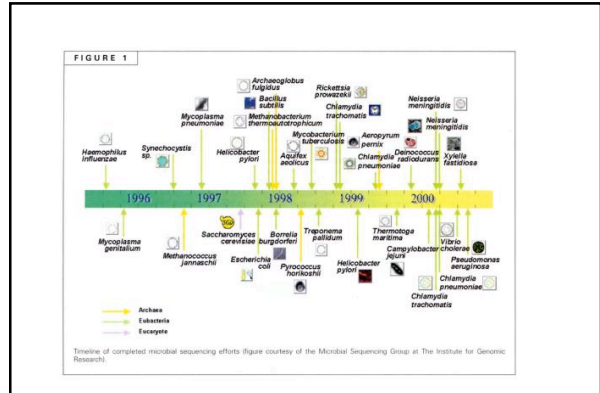
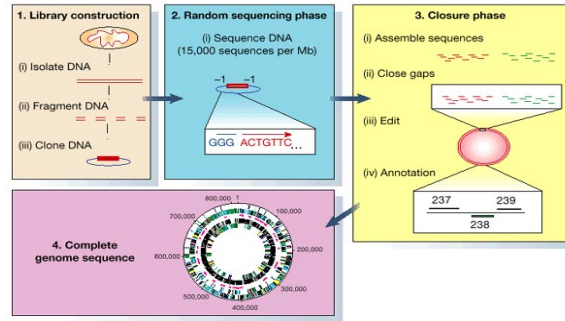


Genomics and marine microbes

- DNA sequencing, then and now...
- Basics of microbial genome analyses
- Marine microbial genomes ...



Fraser, Eisen, and Salzman
Nature **406**, 799-803 (17 August 2000) **Shotgun genome sequencing**

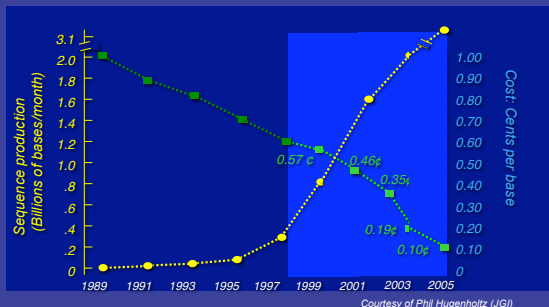


Chromatograms from PJ-C8
 eos_PJL_30601954_29.pl Fragment size: 4125 Size: 125 of 181

1. Trace files (dye signals) are analyzed and bases called to create chromatograms.

2. Chromatograms from opposite strands are reconciled with software to create double-stranded sequence data.

Sanger sequencing has become much cheaper
 Joint Genome Institute (JGI) statistics




New technologies...
454 DNA sequencing

500,000 DNA fragments sequenced at once

- Average read length 250 nt
- 100 million nt per run
- One 454 run = 4 hours
- < \$10,000 reagent cost

Pyrosequencing - 454 GS20 and FLX



OLD:

Genome Sequencer GS20
 ~ 40 million bases/run
 ~ >100bp
 ~ 400,000 reads

NEWER:

Genome Sequencer FLX
 ~ 100 million bases/run
 ~ >250bp
 ~ 400,000 reads

NEWEST: "Titanium"
 ~ 250 million bases/run
 ~ > 450 bp
 ~ 500,000 reads

Pyrosequencing - Library construction

gDNA
cDNA

1. DNA Library Construction* (4.5 h)

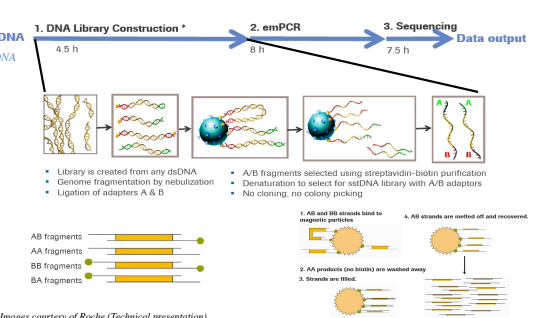
- Library is created from any dsDNA
- Genome fragmentation by nebulization
- Ligation of adapters A & B

2. emPCR (8 h)

- A/B fragments selected using streptavidin-biotin purification
- Denaturation to select for ssDNA library with A/B adaptors
- No cloning, no colony picking

3. Sequencing (7.5 h)

Data output



Images courtesy of Roche (Technical presentation)

Pyrosequencing - emPCR

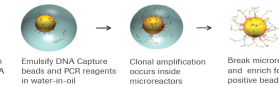
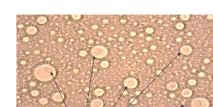
Anneal sstDNA to an excess of 28 µm DNA Capture beads

Emulsify DNA Capture beads and PCR reagents in water-in-oil microreactors

Clonal amplification occurs inside microreactors

Break microreactors and enrich for DNA-positive beads

sstDNA library → Clonally-amplified sstDNA attached to bead

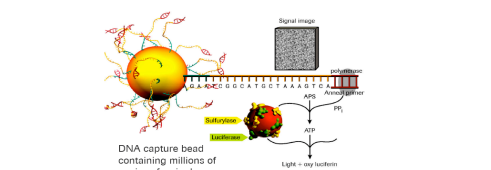
DNA Capture Beads

Pyrosequencing - sequencing

Amplified sstDNA library beads

- Well diameter average for PicoTiterPlate is 44 µm
- A single clonally amplified sstDNA bead is deposited per well.
- A layer of packing and enzyme beads are deposited
- Plate is loaded into instrument for sequencing

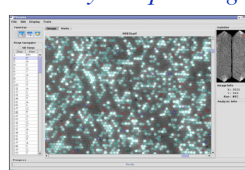
→ **Packed PTP**

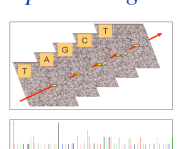


DNA capture bead containing millions of copies of a single clonal fragment

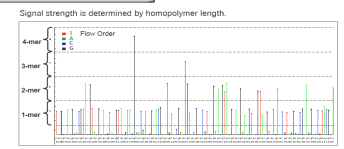
Light = ray location

Pyrosequencing - imagine processing





Signal strength is determined by homopolymer length.



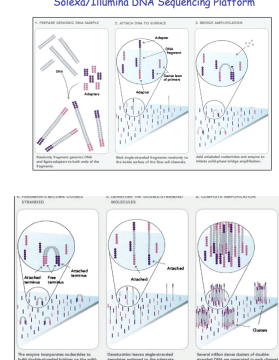
TCAG

Solexa technology

Solexa/Illumina DNA Sequencing Platform

The sequencing of the "DNA clusters" relies on reversible terminators (like dideoxys) on fluorescently labelled nucleotides. Each round of synthesis gives the next base in sequence, by its color.

(sequence by synthesis)



Reversible Terminator Chemistry

a

b

10 million clusters per 100 microns

Bentley et al. Nature 456, 53-, 2008.

Solexa/Illumina DNA Sequencing Platform

1. DETERMINE FIRST BASE

2. IMAGE FIRST BASE

3. DETERMINE SECOND BASE

~ 70 bp reads

7 million reads/lane !

Solid Technology (ABI)

Next Gen Platform Comparisons

| | AB3730 | 454 FLX | Solexa/Illumina |
|-------------------------------|------------------------|--------------------|----------------------|
| Cost per run | ~\$50 | ~\$12K | ~\$5K |
| Bases read/run | 72Kb | 100Mb | 720Mb |
| Bases per read | 750 | 250 | 36 |
| Reads per run | 96 reads/run | 400K reads/run | 20M reads/run |
| \$ per Mbp | \$ 694 | \$ 120 | \$ 7 |
| AB3730 work equivalent | - | 100x AB3730/dy | 300x AB3730/dy |
| Errors | Diverse (cloning bias) | Homopolymeric runs | Diverse (base subn.) |
| Run time | 1 hour | 6.5 hours | 2-4 days* |

Pacific Bioscience's single molecule DNA sequencing strategy (2010)

Attenuated light from the excitation beam in a detection volume of 20 zeptoliters (10^{-21} liters).

Figure 2.1. Highly parallel single molecule sequencing. The detected flash of light is captured into a optical fiber, from which the identity of the incorporated base is determined.

