Metagenomics

Sequences from the Environment

Last Updated: 2006

National Center for Biotechnology Information (US)
Bethesda (MD)

National Center for Biotechnology Information (US), Bethesda (MD)

NLM Citation: Metagenomics: Sequences from the Environment [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2006.

This publication contains a collection of chapters developed by NCBI from metagenome projects submitted to the Genomes Projects database. These chapters provide links to the sequence data and genome project submission as well as to BLAST, taxonomic lineages and publications.

Metagenomics is the functional and sequence-based analysis of the collective microbial genomes that are contained in an environmental sample. The word metagenomics describes "the notion of analysis of a collection of similar but not identical items, as in a meta-analysis, which is an analysis of analyses" (Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev. 2004 Dec;68(4):669-85.). Metadata from each metagenome project has been divided into the following categories for presentation within the Book Chapter:

WGS study Whole Genome Shotgun sequences that have been submitted for each metagenome project. WGS projects consist of contigs (and in some cases singletons) isolated from the described environmental source. See http://www.ncbi.nlm.nih.gov/Genbank/wgs.html for more information about WGS submissions.

Contig A non-redundant sequence formed by joining, based on sequence overlap, one or more smaller sequences. There should be no gaps.

Scaffold A non-redundant sequence formed by joining one or more contig sequences. A sequence overlap is not required to form a scaffold. Typically, a scaffold contains one or more gaps. The identification of genomes using WGS scaffolds is included within the Genome Assembly portion of the project summary.

Trace data Single-pass reads (including DNA sequence chromatograms, base calls, and quality estimates) submitted to the Trace Archive. See http://www.ncbi.nlm.nih.gov/Traces/trace.cgi for more information about the Trace Archive.

16S ribosomal RNA 16S ribosomal RNA sequences that have been submitted to GenBank. These may have been used in determining taxonomic phylotypes within the metagenome.

Table of Contents

Acid Mine Drainage Biofilm

This biofilm sequencing project was designed to explore the distribution and diversity of metabolic pathways in acidophilic biofilms (e.g., nitrogen fixation, sulfur oxidation, iron oxidation), in order to understand the mechanisms by which the microbes tolerate environmental extremes, and to evaluate how this might impact the geochemistry of the environment.

Data

WGS study	AADL00000000
Contigs	AADL01000001-AADL01002534
Scaffolds	CH003520-CH004435
Trace data	180,713 reads available from the Trace Archive
16S ribosomal RNA	No 16S rRNA data available

A total of 76.2 million bp of DNA sequence was generated from 103,462 high quality trimmed reads.
Coverage average depth in raw shotgun data: 10X
85% of shotgun reads were combined into contigs greater then 2 kb with a combined length of 10.83 Mbp.

Isolation Source

Acidophilic biofilms are self-sustaining communities that grow in the deep subsurface and receive no significant inputs of fixed carbon or nitrogen from external sources. While some acid mine drainage is caused by the oxidization of rocks rich in sulfide minerals, this is a very slow process and most acid mine drainage is due directly to microbial activity.

This data was isolated underground from a pink biofilm microbial community growing on the surface of flowing acid mine drainage (AMD) in the five-way region (CG) of the Richmond mine at Iron Mountain, California in March 2002. Iron Mountain is located at 40 deg 40' 38.42" N and 122 deg 31" 19.90 " W (elevation 3,100 ft.). The 5-way biofilm was growing in pH 0.83, 42 degrees C, 317 mM Fe, 14 mM Zn, 4 mM Cu, and 2 mM As solution and was collected from a surface area of approximately 0.05 m^2.

Genome Assembly

The 2455 contigs were subdivided into five sets of WGS scaffolds:

Assembly	Sequences	Notes
Leptospirillum sp. Group II '5-way CG'	CH003520-CH003544	Iron-oxidizing bacteria
Leptospirillum sp. Group III	CH003545-CH003921	Iron-oxidizing bacteria
Ferroplasma acidarmanus Type I	CH003922-CH004070	Iron-oxidizing archaea
Ferroplasma sp. Type II	CH004071-CH004104	Iron-oxidizing archaea
Thermoplasmatales archaeon Gpl (G-plasma)	CH004105-CH004435	Archaea; facultatively anaerobic, thermoacidophilic, autotrophic or heterotrophic organisms

References

1. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment *Nature* 428, 25-26 PubMed PMID: 14961025.

2. Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC 2nd, Shah M, Hettich RL, Banfield JF (2005) Community proteomics of a natural microbial biofilm *Science* 308, 1915-1920 PubMed PMID: 15879173.
3. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM (2005) Comparative metagenomics of microbial communities *Science* 308, 554-557 PubMed PMID: 15845853.
4. Acid Mine Drainage Home Page at the University of California, Berkeley.

Waseca County Farm Soil

The metabolic capabilities of terrestrial and marine microbial communities were compared using largely unassembled sequence data obtained by shotgun sequencing DNA isolated from whale fall and soil. The gene content isolated from soil was compared to environmental sequence data isolated from three whale fall carcasses, acid mine drainage, and the Sargasso Sea. Analysis of these metagenomes identified genes that display environment-dependent characteristics, distribution and expression.

Data

WGS study	AAFX00000000
Contigs	AAFX01000001-AAFX01139340
Scaffolds	No scaffolds available
Trace data	138,347 reads available from the Trace Archive
16S ribosomal RNA	AY921654-AY922179

100 Mbp of sequence was generated from the soil sample. However, less then 1% of nearly 150,000 reads exhibited overlap with reads from independent clones. It is estimated that 2-5 billion base pairs would be necessary to obtain 8X-coverage for draft genome assemblies.

Isolation Source

Agricultural surface soil (0-10 cm) was collected in September 2001 from a farm in Waseca County, Minnesota. The surrounding area had been used for livestock, including sheep, cattle, and pigs, and was in the drainage path of a silage storage bunker that had been used for sweet corn and pea silage water operations from 1990-1997. Biochemical analyses on 20g of soil from this site revealed it to be clay loam, with fair to low organic matter content, high levels of essential elements, and low levels of nonessential elements. Microscopic analysis found the organisms in the sample to be primarily prokaryotic.

Genome Assembly

None

References

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM. Comparative metagenomics of microbial communities. Science 2005, 308: 554-557. PubMed PMID: 15845853.

Soil metagenomic data is also available from the U.S. Department of Energy Joint Genome Institute Integrated Microbial Genomes System. Supporting data and an interactive analysis of the soil, whale fall, and Sargasso Sea communities can be found here.

Whale Fall Community

The metabolic capabilities of terrestrial and marine microbial communities were compared using largely unassembled sequence data obtained by shotgun sequencing DNA isolated from whale fall and soil. The gene content isolated from three whale fall carcasses was compared to environmental sequence data isolated from soil, acid mine drainage, and the Sargasso Sea. Analysis of these metagenomes identified genes that display environment-dependent characteristics, distribution and expression.

Data

WGS study	Whale fall #1	AAFY00000000
	Whale fall #2	AAFZ00000000
	Whale fall #3	AAGA00000000
Contigs	Whale fall #1	AAFY01000001-AAFY01028151
	Whale fall #2	AAFZ01000001-AAFZ01029934
	Whale fall #3	AAGA01000001-AAGA01026232
Scaffolds	No scaffolds available	
Trace data	117,326 reads available from the Trace Archive	
16S ribosomal RNA	Whale fall #1	AY922180-AY922207
	Whale fall #2	AY922208-AY922232
	Whale fall #3	AY922233-AY922252

25 Mbp of sequence was generated for each whale fall library. It is estimated that between 100-700 Mbp would be needed to generate a draft assembly for the most prevalent genome.

Isolation Source

Whale carcasses on the sea floor are a rich source of organic matter for organisms that inhabit the ocean depths. This ecological niche, referred to as a "whale fall" may select for unique species within this nutrient-poor environment. In particular, sulfate-reducing bacteria oxidize organic matter, such as the lipid-rich skeleton, producing sulfides that chemosynthetic microorganisms use for energy.

Whale fall #1	Isolated from a rib bone from a gray whale carcass experimentally sunk in 1998 in the Pacific Ocean, Santa Cruz Basin (N33.30 W119.22) at a depth of 1674 meters
Whale fall #2	Isolated from an orange microbial mat on a gray whale carcass experimentally sunk in 1998 in the Pacific Ocean, Santa Cruz Basin (N33.30 W119.22) at a depth of 1674 meters
Whale fall #3	Isolated from a whale bone of uncertain age and species collected by otter trawl on a muddy seafloor at a depth of 560 meters in the Southern Ocean of the West Antarctic Peninsula Shelf (S65.10 W64.47)

Genome Assembly

None

References

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM. Comparative metagenomics of microbial communities. Science 2005, 308: 554-557. PubMed PMID: 15845853.

