Complete Microbial Genomes, Epigenomes, and Transcriptomes using Long-Read PacBio® Sequencing

Tyson A. Clark
Pacific Biosciences, Menlo Park, CA

PACIFIC BIOSCIENCES®

Single Molecule, Real-Time (SMRT®) DNA Sequencing

High Consensus Accuracy

Achieves >99.999% (QV50)
Lack of systematic bias

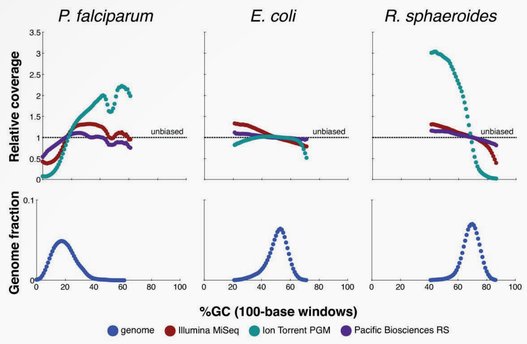
Table 3 Genome assembly continuity and correctness comparison to secondary technologies

Organism	Assembled with	Assembly bp	Contigs	N50	LAP	Discordant bases	QV
E. coli K12	MiSeq 100× 2×150 bp 300 bp (MaSuRCA iCORN)	4,682,345	139	113,852	−9.68E + 07	28	52.23
	454 50×	4,569,757	93	117,490	−9.73E + 07	17	54.29
	PBcR 200×	**4,653,486**	**1**	**4,653,486**	**−9.64E + 07**	**3**	**>60**
E .coli O157:H7	MiSeq 100× 2×150 bp 500 bp (SPAdes iCORN)	5,433,737	413	133,641	−3.67E + 07	62	49.43
	454 22× + 8× 5 kbp + 10× 10 kbp	5,347,050	409	133,665	−3.73E + 07	66	49.09
	PBcR 200×	**5,611,389**	**9**	**4,324,437**	**−3.66E + 07**	**0**	**>60**
B. trehalosi	MiSeq 100× 2×150 bp 500 bp (SPAdes iCORN)	2,377,594	83	222,446	−3.31E + 07	10	53.76
	454 50×	2,364,704	66	117,742	−3.32E + 07	9	54.20
	PBcR 200×	**2,411,068**	**1**	**2,411,068**	**−3.27E + 07**	**0**	**>60**
M. haemolytica	MiSeq 100× 2×150 bp 500 bp (MaSuRCA iCORN)	2,721,965	89	84,094	−3.33E + 07	47	47.63
	PBcR 200×	**2,736,037**	**1**	**2,736,037**	**−3.31E + 07**	**0**	**>60**
F. tularensis	MiSeq 100× 2×250 bp 500 bp (SPAdes iCORN)	1,825,374	130	24,065	−1.33E + 07	0	>60
	454 50×	1,655,657	326	7,316	−1.33E + 07	28	47.72
	PBcR 300×	**1,877,407**	**3**	**573,021**	**−1.33E + 07**	**0**	**>60**
S. enterica Newport	MiSeq 56× 2×150 bp 500 bp (MaSuRCA iCORN)	5,187,269	114	195,780	−2.24E + 07	360	41.59
	454 23× + 2× 10 kbp	5,005,089	172	372,513	−2.25E + 07	39	51.08
	PBcR 200×	**5,029,197**	**2**	**4,919,684**	**−2.24E + 07**	**2**	**>60**

Reducing assembly complexity of microbial genomes with single-molecule sequencing, *Genome Biology*, Koren et al., (2013) **14**:R101 doi:10.1186/gb-2013-14-9-r101