SUMMARY CURRENT 'NEXT GEN' TECHNOLOGIES

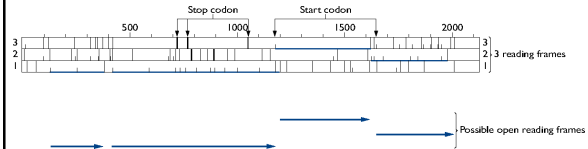
| Platform | Amplification | Sequencing |
|-------------------------------|---------------|------------------------------------|
| Harvard/Danaher/Agencourt/ABI | Emulsion PCR | Ligase |
| 454 Life Sciences / Roche | Emulsion PCR | Polymerase - pyrosequencing |
| Solexa/Illumina | Bridge PCR | Polymerase - reversible terminator |
| Helicos | None | Polymerase - single base extension |
| Pacific Biosciences | None | Polymerase - active site occupancy |

How do you make sense of this ????????

```

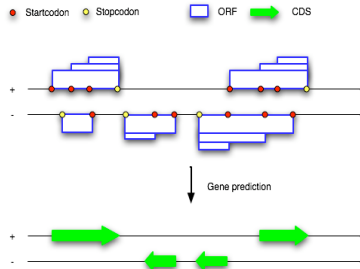
TTCAATTGCCAAATCCACTAGATGAATGATTCTGAATCAGATATAAAATTACGTAATTGTAATAAGA
ATTGAGTTTAAACACATTATGGAAAAAGGATTGATTTTCATGTTGATTTTGACTATTTTATGCACA
GTGTGAAGAGATTAGAGATTCTAAATATAAGGGAAATGTGTGACAGTTGTATATTTCTGATAGAGGA
GGAGATAGTGGAGCAATAGCTACTTGCAAATACAAATGCAAGAAACTTTGGAGTAAATCAGGAATTC
TAATGCTTGCATAAAATAAATAAAGGAGCAAGATTCAATATTTTGGCAGCTGATTTGATTAATTC
AGAAGTATCATCAAAGCAATGGAAATATTGAAAGTATGCAGATGATTTGATATGTTGGAAGAGAT
GAAGCATATCTGTGTTACAAAAAACTGAGAGTATTTCATATATGCGAAGACATAGCGCAACAGT
TGAATAATGAAATAAGAAATAGTCTAAACTACATGTTCTGTAGGGATCAGCGCAATAAACTACTTTC
AAAAATGCTTCAGATTATAAGAACAGATGGATTGCAACTGTAAACCAGAAACAAGTTGATGAGTTT
TTATCACCATTAAAAATAAGAGTAATCCAGGTATAGGTAAAAAACAGAGATTTTTTTGTACAAATGC
ATGTGAATCAATAGAAGACCTAAGAGAAATTAATTTTTGATTTAAACAAAATGTTGGGAGAAAAAC
TGGAGGACATATTTTAAATTTCTTCAAGGAAATGTGATGAAGAAATGGTTAAACGAAAGCGCTACAAAT
CAGTTTAGCAAGATTATCACTTTAAAGAAAATTTCTAAAGAACTTAATTTAAGTAAACATAGAAAA
AATCTATGTGCAATTGAATGAATTTGCAATTTGAAAAACAATAATGATCGCTCAATTTGGAATCCAGTT
TCTAAACGAGGATTTTCCACAAGAACTAATCAAAAATGATAAAAAATCCAGGAAATAATGATTTGAA
TTAAAAAAGTTGTAATCAATCTACTAGAAGAAGCAITGATAGAACAAGAAATGTAGTTAGAAGAAITG
GAGTTAGAATTTTTCAGATTTTTCAGATGAGAGGGTCAGAGAGACATACAAATTAATTTAGATGTTTT
TCAGATTTCTATTTCGCTTGTGAGTCGTTTTGTTTCAACTCTGTCACAAATCAGTAAAAAGATGAATAACC
GAAGAAACATTATTTGCGCTGCTACAGAATGAAATTTCTCCATGCTTCAGGATCAGTATGCTACTTTAG
TGGTAAACAGATAACGAAAAGATTGCAATCAAAATTTACAAATATAGTTCCAAGAATTTCTAAAAATAAA
    
```

Determination of possible open reading frames in a DNA sequence.



A simplistic view...

ORF = open reading frame = start codon...[GATC]n.....stop codon
CDS = coding sequence (produces the actual protein/RNA species)



Annotation of bacterial and archaeal genomes: improving accuracy and consistency. Overbeek et al., Chem Rev. 2007. 107:3431-3447.

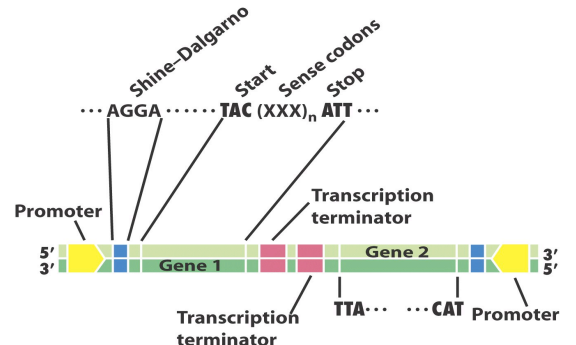
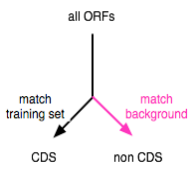


Figure 15-4 Brock Biology of Microorganisms 11/e © 2006 Pearson Prentice Hall, Inc.

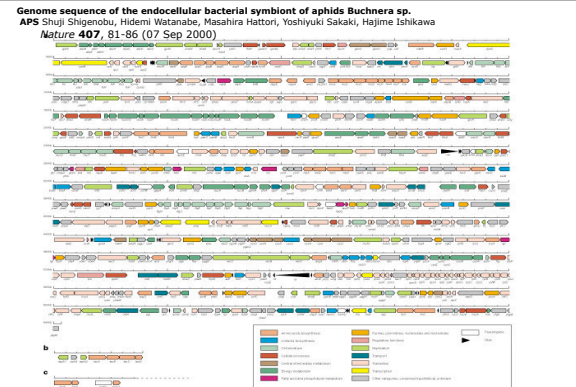
Computing the probability of GATC under a markov chain can be translated into:

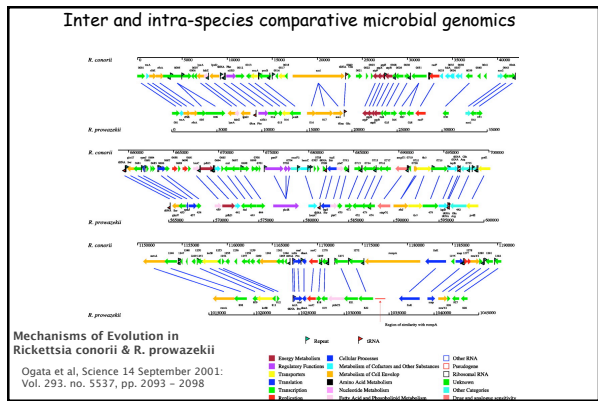
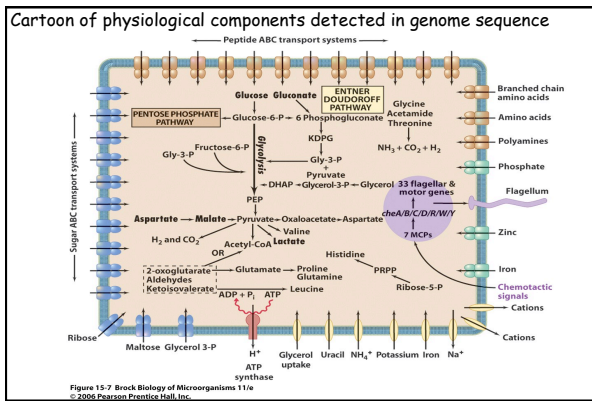
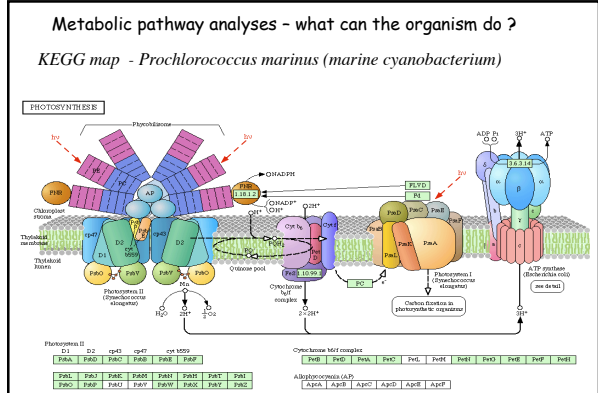
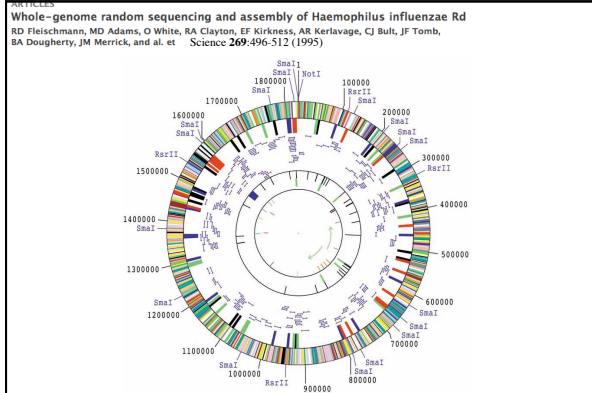
$$P_{\text{chain}}(\text{G}|\langle \text{startsymbol} \rangle) * P_{\text{chain}}(\text{A}|\text{G}) * P_{\text{chain}}(\text{T}|\text{A}) * P_{\text{chain}}(\text{C}|\text{T})$$