Whalefall metagenomic data is also available from the U.S. Department of Energy Joint Genome Institute Integrated Microbial Genomes System. Supporting data and an interactive analysis of the soil, whale fall, and Sargasso Sea communities can be found here.

Methane-Oxidizing Archaea

Anaerobic oxidation of methane (AOM) is estimated to consume a vast quantity of methane annually, however, the biology of this process is not well understood. It is suggested that methane is converted by methanotrophic Archaea to carbon dioxide and reduced by-products. Methane-oxidizing Archaea were isolated from deep-sea methane seeps for genomic analysis in order to test the "reverse-methanogenesis" hypothesis. Genes associated with methanogenesis were identified within an Archaeal group lending to a greater biological understanding of the process of methane oxidation in anoxic marine habitats.

Data

WGS study	No WGS sequences available	
Contigs	Fosmid sequences	AY714814-AY714873
Scaffolds	No scaffolds available	
Trace data	No available trace data	
16S ribosomal RNA	No 16S rRNA data available	

4.6 Mbp of DNA sequence was generated from the fosmid library from 7104 reads (averaging 700 bp per read). 191 fosmids encompassing 7.4 Mbp were selected for subcloning and. Archaeal clones were chosen in order to maximize large-insert coverage dept of anaerobic methane-oxidizing Archaea genomes.

Isolation Source

Samples were obtained from a 6 to 9-cm deep sea sediment pushcore interval (PC45) isolated from the Eel River Basin off the Mendocino California coastline (dive T201, LAT: 40,785 LON: -124,596 at a depth of 520 m).

Genome Assembly

A group of anaerobic methane-oxidizing Archaea ANME-1 dominated the purified cell population. Analysis of gene content showed that these organisms harbor methanogenesis-associated genes, which allowed the fosmids to be divided into two bins, ANME-1 and ANME-2. Assembly of binned fosmids generated 13 unique scaffolds for ANME-1 and one scaffold for ANME-2, with no cross assembly between bins. Genes that were involved in reverse-methanogenesis were found associated mostly with one group of bacteria, ANME-1 (anaerobic methane-oxidizing Archaea), although some were found in ANME-2.

References

Hallam SJ, Putnam N, Preston CM, Detter JC, Rokhsar D, Richardson PM, DeLong EF. Reverse methanogenesis: testing the hypothesis with environmental genomics. Science. 2004 Sep 3;305(5689):1457-1462. PubMed PMID: 15353801.

Additional data are available from the AOM Microbial Community homepage at the Joint Genome Institute.

Human Fecal Virus

The enteric RNA viral community present in healthy humans has not been described, even though many RNA viruses are known to cause gastroenteritis. Using comparative metagenomic analysis, the RNA viruses found in three fecal samples from two healthy human individuals were analyzed. The vast majority were similar to a plant pathogenic RNA viruses, pepper mild mottle virus (PMMV), suggesting that humans may serve as vehicles for the dissemination of plant viruses.

Data

WGS study	Sample 1	AAMG00000000
	Sample 2	AAMH00000000
	Sample 3	AAMI00000000
Contigs	Sample 1	AAMG01000001-AAMG01002373
	Sample 2	AAMH01000001-AAMH01004593
	Sample 3	AAMI01000001-AAMI01003329
Scaffolds	No scaffolds available	
Trace data	39,105 reads available from the Trace Archive	
16S ribosomal RNA	No 16S rRNA data available	

A total of 36,769 sequences were generated from all three samples. 25,779 (76.6%) were most similar to viruses, with 25,040 (97/1%) being homologous to plant viruses.

Isolation Source

Three fecal samples from two healthy adults living in San Diego were used for virus isolation.

Sample 1	Individual 1 fecal sample
Sample 2	Individual 1 fecal sample collected 6 months after collecting Sample 1
Sample 3	Individual 2 fecal sample

Feces-borne viral particles were concentrated and treated with DNase and RNase to eliminate potential contamination with nucleic acids.