Lack of Sequence Context Bias

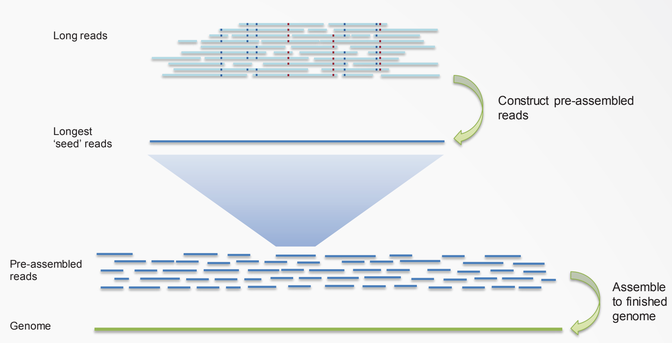
GC content
Low complexity sequence



Characterizing and measuring bias in sequence data, *Genome Biology*, Ross et al., (2013) **14**:R51 doi:10.1186/gb-2013-14-5-r51

Long Sequence Reads

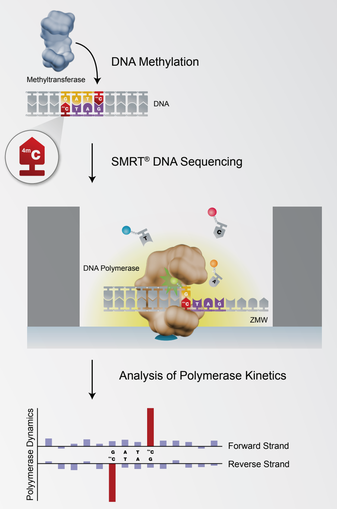
Finished microbial genomes
Improve large, complex genomes
Full-length cDNA sequencing
Long-range haplotype phasing



Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data, *Nature Methods*, Chin et al., (2013) **10**: 563–569

Base Modification Detection

Epigenome characterization

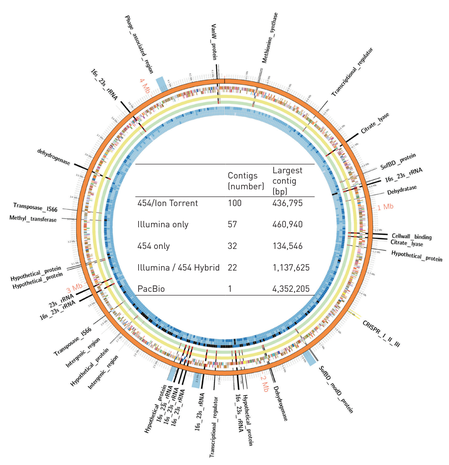


The methylomes of six bacteria, *Nucleic Acids Research*, Murray et al., (2012) **40**: 11450-11462

Applications for Natural Product Discovery

Industrial Microbioloy

- *Clostridium autoethanogenum* is capable of fermenting waste gases (CO, H_2, CO_2) into ethanol & commodity chemicals
- Previous draft genome in ~100 contigs
- Closed, high-quality genome generated from only PacBio data without manual finishing

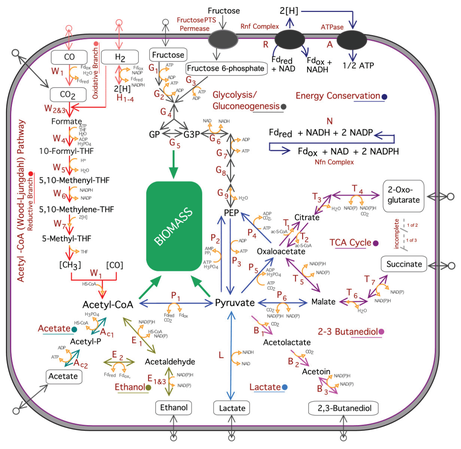


Genes missed by short-read assemblies, but present in PacBio assembly:

Table 3: Regions of low sequence-coverage

Locus tag	Start[a]	End[a]	Product description	Pacbio coverage [×]	454 Coverage [×]	Illumina coverage [×]	454 Hybrid contig coverage[1]	Draft assembly contig coverage
CAETHG_0602	663332	664234	Citrate lyase, beta subunit	111	29	65	None	None
CAETHG_0603	664234	664530	Citrate lyase acyl carrier protein	107	29	63	None	None
CAETHG_0604	664553	665587	Citrate lyase ligase	109	23	63	None	Partial
CAETHG_0605	665628	666806	Malic protein NAD-binding protein	101	27	69	None	None
Intergenic	827340	827520	NA	106	30	53	None	None
rRNA	885055	887942	23s_rRNA	87	77	147	None	None
rRNA	888206	889703	16s_rRNA	102	56	165	None	None
CAETHG_1038	1116305	1121431	Cell wall binding repeat 2-containing protein	127	27	69	Partial	None
CAETHG_1052	1136476	1138017	Citrate lyase, alpha subunit	107	22	53	Partial	None
CAETHG_1055	1139370	1140533	Malic protein NAD-binding protein	107	27	51	Partial	Partial
rRNA	2114155	2117042	23s_rRNA	122	81	161	None	None
rRNA	2117334	2118831	16s_rRNA	118	66	128	None	None
CAETHG_2076	2220169	2221506	Sigma54 specific transcriptional regulator, Fis family	122	32	85	Partial	Partial
CAETHG_2077	2221658	2221885	Transcriptional regulator, Fis family	126	21	92	Partial	None
CAETHG_2078	2222014	2222994	Putative sigma54 specific transcriptional regulator	135	30	77	Partial	Partial
rRNA	2271738	2273235	16s_rRNA	165	10	26	None	None
rRNA	2273527	2276414	23s_rRNA	158	10	26	None	None
rRNA	2355334	2356831	16s_rRNA	145	11	24	None	None
rRNA	2357123	2360010	23s_rRNA	136	13	23	None	None
rRNA	2372238	2373735	16s_rRNA	128	13	21	None	None
rRNA	2374027	2376914	23s_rRNA	126	14	19	None	None
rRNA	2392702	2394199	16s_rRNA	134	12	20	None	None
rRNA	2394596	2397483	23s_rRNA	142	11	21	None	None
CAETHG_2621	2823723	2824328	Transposase IS66	127	30	52	Partial	Partial
rRNA	2935186	2936683	16s_rRNA	127	14	27	None	None
tRNA	2936973	2937045	tRNA_Ala	125	19	51	None	None
tRNA	2937053	2937126	tRNA_Ile	125	26	58	None	None
rRNA	2937443	2940330	23s_rRNA	117	14	28	None	None
rRNA	2966992	2968489	16s_rRNA	126	11	20	None	None
tRNA	2968779	2968851	tRNA_Ala	132	20	50	None	None
tRNA	2968859	2968932	tRNA_Ile	131	23	70	None	None
rRNA	2969222	2972109	23s_rRNA	128	10	19	None	None
rRNA	3872016	3873511	16s_rRNA	98	10	18	None	None
rRNA	3873937	3876824	23s_rRNA	107	14	21	None	None

Full metabolic reconstruction possible



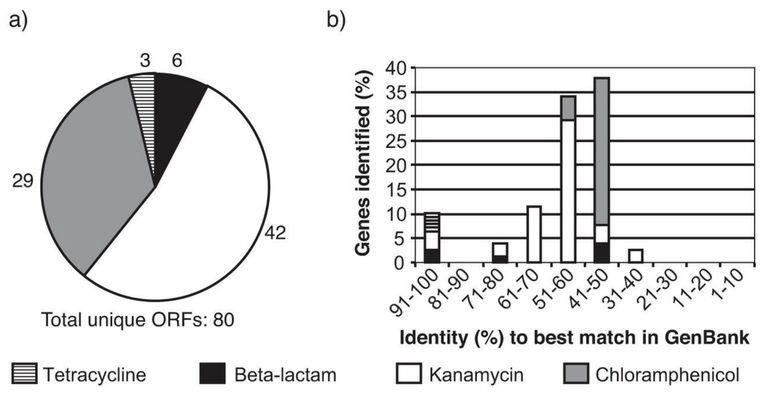
Inferred metabolism of *C. autoethanogenum*.