Factors affecting probabilities include :
 Different GC content in coding genes
 The GC content of the third position in a codon
 Nucleotide frequency
 Tri-nucleotide frequency
 Di-nucleotide frequency and other frequencies
 Presence of ribosomal binding sites (RBS)

Annotation of bacterial and archaeal genomes: improving accuracy and consistency. Overbeek et al., Chem Rev. 2007. 107:3431-3447.





Genome size in representative Bacteria

| Organism | Size (base pairs) | ORFs ^a | Comments |
|-----------------------------------|-------------------|-------------------|---|
| <i>Mycoplasma genitalium</i> | 980,070 | 470 | Smallest known cellular genome (ORF Section 12.21) |
| <i>Mycoplasma pranii</i> | 816,794 | 477 | <i>Causa pneumoniae</i> (ORF Section 12.21) |
| <i>Bradyrhizobium lotus</i> | 1,012,725 | 493 | <i>Symbiote</i> , has linear chromosomes, causes Lysine disease (ORF Sections 12.13 and 27.4) |
| <i>Chlamydia trachomatis</i> | 1,042,519 | 494 | Chlamydiae intracellular parasite, causes human pathogen (ORF Sections 12.27 and 26.13) |
| <i>Rickettsia prowazekii</i> | 1,111,533 | 454 | Chlamydiae intracellular parasite, causes epidemic typhus (ORF Sections 12.13 and 27.4) |
| <i>Thiomargarita</i> | 1,136,006 | 1041 | Symbiote, causes whiptail (ORF Sections 12.13 and 26.12) |
| <i>Agrobacterium tumefaciens</i> | 1,251,335 | 1544 | Hyperthermophilic, autotroph (ORF Section 12.27) |
| <i>Prochlorococcus marinus</i> | 1,657,699 | 1716 | Marine cyanobacterium, photosynthetic |
| <i>Halobacterium salinarum</i> | 1,667,467 | 1560 | Causes purple color (ORF Section 26.10) |
| <i>Staphylococcus aureus</i> | 1,852,414 | 1752 | Pathogen, causes staph and scalded liver (ORF Section 26.2) |
| <i>Haemophilus influenzae</i> | 1,860,725 | 1027 | Hyperthermophilic (ORF Section 12.26; See also Figure 15.7) |
| <i>Chlamydomonas reinhardtii</i> | 2,154,446 | 2288 | Model photosynthetic bacterium (ORF Section 12.22) |
| <i>Streptococcus pneumoniae</i> | 2,414,616 | 2503 | Major cause of pneumococcal infections (ORF Section 12.19 and 18.4) |
| <i>Deinococcus radiodurans</i> | 3,284,156 | 2185 | Resistant to radiation, multiple chromosomes (ORF Section 12.34) |
| <i>Synsphyxys sp.</i> | 3,272,470 | 3168 | Cyanobacterium (ORF Section 12.25) |
| <i>Helicobacter pylori</i> | 3,292,650 | 3364 | Relative resistant, multiple chromosomes (ORF Section 12.14) |
| <i>Geobacter sulfurreducens</i> | 3,314,139 | 3427 | Model for bioremediation (ORF Section 17.18) |
| <i>Conditiformis sp.</i> | 4,016,442 | 3527 | Complex life cycle (ORF Section 12.10) |
| <i>Bacillus subtilis</i> | 4,214,210 | 4109 | Gram-positive genetic model (ORF Section 12.20) |
| <i>Mycobacterium tuberculosis</i> | 4,411,259 | 3924 | Causes tuberculosis (ORF Section 12.13 and 26.5) |
| <i>Escherichia coli</i> | 4,639,221 | 4288 | Gram-negative genetic model (ORF Section 12.11) |
| <i>Bacillus anthracis</i> | 5,227,251 | 4758 | Pathogen, biothreat agent (ORF Section 12.28) |
| <i>Rhodospirillum rubrum</i> | 5,495,213 | 4856 | Metabolically versatile anoxygenic phototroph (ORF Section 12.8) |
| <i>Paracoccus aminovorans</i> | 6,264,403 | 5970 | Metabolically versatile opportunistic pathogen (ORF Section 12.2) |
| <i>Streptococcus carnosus</i> | 6,667,507 | 7825 | Has linear chromosome, produces antibiotics (ORF Section 12.24) |
| <i>Bradyrhizobium japonicum</i> | 9,105,828 | 8337 | N ₂ fixation, nodulates soybeans (ORF Section 19.22) |

^aIndicates on these and hundreds of other prokaryotic genomes can be found in the TIGR Database (www.tigr.org/tdb), a web site maintained by The Institute for Genomic Research (TIGR), Rockville, MD, a non-for-profit research institute, and at <http://www.genomesonline.org>. Links are listed there to other relevant web sites.

^bOpen reading frames. The purpose of reporting ORFs is to predict the total number of proteins that an organism might encode. Of course, genes encoding known proteins are included, as are all ORFs that could encode a protein greater than 100 amino acid residues. Smaller ORFs are typically not included unless they show similarity to a gene from another organism or unless the codon usage is typical of the organism being studied.

Table 15-1 part 2. Brock Biology of Microorganisms 11/e © 2006 Pearson Prentice Hall, Inc.

Genome size in representative Archaea

| Organism | Size (base pairs) | ORFs ^a | Comments |
|---|-------------------|-------------------|--|
| <i>Nanoarchaeum equitans</i> | 490,885 | 552 | Smallest known genome (ORF Section 13.11) |
| <i>Thermoplasma acidophilum</i> | 1,564,905 | 1509 | Thermophilic, acidophilic (ORF Section 13.5) |
| <i>Methanocaldococcus jannaschii</i> | 1,664,976 | 1738 | Methanogen (ORF Section 13.4) |
| <i>Archaeoglobus fulgidus</i> | 1,669,695 | 1841 | Hyperthermophilic (ORF Section 13.9) |
| <i>Pyrococcus horikoshii</i> | 1,738,505 | 2061 | Hyperthermophilic (ORF Section 13.6) |
| <i>Methanohalobium thermohalophilus</i> | 1,751,377 | 1855 | Methanogen (ORF Section 13.4) |
| <i>Halobacterium salinarum</i> | 2,178,400 | 2436 | Hyperthermophilic (ORF Section 13.7) |
| <i>Halobacterium salinarum</i> | 2,571,010 | 2630 | Extreme halophile, bacteriorhodopsin (ORF Section 13.3) |
| <i>Sulfolobus solfataricus</i> | 2,992,245 | 2977 | Hyperthermophilic, sulfur chemolithotroph (ORF Section 13.9) |

^aIndicates on these and hundreds of other prokaryotic genomes can be found in the TIGR Database (www.tigr.org/tdb), a web site maintained by The Institute for Genomic Research (TIGR), Rockville, MD, a non-for-profit research institute, and at <http://www.genomesonline.org>. Links are listed there to other relevant web sites.