Genome Assembly

None

References

Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, Soh SW, Hibberd ML, Liu ET, Rohwer F, Ruan Y. RNA viral community in human feces: prevalence of plant pathogenic viruses. PLoS Biol. 2006 Jan;4(1):e3. PubMed PMID: 16336043.

More information about the Human Fecal Virus project is available at the Genome Institute of Singapore.

Global Ocean Sampling Expedition Microbial Metagenome

Microbial DNA sequences were isolated from seawater samples collected during a global circumnavigation aboard the Sorcerer II in order to analyze the genomic and functional diversity within this particular marine environment. Starting in Halifax, Canada, samples were collected at sites along the U.S. east coast, Gulf of Mexico, Galapagos Islands, central and south Pacific Oceans, Australia, Indian Ocean, South Africa, and across the Atlantic back to the U.S. This large dataset allows investigators to study genetic and biochemical microbial diversity within the marine environment.

Data

WGS study	AACY000000000
Contigs	AACY020000001-AACY024124495
Scaffolds	EM000001-EM999999
	EN000001-EN999999
	EP000001-EP999999
	EQ000001-EQ087209
Trace data	7,521,215 reads available from the Trace Archive
16S ribosomal RNA	EU798997-EU805409

A total of 41 different samples were taken from a variety of aquatic habitats collected over 8,000 km. 7.7 million sequencing reads were obtained from size-fractionated samples, yielding 6.4 million contiguous sequences, totaling 5.9 Gbp of nonredundant sequence. These were further processed into about 3 million assemblies (scaffolds).

Isolation Source

Samples were collected as part of the Sorcerer II expedition between August 8, 2003, and May 22, 2004. Most specimens were collected from surface water marine environments at approximately 320 km intervals. 44 samples were obtained from 41 sites, covering a wide range of distinct surface marine environments as well as a few nonmarine aquatic samples for contrast.

Genome Assembly

The 4,124,495 contigs were further assembled into 3,087,206 WGS scaffolds using a overlap cutoff of 98%. 85% of the assembled sequences and 57% of unassembled data is unique at the 98% identity cutoff. Based on clustering and HMM profiling, more than 6.1 million proteins were annotated on this dataset (includes bacterial as well as viral sequences). These are defined as "marine metagenome" within the source. 60 highly abundant ribotypes were identified and found to be associated with open ocean and aquatic samples.

Scaffolds	EM000001-EM999999
	EN000001-EN999999
	EP000001-EP999999
	EQ000001-EQ087209

References

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W, Fouts DE, Levy S, Knap AH, Lomas MW, Nealson K, White O, Peterson J, Hoffman J, Parsons R, Baden-

Tillson H, Pfannkoch C, Rogers YH, Smith HO. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 2004 Apr 2;304(5667):58-60. PubMed PMID: 15001713.

Tringe SC, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM. Comparative metagenomics of microbial communities. Science. 2005 Apr 22;308(5721):5547. PubMed PMID: 15845853.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers YH, Falcon LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Nealson K, Friedman R, Frazier M, Venter JC. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol. 2007 Mar 13;5(3):e77. PubMed PMID: 17355176.

Yooseph S, Sutton G, Rusch DB, Halpern Al, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W, Jaroszewski L, Cieplak P, Miller CS, Li H, Mashiyama ST, Joachimiak MP, van Belle C, Chandonia JM, Soergel DA, Zhai Y, Natarajan K, Lee S, Raphael BJ, Bafna V, Friedman R, Brenner SE, Godzik A, Eisenberg D, Dixon JE, Taylor SS, Strausberg RL, Frazier M, Venter JC. The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families. PLoS Biol. 2007 Mar 13;5(3):e16. PubMed PMID: 17355171.

Kannan N, Taylor SS, Zhai Y, Venter JC, Manning G. Structural and Functional Diversity of the Microbial Kinome. PLoS Biol. 2007 Mar 13;5(3):e17. PubMed PMID: 17355172.

Human Distal Gut Microbiome

A metagenomics approach was used to reveal microbial genomic and genetic diversity and to identify some of the distinctive functional attributes encoded in the human distal gut microbiome. Insight into the diversity within these samples was obtained by comparison of a subset of the shotgun reads to a completed bacterial genome. The analysis of the gene content and functional attributes of the gut microbiome in healthy humans defines them as superorganisms whose metabolism represents an amalgamation of microbial and human attributes.

