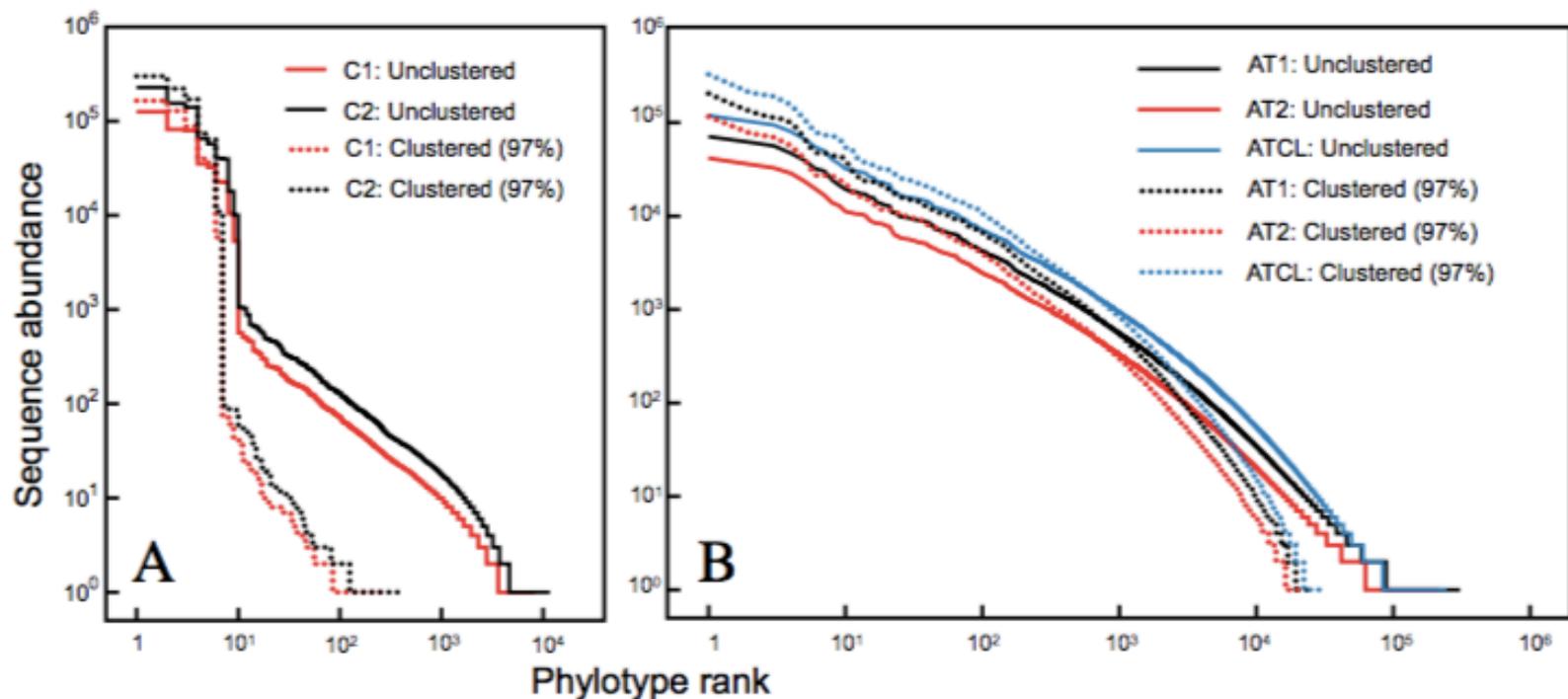


FIG. 3. Rank-abundance curves for duplicate control libraries (A) and Alert Arctic tundra libraries (B). The data shown are the raw data and also the data clustered using CD-HIT at a cutoff of 97%. Note that the Alert Illumina library was considered as separate replicates (AT1 and AT2) and also as a composite library (ATCL), which represents the combined replicates.