^bOpen reading frames. The purpose of reporting ORFs is to predict the total number of proteins that an organism might encode. Of course, genes encoding known proteins are included, as are all ORFs that could encode a protein greater than 100 amino acid residues. Smaller ORFs are typically not included unless they show similarity to a gene from another organism or unless the codon usage is typical of the organism being studied.

Table 15-1 part 2. Brock Biology of Microorganisms 11/e © 2006 Pearson Prentice Hall, Inc.

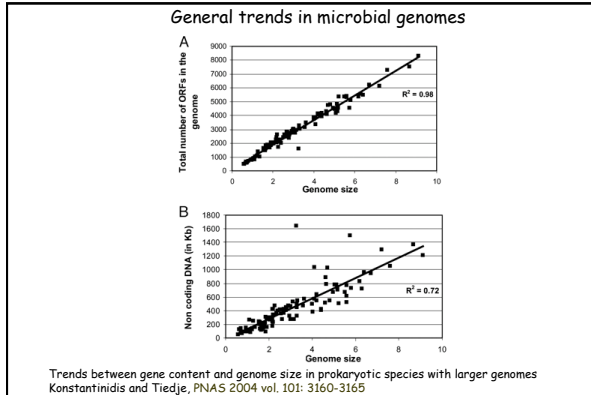


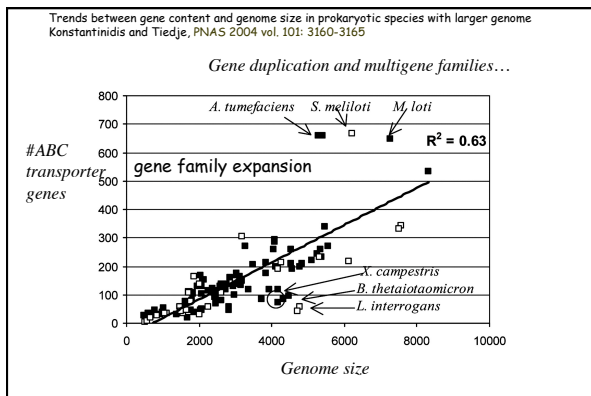
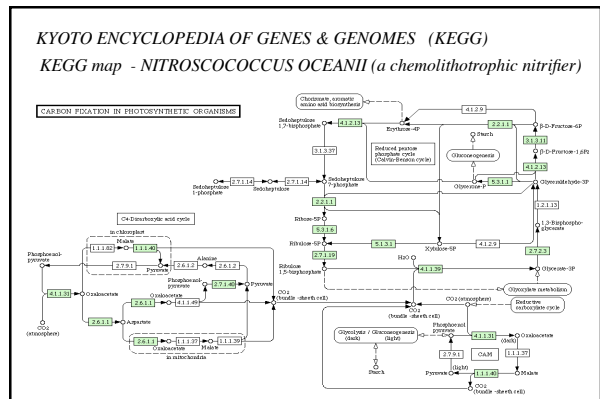
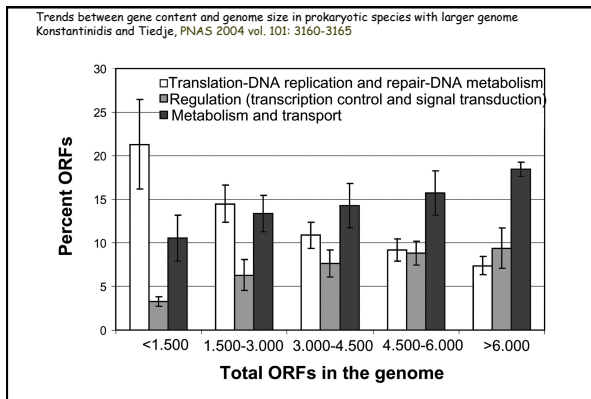
Table 15.2 Gene function in bacterial genomes

Percentage of genes on chromosome in that category

| Functional categories | <i>Escherichia coli</i> (4.64 Mbp) ^a | <i>Haemophilus influenzae</i> (1.83 Mbp) ^a | <i>Mycoplasma genitalium</i> (0.58 Mbp) ^a |
|-----------------------|---|---|--|
| Metabolism | 21.0 | 19.0 | 14.6 |
| Structural | 5.5 | 4.7 | 3.6 |
| Transport | 10.0 | 7.0 | 7.3 |
| Regulation | 8.5 | 6.6 | 6.0 |
| Translation | 4.5 | 8.0 | 21.6 |
| Transcription | 1.3 | 1.5 | 2.6 |
| Replication | 2.7 | 4.9 | 6.8 |
| Other, known | 8.5 | 5.2 | 5.8 |
| Unknown | 38.1 | 43.0 | 32.0 |

^a Chromosome size. Each organism listed contains only a single circular chromosome.

Table 15.2 Brock Biology of Microorganisms 11/e
© 2006 Pearson Prentice Hall, Inc.



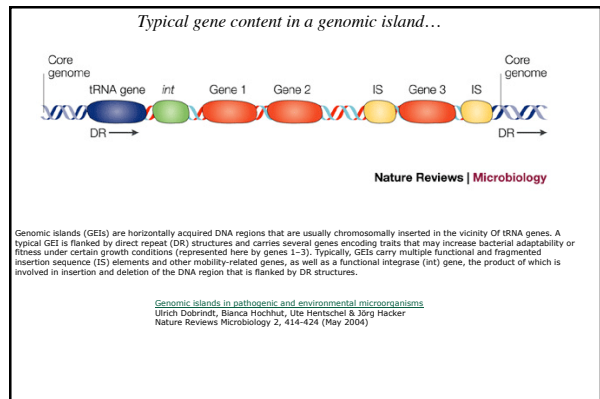
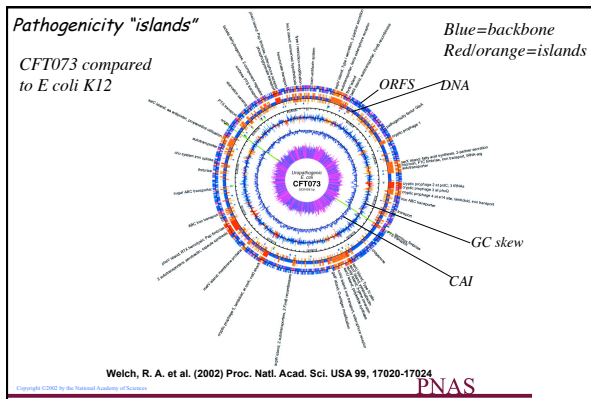
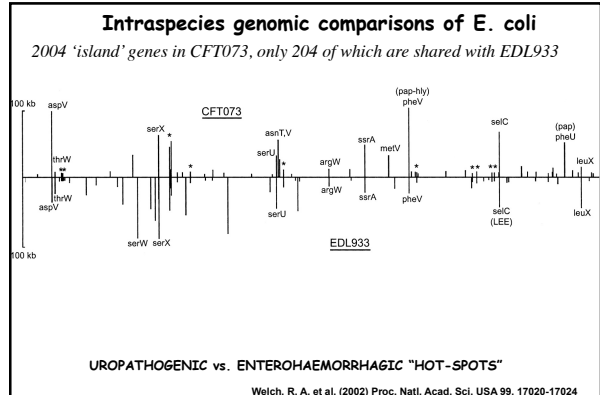
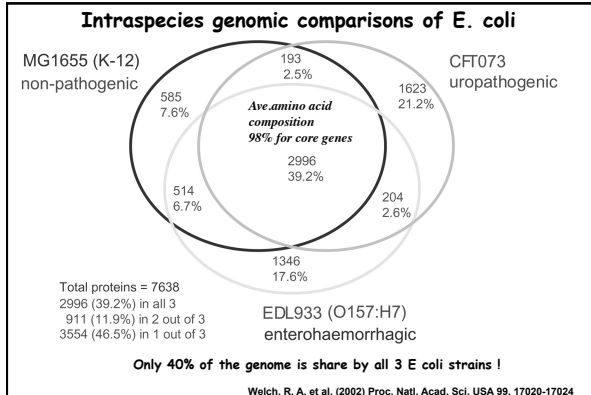
Evolution in genome sequence and content

Genome expansion

- Gene duplication and gene family expansion
- Genomic Islands
- Lateral gene transfer (but NOT in obligate symbionts)

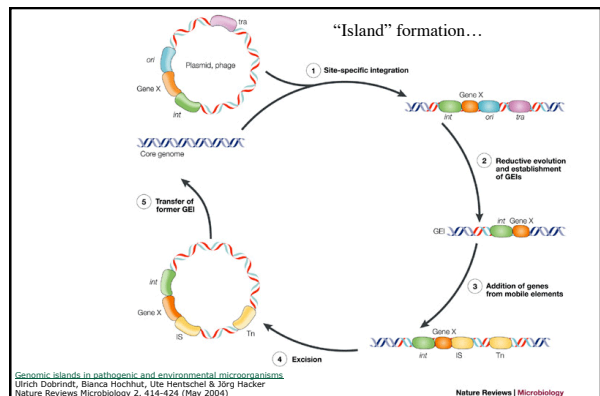
And contraction ...