Brown ***et al.*** (2014) Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of Clostridium autoethanogenum and analysis of CRISPR systems in industrial relevant Clostridia. ***Biotechnology for Biofuels*** **7**:40

Livestock-Microbe Interactions

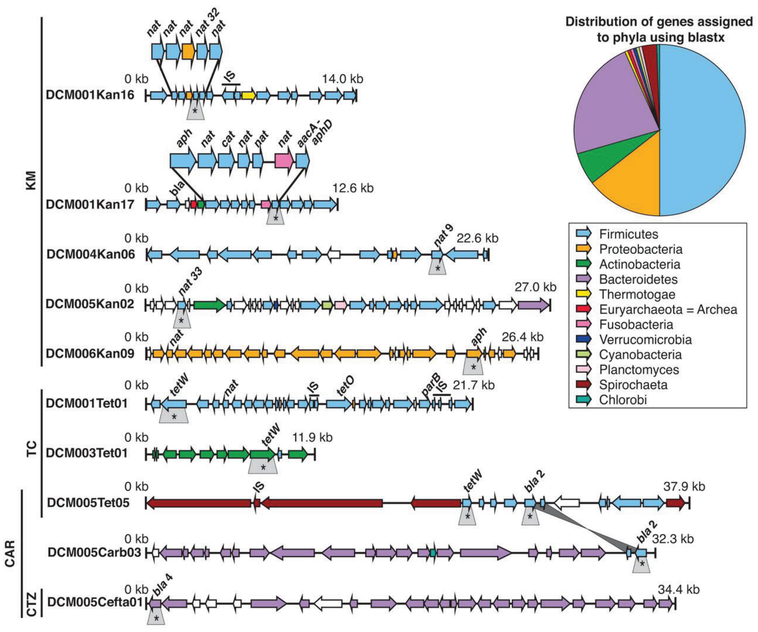
- Microbiomes of farm animals are reservoirs of antibiotic resistance genes
- Discovered a novel clade of chloramphenicol acetyltransferases
- Significantly extends the roster of functional antibiotic resistance genes found in animal gut bacteria

Antibiotic resistance genes from dairy cow manure.



Antibiotic resistance genes from dairy cow manure. (a) Distribution of 80 unique antibiotic resistance (AR) genes among four classes of antibiotics. (b) Distribution of similarity of 80 unique AR genes from manure compared to homologues in GenBank.

Organization of genes on ten metagenomic fosmid clones conferring resistance to kanamycin (KM), tetracycline (TC), carbenicillin (CAR), or ceftazidime (CTZ) assessed using Pacific Biosciences RS sequencing technology.

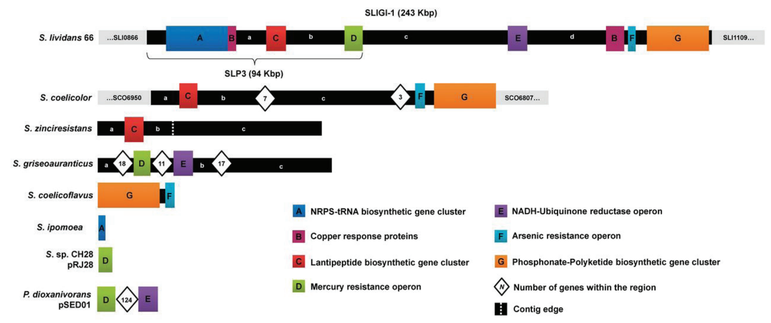


Organization of genes on ten metagenomic fosmid clones conferring resistance to kanamycin (KM), tetracycline (TC), carbenicillin (CAR), or ceftazidime (CTZ) assessed using Pacific Biosciences RS sequencing technology. Genes are as follows: *nat*, *N*-acetyltransferases; aph, aminoglycoside phosphotransferases; aacA-*aphD*, bifunctional aminoglycoside-modifying enzyme; *tetW* and *tet0*, tetracycline resistance genes; *cat*, chloramphenicol acetyltransferase; *bla*, beta-lactamases. Genes with annotations related to mobile genetic elements are labeled as IS (insertion sequence elements/transposases) or*parB* (plasmid-partitioning protein). Identical genes are connected with gray lines. Antibiotic resistance genes for which we demonstrated a function by subcloning are indicated with gray boxes and asterisks. All other labels were assigned based on sequence homology.