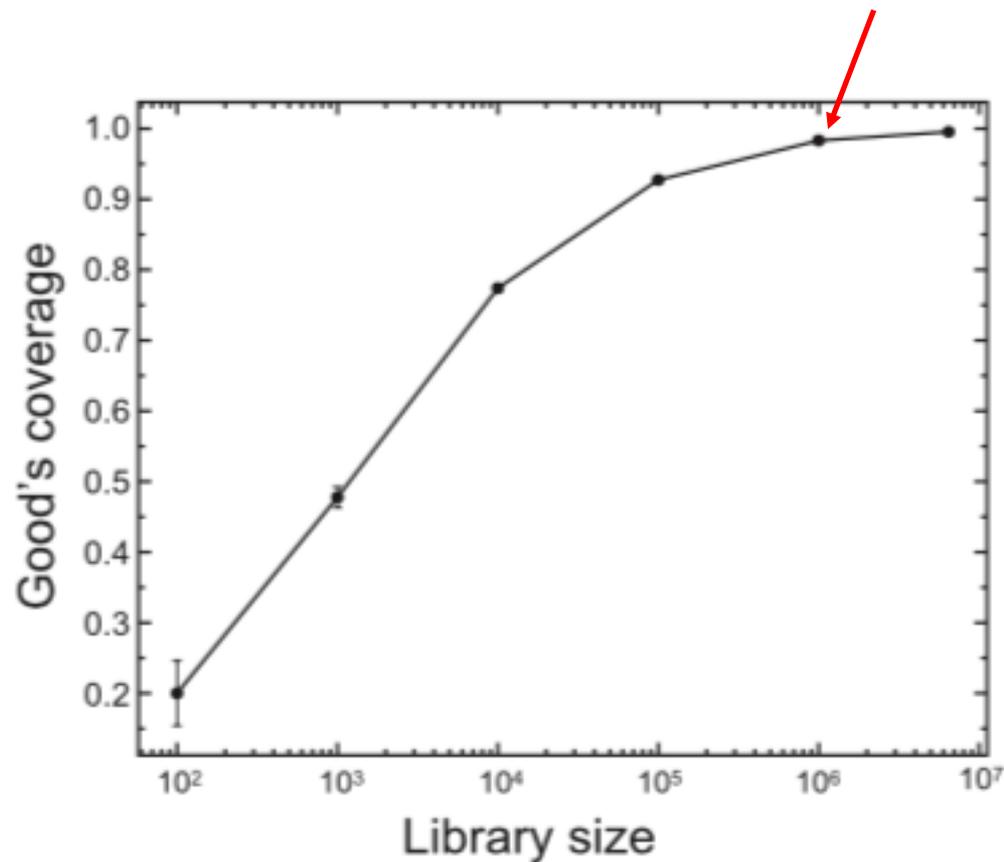
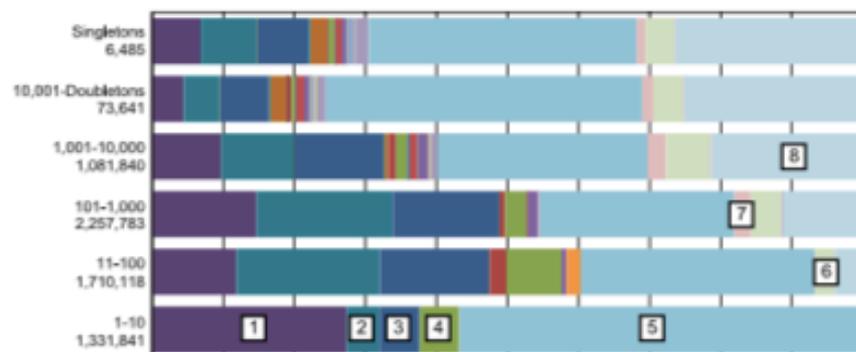


FIG. 4. Effect of library size on phylotype coverage. Randomly subsampled libraries were drawn in triplicate from combined AT libraries and used to calculate Good's coverage estimates. Averages for triplicates were plotted with standard deviations.