- Gene deletion
- Genome reduction

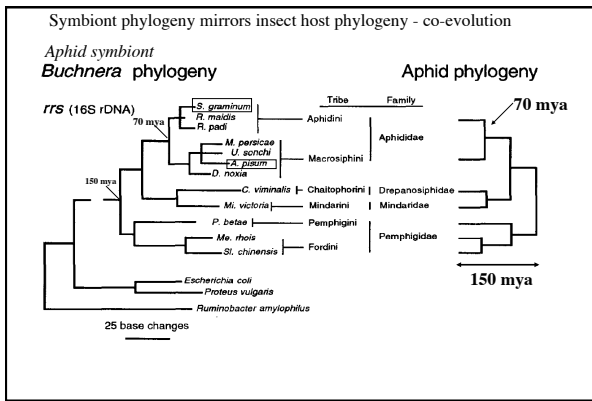
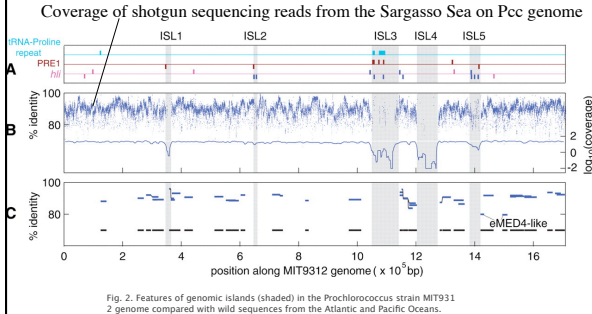


| Function | Organism | Genomic localization | Advantage conferred | References |
|---|---|----------------------|---|---------------------------|
| Invasibility | <i>Facial</i> <i>Escherichia coli</i> , <i>Klebsiella</i> spp., <i>Salmonella enterica</i> subgroups II and VI | Chromosome | Increased adaptability | 71-73 |
| Invasibility, antibiotic resistance, production of enterotoxins | Bacterial viruses | Chromosome | Increased adaptability, competitiveness | 82 |
| Expression of adhesins | <i>Facial</i> <i>Escherichia coli</i> | Chromosome | Adhesion to gastrointestinal epithelium, colonization | 9 |
| Expression of iron acquisition proteins | <i>Proteobacteria</i> <i>luminescens</i> | Chromosome | Killing of macrophages, competitiveness | 85,91 |
| Expression of type II secretion system | <i>Bacteroides</i> <i>glaberrimus</i> , <i>Mycobacterium</i> <i>goodii</i> , <i>Serratia</i> <i>profaecialis</i> | Chromosome, plasmid | Targeted delivery of effector proteins, interference with host-cell functions | 80,82,83, 86,90,93 |
| Expression of type IV secretion system | <i>Agrobacterium tumefaciens</i> , <i>Shigella</i> <i>flexneri</i> , <i>Brucella</i> , <i>Brucella abortus</i> , <i>Yersinia enterocolitica</i> , <i>Yersinia pseudotuberculosis</i> , <i>Yersinia enterocolitica</i> | Chromosome, plasmid | Targeted delivery of DNA and effector proteins into host cells, interference with host-cell functions | 34,50,82,83, 83,85, 89-91 |
| Nitrogen fixation, nodulation | <i>Rhizobium</i> | Chromosome, plasmid | Increased metabolic versatility, bacteria-host interactions | 84,85 |
| Nitrogen fixation | <i>Hydrothermal vent</i> <i>Archaea</i> | Chromosome | Increased metabolic versatility | 24 |
| Sucrose uptake | <i>Salmonella enteritidis</i> | Chromosome | Increased metabolic versatility | 29 |
| Magnetic phenotype | <i>Magnetospirillum magnetotacticum</i> | Chromosome | Magnetotaxis, increased adaptability | 95 |
| Resistance to methicillin | <i>Staphylococcus aureus</i> | Chromosome | Increased adaptability | 63,64,65 |
| Resistance to rifampin antibiotics | <i>Staphylococcus aureus</i> , <i>Salmonella enterica</i> DT104 | Chromosome | Increased adaptability | 49,56,96 |
| Resistance to mercury and vancomycin | <i>Providencia stuartii</i> | Chromosome | Increased adaptability | 58 |
| Degradation of phenolic compounds | <i>Pseudomonas putida</i> | Chromosome | Increased adaptability | 39,42 |
| Expression of photophosphorylation, toxins, fibrinase | <i>Proteobacteria</i> <i>luminescens</i> | Chromosome | Bacteria-mammalian interactions, exploitation of nutrients | 89-91 |
| Expression of toxins and bacteriocins, antibiotic biosynthesis | <i>Proteobacteria</i> <i>luminescens</i> | Chromosome | Killing of macrophages, competitiveness | 89-91 |
| Myxobacterial catenation | <i>Proteobacteria</i> <i>luminescens</i> | Chromosome | Increased fitness, virulence | 89-91 |
| Diplosporin biosynthesis and anti-hepatocellular carcinoma | <i>Proteobacteria</i> <i>luminescens</i> | Chromosome | Increased fitness, virulence | 89-91 |

Genomic islands in pathogenic and environmental microorganisms
 Ulrich Dobrindt, Bianca Hochhut, Ute Hentschel & Jörg Hacker
 Nature Reviews Microbiology 2, 414-424 (May 2004)

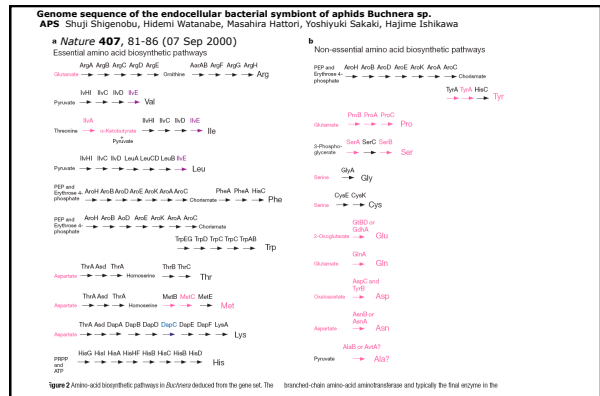
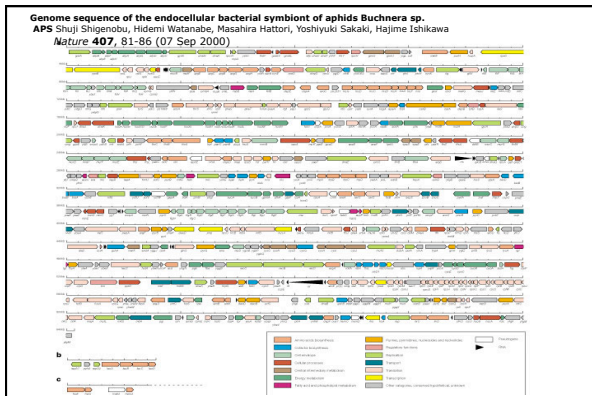


Genomic Islands and the Ecology and Evolution of Prochlorococcus
Maureen L. Coleman et al., Science 2006:311, pp. 1768 - 1770



Essential amino acids are not in the bugs diet !