Wichmann *et al.* (2014) Diverse antibiotic resistance genes in dairy cow manure. *mBio* **5**:e01017-13

Bioremediation

- High-quality genome sequence of a widely used actinomycete model organism, *Streptomyces lividans 66*
- High-quality genomes aid in identifying genetic diversity amongst closely related strains that otherwise would have been lost
- Resolves large mobile genomic island related to metal metabolism, which harbors the elusive plasmid SLP3
- SLP3 encodes enzymes that direct the synthesis of natural products that remain to be discovered, even in such well-studied strains as *S. lividans 66*



Cruz-Morales *et al.* (2013) The genome sequence of Streptomyces lividans 66 reveals a novel tRNA-dependent peptide biosynthetic system within a metal-related genomic island. *Genome Biology & Evolution* **5**:1165-1175

Other Examples

Biocommodities:

Identification of Restriction-Modification Systems of Bifidobacterium animalis subsp. lactis CNCM I-2494 by SMRT Sequencing and Associated Methylome Analysis http://dx.plos.org/10.1371/journal.pone.0094875.g006

Complete Genome Sequence of Enterococcus mundtii QU 25, an Efficient L-(+)-Lactic Acid-Producing Bacterium http://dnaresearch.oxfordjournals.org/content/early/2014/03/10/dnares.dsu003.full

Biofuels:

The genome of the anaerobic fungus Orpinomyces sp. strain C1A reveals the unique evolutionary history of a remarkable plant biomass degrader http://aem.asm.org/content/early/2013/05/20/AEM.00821-13.full.pdf (fungal)

Localized electron transfer rates and microelectrode-based enrichment of microbial communities within a phototrophic microbial mat http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3902354/ (16S)

Long-term operation of microbial electrosynthesis systems improves acetate production by autotrophic microbiomes http://pubs.acs.org/doi/abs/10.1021/es400341b (16S)

Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of Verrucomicrobia http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0035314

Bioremediation:

Exploring the roles of DNA methylation in the metal-reducing bacterium Shewanella oneidensis MR-1 http://jb.asm.org/content/195/21/4966.short

A crowd sourced funded project to sequence a fern. https://experiment.com/projects/azolla-a-little-fern-with-massive-green-potential/community

Genome sequence of Candidatus Microthrix parvicella Bio17-1, a long-chain-fatty-acid-accumulating filamentous actinobacterium from a biological wastewater treatment plant http://jb.asm.org/content/194/23/6670.long

New bacteria with potential commercial applications:

Single cell genomic study of Dehalococcoidetes species from deep-sea sediments of the Peruvian Margin http://www.wafergen.com/wp-content/uploads/2014/03/ismej201424a.pdf

Genome Sequence of Bacillus pumilus MTCC B6033 http://genomea.asm.org/content/2/2/e00327-14.full

Antibiotic discovery:

De Novo Assembly of the Streptomyces sp. Strain Mg1 Genome UsingPacBio Single-Molecule Sequencing. http://genomea.asm.org/content/1/4/e00535-13.full.pdf+html

Value of a newly sequenced bacterial genome. http://www.wjgnet.com/1949-8454/full/v5/i2/161.htm

Livestock/plant microbiome interactions:

Genome Sequence and Methylome of Soil Bacterium *Gemmatirosa kalamazoonensis* KBS708T, a Member of the Rarely Cultivated *Gemmatimonadetes* Phylum http://genomea.asm.org/content/2/2/e00226-14.full

Genome Sequence of Pseudomonas sp. Strain P482, a Tomato Rhizosphere Isolate with Broad-Spectrum Antimicrobial Activity. http://www.ncbi.nlm.nih.gov/pubmed/24970823

pacb.com

See over 200 customer publications at www.pacb.com/publications