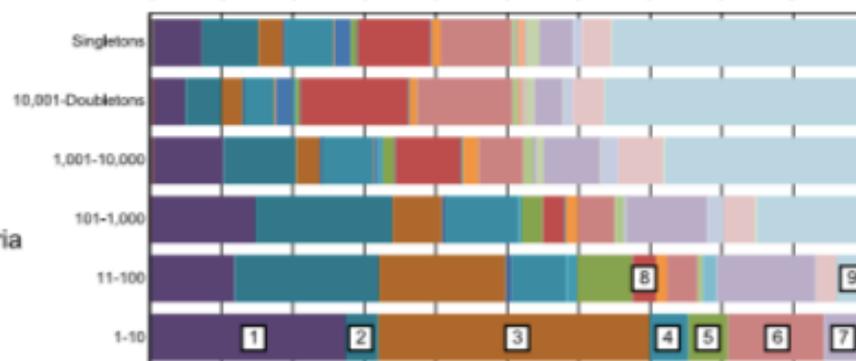
Phylum

- 1 Acidobacteria
- 2 Actinobacteria
- 3 Bacteroidetes
- 4 Cyanobacteria
- 5 Proteobacteria
- 6 Verrucomicrobia
- 7 TM7
- 8 Unclassified



Class

- 1 Acidobacteria
- 2 Actinobacteria
- 3 Alphaproteobacteria
- 4 Betaproteobacteria
- 5 Cyanobacteria
- 6 Gammaproteobacteria
- 7 Sphingobacteria
- 8 Deltaproteobacteria
- 9 Unclassified



Order

- 1 Acidobacteriales
- 2 Burkholderiales
- 3 Chromatiales
- 4 Family I (Cyanobacteria)
- 5 Rhizobiales
- 6 Sphingobacteria
- 7 Sphingomonadales
- 8 Unclassified

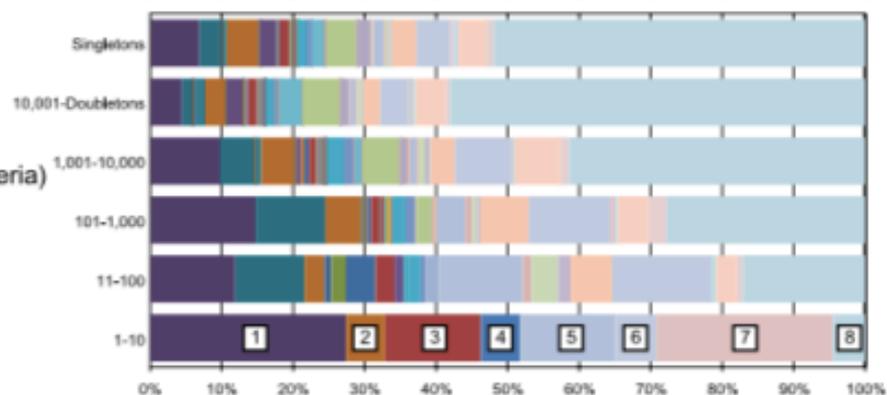


FIG. 5. Taxonomic affiliations at the levels of phylum, class, and order for consecutive abundance ranks of sequence data clustered at 97% with CD-HIT. Predominant taxa are represented in the bottom row, and singletons are at the top for each taxonomic level. Full details of RDP affiliations are summarized in Tables S3, S4, and S5 in the supplemental material.

Supplementary Table S3. Taxonomic affiliations of phyla associated with distinct abundance ranks shown in Figure 4 for the combined Arctic tundra library. Numbers in brackets represent the phylum proportion within the total library size for each rank (%).

Taxonomic affiliation	Abundance ranks					
	1-10	11-100	101-1000	1001-10000	10001-Doubletons	Singletons
Acidobacteria	365179 (27.4)	203500 (11.9)	332280 (14.7)	105752 (9.8)	3280 (4.5)	444 (6.8)
Actinobacteria	60444 (4.5)	344566 (20.1)	431493 (19.1)	109524 (10.1)	3689 (5.0)	509 (7.8)
Aquificae				209 (<0.05)	45 (0.1)	3 (<0.05)
Bacteroidetes	76038 (5.7)	263945 (15.4)	340140 (15.1)	137636 (12.7)	5044 (6.8)	477 (7.4)
BRC1				825 (0.1)	196 (0.3)	12 (0.2)
Chlamydiae				5920 (0.5)	1710 (2.3)	144 (2.2)
Chlorobi					11 (<0.05)	
Chloroflexi		40305 (2.4)	17990 (0.8)	11097 (1.0)	418 (0.6)	16 (0.2)
Cyanobacteria	73754 (5.5)	132774 (7.8)	67597 (3.0)	17804 (1.6)	447 (0.6)	59 (0.9)
Deinococcus-Thermus				1248 (0.1)	43 (0.1)	3 (<0.05)
Dictyoglomi				142 (<0.05)	13 (<0.05)	
Euryarchaeota					6 (<0.05)	2 (<0.05)
Fibrobacteres			1822 (0.1)	1716 (0.2)	49 (0.1)	5 (0.1)
Firmicutes			3762 (0.2)	12055 (1.1)	785 (1.1)	53 (0.8)
Fusobacteria				22 (<0.05)	3 (<0.05)	1 (<0.05)