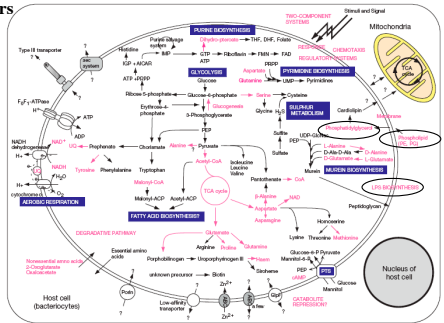
| Name | Recommended daily intake in human Adults mg per Kg body weight WHO | for 70Kg human (mg) |
|-----------------|--|---------------------|
| F Phenylalanine | 14 (sum with Tyrosine) | 980 |
| L Leucine | 14 | 980 |
| M Methionine | 13 (sum with Cysteine) | 910 |
| K Lysine | 12 | 840 |
| I Isoleucine | 10 | 700 |
| V Valine | 10 | 700 |
| T Threonine | 7 | 490 |
| W Tryptophan | 3 | 245 |
| H Histidine | unknown, 28 in infants (? sum with arginine) | (? 1960) |
| R Arginine | unknown, required for infants, maybe seniors | (?) |



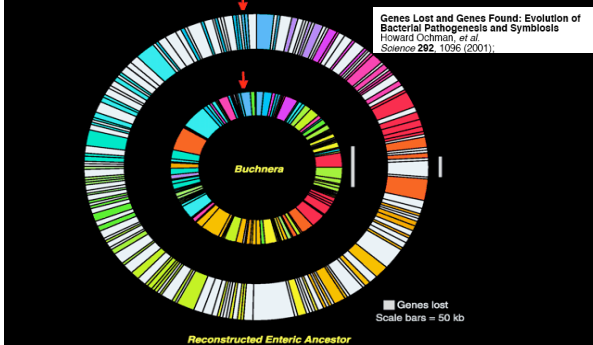
Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. *Nature* **407**, 81-86 (07 Sep 2000)
 APS* Shuji Shigenobu, Hidemi Watanabe, Masahira Hattori, Yoshiyuki Sakaki, Hajime Ishikawa

LITTLE CAPACITY FOR GENE REGULATION !
 Only 2 sigma factors

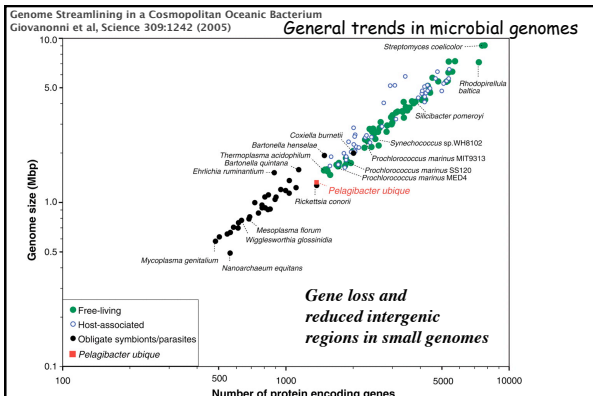
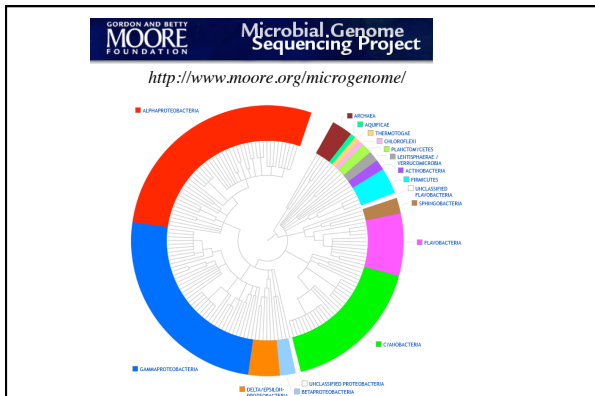
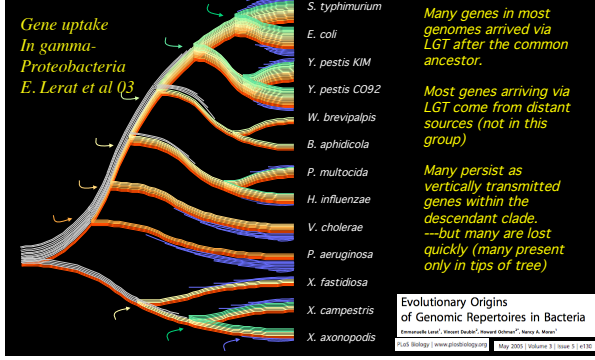
No leader seqs
 No attenuation
 No 2 component regulatory systems
 No cAMP/crp

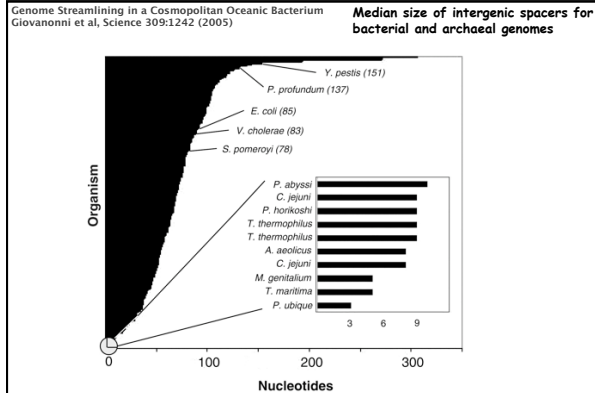


2000 genes lost from ancestor, to Sg/Ap divergence
 Must have been rapid evolution and gene loss !



Example of genome dynamics over time due to Lateral Gene Transfer

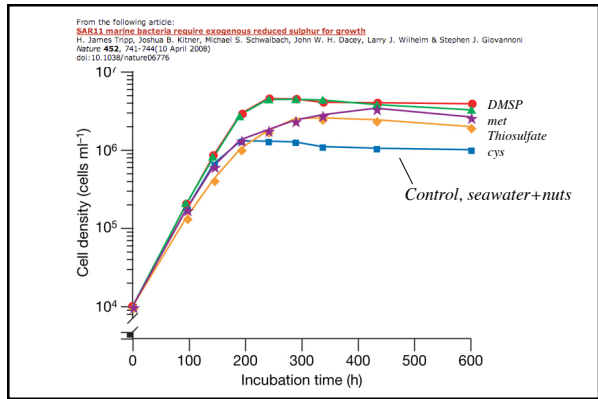
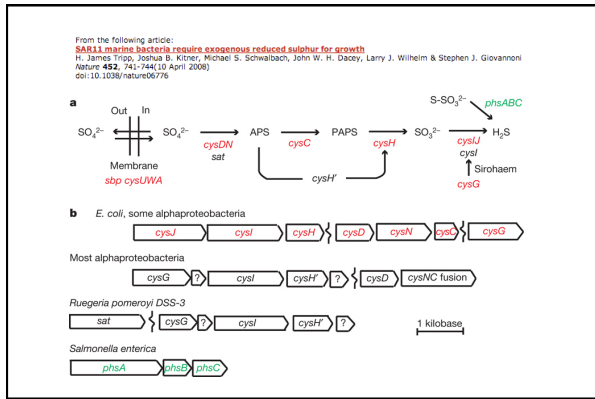




Genome Streamlining in a Cosmopolitan Oceanic Bacterium
Giovannoni et al, Science 309:1242 (2005)

Table 1. Metabolic pathways in Pelagibacter.

| Pathway | Prediction |
|------------------------------------|------------|
| Glycolysis | Uncertain |
| TCA cycle | Present |
| Cysteine shunt | Present |
| Respiration | Present |
| Penicillin phosphate cycle | Present |
| Fatty acid biosynthesis | Present |
| Cell wall biosynthesis | Present |
| Biosynthesis of all 20 amino acids | Present |
| Heme biosynthesis | Present |
| Ubiquinone | Present |
| Nicotinate and nicotinamide | Present |
| Folate | Present |
| Riboflavin | Present |
| Paracetamol | Absent |
| flg | Absent |
| Thiamine | Absent |
| Biotin | Absent |
| B ₁₂ | Absent |
| Retinal | Present |