Data

WGS study	Human subject 7	AAQK00000000
	Human subject 8	AAQL00000000
Contigs	Human subject 7	AAQK01000001-AAQK01010488
	Human subject 8	AAQL01000001-AAQL01012020
Scaffolds	No scaffolds available	
Trace data	156,528 reads available from the Trace Archive	
16S ribosomal RNA	Human subjects 7 and 8	DQ325545-DQ327606

A total of 74,462 high-quality sequence reads were obtained from two healthy humans. 17,668 contigs were assembled into 14,572 scaffolds, totaling 33 million bp. 40% of reads could not be assembled into contigs due to low depth of coverage (singletons). Coverage average depth in contigs: 2.13X

Isolation Source

The human intestinal microbiota is composed of 10^{13} to 10^{14} microorganisms whose collective genome ("microbiome") contains at least 100 times as many genes as our own genome. 78 million base pairs of unique DNA sequence and 2062 polymerase chain reaction-amplified 16S ribosomal DNA sequences were obtained from the fecal DNAs of two healthy adults.

Subject 7	Fecal specimen from 28 year old healthy female
Subject 8	Fecal specimen from 37 year old healthy male

Both subjects had not used antibiotics of any other medications during the year before specimen collection.

Genome Assembly

None

References

Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE. Metagenomic analysis of the human distal gut microbiome. Science. 2006 Jun 2;312(5778): 1355-9. PubMed PMID: 16741115.

Microbial Genomics Home page at The Institute for Genomic Research

Enhanced Biological Phosphorus Removal (EBPR) Sludge Community

Enhanced biological phosphorus removal (EBPR) is a treatment process in which microorganisms remove excessive inorganic phosphate from wastewater. However, the metabolic traits of this process are not well understood. Metagenomic analysis of two sludge samples allowed for the comparison of the microbial communities as well as the isolation of a near-complete genome of Acinetobacter phosphatis.

Data

WGS study	EBPR sludge from Australia	AATN00000000
	EBPR sludge from USA	AATO00000000
Contigs	EBPR sludge from Australia	AATN01000001-AATN01011188
	EBPR sludge from USA	AATO01000001-AATO01015866
Scaffolds	No scaffolds available	
Trace data	EBPR sludge from Australia	96,563 reads available from the Trace Archive
	EBPR sludge from USA	127,953 reads available from the Trace Archive
16S ribosomal RNA	No 16S rRNA sequences available	

Approximately 98 and 78 Mbp of shotgun sequence data were obtained from the US and OZ (Australia) sludge, respectively.

Isolation Source

Australia	Activated sludge from Thornside Sewage Treatment Plant in Queensland, Australia on August 18, 2004. The reactor was operated in four cycles of 6h per day, including 150 min anaerobic period, 180 min aerobic period, 30 min settling, and 5 min effluent withdrawing. The pH was kept at 7.0. The SBR was fed with propionate and a synthetic feed and nitrogen gas and air was bubbled through the liquid.
USA	Activated sludge mixed liquor from the Nine Springs Wastewater Treatment Plant in Madison, WI, USA on July 3, 2004. The reactor was operated in four cycles of 6h per day, including 130 min anaerobic phase, 190 min aerobic phase, 30 min settling, and 10 min effluent withdrawing. The pH was in the range of 7.0-7.3. The sequencing batch reactor (SBR) was fed with acetate, casamino acids, and a mineral salts medium with sodium phosphate. The SBR was operating for 11 months at the time of sampling.

Genome Assembly

None

References

Martin HG, Ivanova N, Kunin V, Warnecke F, Barry KW, McHardy AC, Yeates C, He S, Salamov AA, Szeto E, Dalin E, Putnam NH, Shapiro HJ, Pangilinan JL, Rigoutsos I, Kyrpides NC, Blackall LL, McMahon KD, Hugenholtz P. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat Biotechnol. 2006 Oct;24(10):1263-1269. PubMed PMID: 16998472.

Mouse Gut Microbiota Metagenome

Comparisons of the distal gut microbiota of genetically obese mice and their lean littermates, as well as those of obese and lean human volunteers have revealed that obesity is associated with changes in the relative abundance of the two dominant bacterial divisions, the Bacteroidetes and the Firmicutes. Comparative metagenomic analyses was used to examine how these changes affect the metabolic potential of the mouse gut microbiome.