Supplementary Table S4. Taxonomic affiliations of classes associated with distinct abundance ranks shown in Figure 4 for the combined Arctic tundra library. Numbers in brackets represent the class proportion within the total library size for each rank (%).

Taxonomic affiliations	Abundance Rank					
	1-10	11-100	101-1000	1001-10000	10001-Doubletons	Singletons
"Bacilli"				2801 (0.3)	178 (0.2)	15 (0.2)
"Clostridia"			2229 (0.1)	3524 (0.3)	255 (0.3)	17 (0.3)
"Erysipelotrichi"						1 (<0.05)
Acidobacteria	365179 (27.4)	203500 (11.9)	332280 (14.7)	105752 (9.8)	3280 (4.5)	444 (6.8)
Actinobacteria	60444 (4.5)	344566 (20.1)	431493 (19.1)	109524 (10.1)	3689 (5.0)	509 (7.8)
Alphaproteobacteria	505636 (38.0)	303606 (17.8)	157245 (7.0)	35290 (3.3)	2108 (2.9)	219 (3.4)
Anaerolineae		15823 (0.9)	8916 (0.4)	2987 (0.3)	48 (0.1)	7 (0.1)
Aquificae				209 (<0.05)	45 (0.1)	3 (<0.05)
Archaeoglobi						
Bacteroidetes			1574 (0.1)	388 (<0.05)	65 (0.1)	8 (0.1)
Betaproteobacteria	72205 (5.4)	129995 (7.6)	231774 (10.3)	78076 (7.2)	3117 (4.2)	439 (6.8)
BRC1_genera_incertae_sedis				825 (0.1)	196 (0.3)	12 (0.2)
Chlamydiae				5920 (0.5)	1710 (2.3)	144 (2.2)
Chlorobia					11 (<0.05)	
Chloroflexi		24482 (1.4)	9074 (0.4)	7348 (0.7)	336 (0.5)	8 (0.1)
Chrysiogenetes						

Supplementary Table S5. Taxonomic affiliations of orders associated with distinct abundance ranks shown in Figure 4 for the combined Arctic tundra library. Numbers in brackets represent the order proportion within the total library size for each rank (%).

Taxonomic affiliations	Abundance Rank					
	1-10	11-100	101-1000	1001-10000	10001-Doubletons	Singletons
"Erysipelotrichales"						1 (<0.05)
"Lactobacillales"				258 (<0.05)	27 (<0.05)	4 (0.1)
"Thermoanaerobacterales"			1352 (0.1)	86 (<0.05)	6 (<0.05)	1 (<0.05)
Acidobacteriales	365179 (27.4)	203500 (11.9)	332280 (14.7)	105752 (9.8)	3280 (4.5)	444 (6.8)
Actinobacteridae		165436 (9.7)	214794 (9.5)	51873 (4.8)	1060 (1.4)	186 (2.9)
Aeromonadales				99 (<0.05)	12 (<0.05)	
Alteromonadales					5 (<0.05)	
Aquificales				209 (<0.05)	45 (0.1)	3 (<0.05)
Bacillales				2133 (0.2)	132 (0.2)	10 (0.2)
Bacteroidales			1574 (0.1)	388 (<0.05)	65 (0.1)	8 (0.1)
Bdellovibrionales			2945 (0.1)	7033 (0.7)	1012 (1.4)	35 (0.5)
Burkholderiales	72205 (5.4)	48516 (2.8)	112516 (5)	51030 (4.7)	2044 (2.8)	291 (4.5)
Caldilineae		15823 (0.9)	6123 (0.3)	2624 (0.2)	46 (0.1)	7 (0.1)
Campylobacterales				666 (0.1)	23 (<0.05)	3 (<0.05)
Caulobacterales		34564 (2)	5145 (0.2)	277 (<0.05)	88 (0.1)	3 (<0.05)
Chlamydiales				5920 (0.5)	1710 (2.3)	144 (2.2)
Chlorobiales					11 (<0.05)	