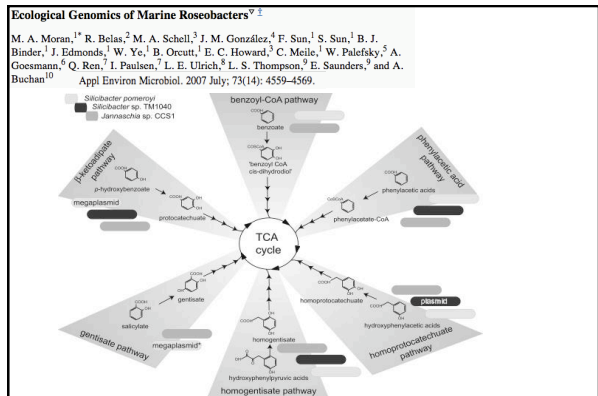
Ecological Genomics of Marine Roseobacters^{7, 2}

M. A. Moran,¹ R. Belas,² M. A. Schell,³ J. M. González,⁴ F. Sun,¹ S. Sun,¹ B. J. Binder,¹ J. Edmonds,¹ W. Ye,¹ B. Orcutt,¹ E. C. Howard,² C. Meile,¹ W. Palefsky,² A. Goesmann,⁶ Q. Ren,¹ I. Paulsen,⁷ L. E. Ulrich,⁸ L. S. Thompson,⁹ E. Saunders,⁹ and A. Buchan¹⁰
Appl Environ Microbiol. 2007 July; 73(14): 4559-4569.

Features of three Roseobacter genomes

| Feature | Result for: | | |
|-------------------------------------|--------------------------|---------------------------------------|-----------------------------------|
| | <i>S. pomeroyi</i> DSS-3 | <i>Silicibacter</i> sp. strain TM1040 | <i>Jannaschia</i> sp. strain CCS1 |
| Total no. of coding sequences | 4,283 | 3,863 | 4,283 |
| G+C content (%) | 64.0 | 60.0 | 62.2 |
| No. of rRNA operons | 3 | 5 | 1 |
| No. of replicons | 2 | 3 | 2 |
| Molecule length by replicon (bp) | | | |
| Main chromosome | 4,109,442 | 3,201,640 | 4,317,977 |
| Megaplasmid | 491,611 | 823,032 | NA |
| Plasmid | NA | 131,885 | 86,072 |
| No. of coding sequences by replicon | | | |
| Main chromosome | 3,838 | 3,013 | 4,212 |
| Megaplasmid | 445 | 747 | NA |
| Plasmid | NA | 103 | 71 |
| G+C content by replicon (%) | | | |
| Main chromosome | 64.2 | 60.4 | 62.3 |
| Megaplasmid | 62.8 | 59.3 | NA |
| Plasmid | NA | 55.3 | 57.8 |
| No. of rRNA operons by replicon | | | |
| Main chromosome | 3 | 1 | 1 |
| Megaplasmid | 0 | 3 | NA |
| Plasmid | NA | 1 | 0 |

^{NA}, not applicable.



Ecological Genomics of Marine Roseobacters^{9,12}

M. A. Moran,^{1*} R. Belas,² M. A. Schell,³ J. M. González,⁴ F. Sun,¹ S. Sun,¹ B. J. Binder,¹ J. Edmonds,¹ W. Ye,¹ B. Orcutt,¹ E. C. Howard,³ C. Meile,¹ W. Palefsky,² A. Goessmann,⁶ Q. Ren,⁷ I. Paulsen,⁸ L. E. Ulrich,⁸ L. S. Thompson,² E. Saunders,⁹ and A. Buchan¹⁰ Appl Environ Microbiol. 2007 July; 73(14): 4559-4569.

Numbers and percentages of shared and unique genes in three *Roseobacter* genomes

| Method | No. (%) of genes that are: | | | |
|---------------|----------------------------|------------------------------------|---|---|
| | Shared 3 ways ^a | Unique to <i>S. pomeroyi</i> DSS-3 | Unique to <i>Silicibacter</i> sp. strain TM1040 | Unique to <i>Jannaschia</i> sp. strain CCS1 |
| RBH analysis | 1,839 | 1,336 (31) | 1,027 (27) | 1,685 (39) |
| ISGS analysis | 121 | 1,260 (29) | 983 (25) | 1,600 (37) |

^aGenes shared three ways are those with orthologs in all three genomes. RBH analysis used the criterion of RBH between pairs of genomes with a corresponding BLAST E-value of <10⁻⁵ and amino acid identity of >90%. ISGS analysis used one-way BLASTy and iBLASTy analysis to identify genes shared by the three roseobacters but missing from four nonmarine alphaproteobacterial relatives (*R. sphaeroides*, *P. denitrificans*, *S. meliloti*, and *M. luti*).

Ecological Genomics of Marine Roseobacters^{9,12}

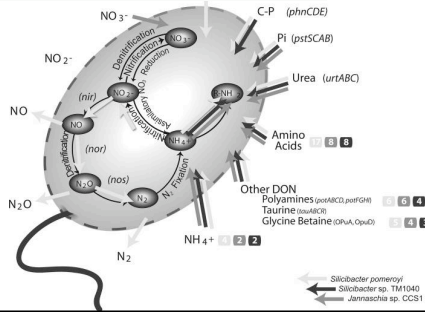
M. A. Moran,^{1*} R. Belas,² M. A. Schell,³ J. M. González,⁴ F. Sun,¹ S. Sun,¹ B. J. Binder,¹ J. Edmonds,¹ W. Ye,¹ B. Orcutt,¹ E. C. Howard,³ C. Meile,¹ W. Palefsky,² A. Goessmann,⁶ Q. Ren,⁷ I. Paulsen,⁸ L. E. Ulrich,⁸ L. S. Thompson,² E. Saunders,⁹ and A. Buchan¹⁰ Appl Environ Microbiol. 2007 July; 73(14): 4559-4569.

| Genome or parameter | Result for: | | | | | | | | | |
|--|-------------|-----------|----------|------|----------|----------|----------------|----------------|----------|---------|
| | AANP genes | caa genes | psrGII | hucC | suB | nit5K | nosI | ghe genes | vg genes | dmfA |
| Genomes | | | | | | | | | | |
| Complete | | | | | | | | | | |
| <i>Silicibacter pomeroyi</i> DSS-3 | + | + | + | + | + | + | + | + | + | + |
| <i>Silicibacter</i> sp. strain TM1040 | + | + | + | + | + | + | + | + | + | + |
| <i>Jannaschia</i> sp. strain CCS1 | + | + | + | + | + | + | + | + | + | + |
| Draft | | | | | | | | | | |
| <i>Silicibacter</i> sp. strain EE-36 | + | + | + | + | + | + | + | + | + | + |
| <i>Silicibacter</i> sp. strain NAS-14.1 | + | + | + | + | + | + | + | + | + | + |
| <i>Alteromonas macleodii</i> DSM | + | + | + | + | + | + | + | + | + | + |
| <i>Roseobacter</i> sp. strain 217 | + | + | + | + | + | + | + | + | + | + |
| <i>Oceanicola</i> <i>Antoniou</i> HTCC2397 | + | + | + | + | + | + | + | + | + | + |
| <i>Oceanicola</i> <i>grandis</i> HTCC2016 | + | + | + | + | + | + | + | + | + | + |
| <i>Roseobacter</i> <i>bacterium</i> HTCC2054 | + | + | + | + | + | + | + | + | + | + |
| <i>Leptanillus</i> <i>veroffii</i> SKA53 | + | + | + | + | + | + | + | + | + | + |
| <i>Roseobacter</i> sp. strain MED193 | + | + | + | + | + | + | + | + | + | + |
| Percentages | | | | | | | | | | |
| % of <i>Roseobacter</i> genomes | 25 | 92 | 83 | 25 | 58 | 25 | 50 | 92 | 50 | 50 |
| Experimental evidence (reference) | ND | Yes (63) | Yes (13) | ND | Yes (21) | Yes (17) | Yes (Fig. S4F) | Yes (Fig. S5F) | ND | Yes (2) |
| Ortholog in <i>P. vulgaris</i> HTCC2042 | No | No | No | No | No | No | No | No | No | Yes (2) |

^aThe nine draft *Roseobacter* genomes include those for aerobic anoxygenic phototrophy (AANP) genes, carbon monoxide oxidase (*cox* genes), aromatic compound degradation (*psrGII* and *hucC*), sulfur metabolism (*suB*), denitrification (*nit5K*), nitrite assimilation (*nosI*), phosphorus use (*ghe* genes), type IV secretion (*vg* genes), and DMSP demethylase (*dmfA*). + indicates the presence of a homolog with an E-value of <10⁻¹⁰ and amino acid sequence identity of >60%. Phenotypes that have been demonstrated experimentally for *Roseobacter* strains are indicated by "yes," "ND," not determined. The presence of an ortholog in the genome of *SAR11* member *P. vitreus* is indicated.