Data

WGS study	Lean mouse 1 (+/+; Litter 1)	AATA0000000
	Lean mouse 2 (ob/+; Litter 2)	AATB0000000
	Lean mouse 3 (+/+; Litter 2)	AATC00000000
	Obese mouse 1 (ob/ob; Litter 1)	AATD00000000
	Obese mouse 2 (ob/ob; Litter 2)	AATE00000000
	Combined datasets	AATF00000000
Contigs	Lean mouse 1 (+/+; Litter 1)	AATA0100001-AATA01010733
	Lean mouse 2 (ob/+; Litter 2)	AATB0100001-AATB01011335
	Lean mouse 3 (+/+; Litter 2)	AATC01000001-AATC01010377
	Obese mouse 1 (ob/ob; Litter 1)	AATD01000001-AATD01010948
	Obese mouse 2 (ob/ob; Litter 2)	AATE01000001-AATE01008828
	Combined datasets	AATF01000001-AATF01013667
Scaffolds	No scaffolds available	
Trace Data	Lean mouse 1	1,057,022 reads available from the Trace Archive
	Obese mouse 1	687,261 reads available from the Trace Archive
16S ribosomal RNA	Microbiota transplant donors and recipients	EF095962-EF100118

A total of 54,991 high-quality sequence reads were obtained from three lean (+/+, ob/+) mice and two genetically obese (ob/ob) mice.

.

Isolation Source

DNA was isolated from the distal gut (ceca) of eight-week old C57BL/6J ob/ob, ob/+, and +/+ mice using a bead beater to mechanically disrupt cells, followed by phenol-chloroform extraction. Mice were housed individually in microisolater cages where they were maintained in a specified pathogen-free state, under a 12h light cycle, and fed a standard polysaccharide-rich diet (Picolab, Purina) ad libitum.

Genome Assembly

None

References

Turnbaugh PJ, Ley RE, Mahowald MA, Mabrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006 Dec 21;444(7122):1027-1031. PubMed PMID: 17183312.

Mediterranean Gutless Worm Metagenome

Olavius algarvensis is a marine worm that belongs to a group of oligochaetes that lack a mouth, gut, anus and nephridia (kidney-like organs). These worms live in obligate and species-specific associations with multiple extracellular bacterial endosymbionts that are located just below the worm cuticle. A metagenomic approach was used to isolate four co-occurring bacterial symbionts. These symbionts were taxonomically identified and shown to be capable of carbon fixation, which provides the host with multiple sources of nutrition

Data

WGS study	AASZ00000000
Contigs	AASZ01000001-AASZ01005597
Scaffolds	DS021107-DS021197
	DS021198-DS021219
	DS021220-DS021445
	DS021446-DS021617
	DS021618-DS022223
Trace data	313,773 reads available from the Trace Archive
16S ribosomal RNA	No 16S rRNA data available

Approximately 60 ug of metagenomic DNA was purified per 200 frozen specimens. 185 million bases of 3 kb library end sequence and 19 Mb of fosmid end sequence was generated for a total of 204 Mb of high-quality shotgun sequence data. These metagenomic shotgun libraries contained a total of 281,448 trimmed sequence reads. 61% of the reads were assembled into a filtered set containing 2,286 scaffolds with 39 Mb of sequence, of which 40% was gap. Binning of the Olavius spp. symbionts' metagenome results in 511 scaffolds forming four distinct clusters.

Isolation Source

Juvenile and adult specimens of the marine oligochaetes Olavius algarvenis and Olavius ilvae (co-occurring gutless oligochaete species) were collected in May and September 2004 from 5.6 m water depth in silicate sediments around sea grass beds in a bay off Capo di Sant' Andrea, Elba, Italy (42 deg 48' 26" N, 010 deg 08' 28" E). Samples were enriched for bacterial cells from approximately 200 live worms.

Genome Assembly

Binning of the 2,286 scaffolds was based on GC-content, dinucleotide relative abundance, Markov model-based statistical evaluation of tri-, tetra and pentamer over- and under representation, and normalized chaos game representations for tri- to hexamers Deschavanne. This resulted in 511 scaffolds distributed into four distinct clusters, identified based on 16S ribosomal RNA genes and phylogenetic analysis of predicted proteins within each cluster of scaffolds.