Discusión y conclusiones

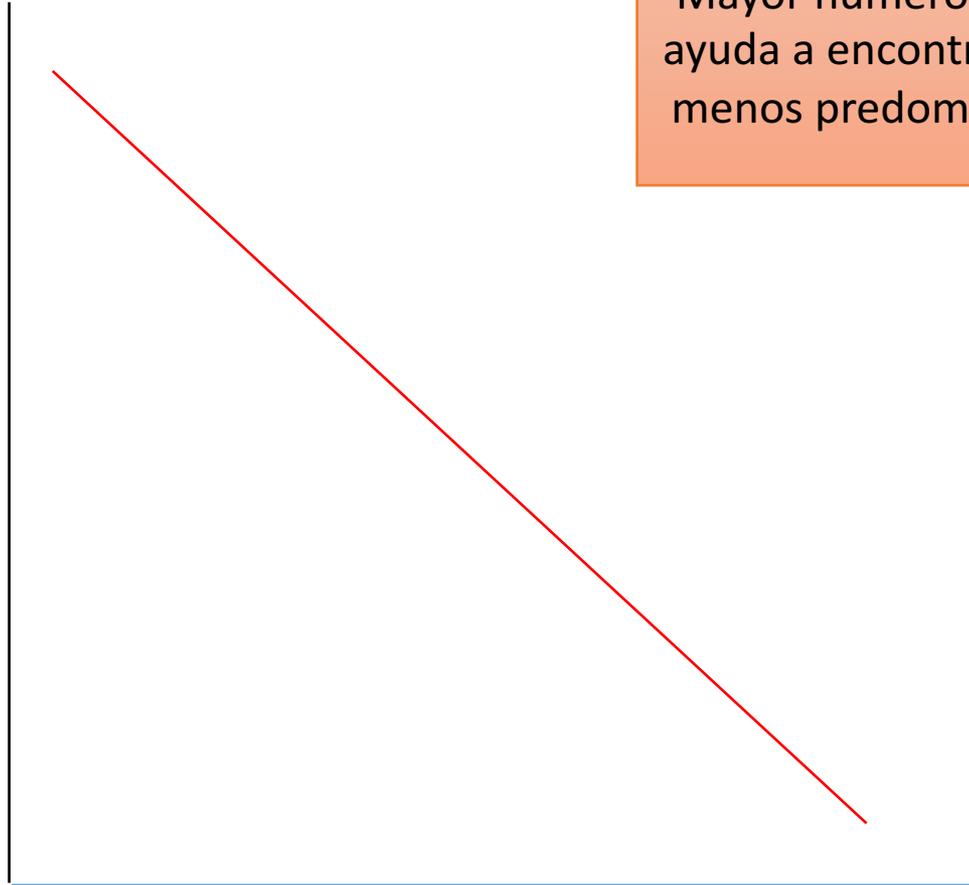
- La calidad de las librerías generadas en Illumina es directamente proporcional a la longitud de las lecturas (~ 125 bases) que se ensamblen correctamente
- Agregar bases adyacentes al primer forward, con el fin de mejorar la secuenciación

El algoritmo identifica más rápidamente los clusters en la celda fluida con mayor diversidad nucleotídica dentro de las 4 primeras bases

Divergencia

No. de secuencias

Mayor número de secuencias ayuda a encontrar poblaciones menos predominantes o raras



Con este método se puede comparar y caracterizar taxa de baja y alta abundancia, debido a que:

- Los primers y adaptadores son modulares
- El algoritmo que utiliza RDP incrementa la profundidad de estudio de especies de baja abundancia.