Ecological Genomics of Marine Roseobacters^{9,12}

M. A. Moran,^{1*} R. Belas,² M. A. Schell,³ J. M. González,⁴ F. Sun,¹ S. Sun,¹ B. J. Binder,¹ J. Edmonds,¹ W. Ye,¹ B. Orcutt,¹ E. C. Howard,³ C. Meile,¹ W. Palefsky,² A. Goessmann,⁶ Q. Ren,⁷ I. Paulsen,⁸ L. E. Ulrich,⁸ L. S. Thompson,² E. Saunders,⁹ and A. Buchan¹⁰ Appl Environ Microbiol. 2007 July; 73(14): 4559-4569.



Comparative genomics of two ecotypes of the marine planktonic copiotroph *Alteromonas macleodii* suggests alternative lifestyles associated with different kinds of particulate organic matter

Diego Ivars-Martinez, Ana-Belen Martin-Cabrero, Giuseppe D'Auria, Alex Mira, Steve Fenner, Justin Johnson, Robert Friedman and Francisco Rodriguez-Velazco

Table 1 General features of genomes

| | <i>A. macleodii</i> AITDE | <i>A. macleodii</i> ATCC27126 ^a | <i>P. atlantica</i> Tbc |
|-------------------------|---------------------------|--|-------------------------|
| Size (bp) | 4,412,285 | 4,607,844 | 5,187,010 |
| GC content% | 44.50 | 44.62 | 44.6 |
| Coverage | 1 | 716 | 1 |
| Total ORFs | 4102 | 4444 | 4261 |
| Function assigned | 3057 | 3319 | 3408 |
| Hypothetical proteins | 1045 | 1086 | 873 |
| Genes shared with AITDE | 2996 | 1988 | 1783 |
| Exclusive genes | 1242 | 1988 | 2368 |
| rRNA operons | 5 | 5 | 5 |
| tRNAs | 60 | 48 | 68 |
| ANP | 81.24 | 81.24 | 88.11 |
| Coding density | (86.63%) | (87.29%) | (86.14%) |
| Average spacers size | 543 | 181 | 167 |
| Average ORF size | 691 | 1343 | 655 |

Abbreviations: ANI, average nucleotide identity; *A. macleodii*, *Alteromonas macleodii*; AITDE, *A. macleodii* 'deep ecotype'; *P. atlantica*, *Pseudoalteromonas atlantica*.
^aOnly automatic annotation.
^bANI (Korosteleva and Tiedje, 2005).

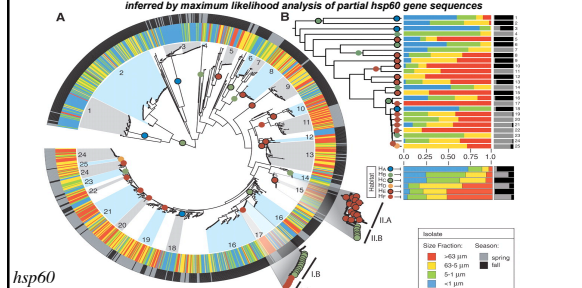
Comparative genomics of two ecotypes of the marine planktonic copiotroph *Alteromonas macleodii* suggests alternative lifestyles associated with different kinds of particulate organic matter

Diego Ivars-Martinez, Ana-Belen Martin-Cabrero, Giuseppe D'Auria, Alex Mira, Steve Fenner, Justin Johnson, Robert Friedman and Francisco Rodriguez-Velazco

| Genes categories | ATCC 27126 | AITDE |
|-----------------------------|------------|---------|
| Phage integrase | 12 (9) | 32 (23) |
| Transposase and IS elements | 3 (1) | 65 (63) |
| Chaperones | 9 (1) | 13 (5) |
| Sigma factors | 10 (0) | 8 (0) |
| Histidine kinases | 53 (18) | 40 (5) |
| Dioxygenases | 8 (0) | 11 (3) |
| Heavy metals resistance | 24 (1) | 36 (18) |
| TonB receptors | 82 (31) | 52 (7) |
| Toxin-antitoxin | 7 (1) | 12 (5) |
| Acr systems | 50 (11) | 45 (11) |
| ABC transporters | 79 (19) | 82 (22) |

Abbreviations: *A. macleodii*, *Alteromonas macleodii*; AITDE, *A. macleodii* 'deep ecotype'.

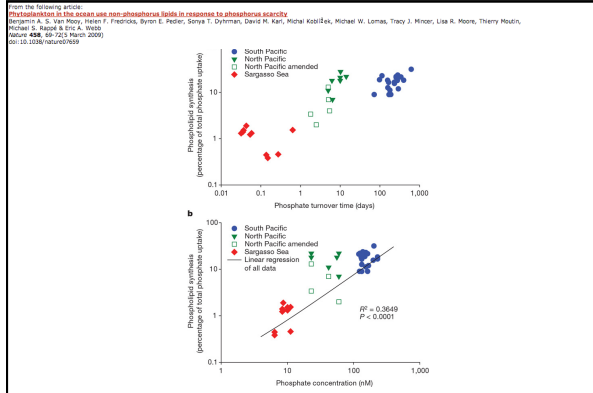
Fig. 1. Season and size fraction distributions and habitat predictions mapped onto *Vibrionaceae* isolate phylogeny inferred by maximum likelihood analysis of partial *hsp60* gene sequences



D. E. Hunt et al., Science 320, 1081-1085 (2008) Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton

Published by AAAS





From the following article:
 Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity
 Benjamin A. S. Van Mooy, Heben F. Friedrichs, Byron E. Pedler, Sonya T. Dyhrman, David M. Karl, Michal Kozlowski, Michael W. Lomas, Tracy J. Minner, Lisa R. Moore, Thierry Moutin, Michael D. Rappel & Eric A. West
 Nature 458, 667–723 March 2009
 doi:10.1038/nature07939

Table 1. Ratios of substitute lipids to phospholipids in phytoplankton cultures and environmental samples.

* Figures & Tables Index

| | SQDG/PG ratio, P-replete | SQDG/PG ratio, P-limited | BL/PC ratio, P-replete | BL/PC ratio, P-limited |
|---------------------------------|--------------------------|--------------------------|------------------------|------------------------|
| Cyanobacteria | | | | |
| <i>Synechococcus</i> WH8102 | 9.9 ± 2.0 | 120.5 ± 7.1 | 0 [†] | 0 [†] |
| <i>Synechococcus</i> WH7803 | 10.3 ± 0.3 | 61.6 ± 15.4 | 0 [†] | 0 [†] |
| <i>Synechococcus</i> WH5701 | 6.2 ± 0.5 | 132.0 ± 31.0 | 0 [†] | 0 [†] |
| <i>Prochlorococcus</i> MED4 | 20.0 ± 1.3 | 34.1 ± 1.6 | 0 [†] | 0 [†] |
| <i>Crocosphaera</i> watsonii | 4.0 | 5.8 | 0 [†] | 0 [†] |
| <i>Trichodesmium</i> erythraeum | 7.8 ± 1.0 | 18.5 ± 4.9 | 0 [†] | 0 [†] |
| Eukaryotic phytoplankton | | | | |
| <i>Thalassiosira</i> pseudonana | 3.0 ± 0.9 | 394.8 ± 48.2 | <0.01 [‡] | >500 [§] |
| <i>Chaetoceros</i> affinis | 10.5 ± 3.6 | 26.3 ± 9.0 | 0.9 ± 0.2 | 27.8 ± 8.3 |
| <i>Emiliania</i> huxleyi | <0.01 [‡] | <0.01 [‡] | 0.7 | 1.3 |
| Communities | | | | |
| South Pacific | 3.6 ± 0.8 | NA | 3.6 ± 1.7 | NA |
| Sargasso Sea | NA | 4.5 ± 1.1 | NA | 13.1 ± 4.0 |