Assembly	Sequences	Notes
Olavius algarvensis Gamma 1 endosymbiont	DS021107-DS021197	Contains genes required for autotrophic CO2 fixation, reduced sulphur oxidation and globule storage
Olavius algarvensis Gamma 3 endosymbiont	DS021198-DS021219	Sulphur-oxidizing chemoautotroph
Olavius algarvensis Delta 1 endosymbiont	DS021220-DS021445	Contains genes coding for the transport and utilization of a variety of carbohydrate substrates

Table continued from previous page.

Olavius algarvensis Delta 4 endosymbiont	DS021446-DS021617	Reduces sulphur compounds of intermediate oxidation states
symbiont metagenome	DS021618-DS022223	Possibly derived from endosymbionts of Olavius ilvae and/or other "contaminating" species

References

Woyke T, Teeling H, Inanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson IJ, Barry KW, Shapiro HJ, Szeto E, Kyrpides NC, Mussmann M, Amann R, Bergin C, Ruehland C, Rubin EM, Dubilier N. Symbiosis insights through metagenomic analysis of a microbial consortium. Nature. 2006 Oct 26;443(7114): 950-5.

Mammuthus primigenius (Woolly Mammoth) Metagenome

Fossil DNA from preserved remains of woolly mammoth samples was isolated and compared to the genomes of different mammals in order to identify sequence reads derived from the woolly mammoth and to eliminate reads that were contamination. Comparisons to the African elephant (Loxodonta africana) support the paleontologically based divergence date of 5-6 million years.

Data

WGS study	CAAM00000000
Contigs	CAAM02000001-CAAM02064265
Scaffolds	No scaffolds available
Trace data	366,957 reads available from the Trace Archive
16S ribosomal RNA	No 16S rRNA data available

Sequence reads were generated by 454 sequencing and queried to the elephant genome. Sequence reads whose best hits were against the elephant genome (or against the human genome to screen for human contaminants) were further analyzed. Only alignments greater then 30 nt were studied.

DQ188829 represents a complete mitochondrial genome derived from a mammoth bone (calibrated age between 11,900-13,400 years) found in Berelekh, Yakutia (71 deg N, 145 deg E) Russia.

EU153446, EU153448, EU153450, EU153451, EU153453, and EU155210 are additional complete mitochondrial genomes derived from mammoth.

Isolation Source

Contributor	Isolation source	Notes
CCGB	Edentulous mandible remains (dated 27,740 +/- 220 ^{14}C years) from the shore of Baikura-turku, a large bay on the southeastern side of Lake Taimyr in Siberia	302,692 reads averaging 95 bp and resulting in 28 million bp were isolated and submitted to the Trace Archive. 13 million bp (45.4%) of the sequencing reads were identified as mammoth DNA.
Max Planck Institute	43,000 year old mammoth femur from Bol'shaya Kolopatkaya River (70 deg N, 151 deg E) in Russia	64,265 reads were submitted to the Trace Archive and included in project CAAM00000000

Genome Assembly

None

References

Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buiques B, Tikhonov A, Huson DH, Tomsho LP, Auch A, Rampp M, Miller W, Schuster SC. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science. 2006;311(5759):392–4. PubMed PMID: 16368896.

Krause J, Dear PH, Pollack JL, Slatkin M, Spriggs H, Barnes I, Lister AM, Ebersberger I, Paabo S, Hofreiter M. Multiplex amplification of the mammoth mitochondrial genomes and the evolution of Elephantidae. Nature. 2006;439(7077):724–7. PubMed PMID: 16362058.

Stiller M, Green RE, Ronan M, Simons JF, Du L, He W, Egholm M, Rothberg JM, Keates SG, Ovodov ND, Antipina EE, Baryshnikov GF, Kuzmin YV, Vasilevski AA, Wuenschell GE, Termini J, Hofreiter M, Jaenicke-Despres V, Paabo S. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. Proc Natl Acad Sci USA. 2006;103(37):14977.

Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prufer K, Meyer M, Krause J, Ronan MT, Lachmann M, Paabo S. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci USA. 2007;104(37):14616–21. PubMed PMID: 17715061.