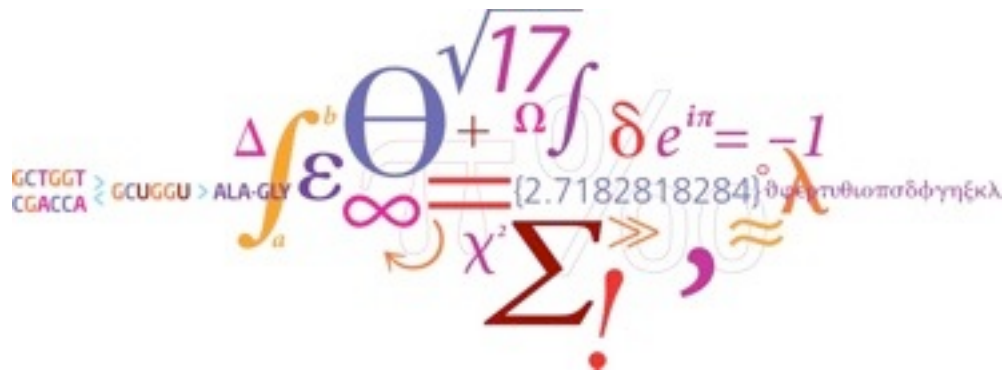


2011 - Bacterial Genome Groups



Dave Ussery

DTU course 27104 - Comparative Genomics
Collections of genomes for groups

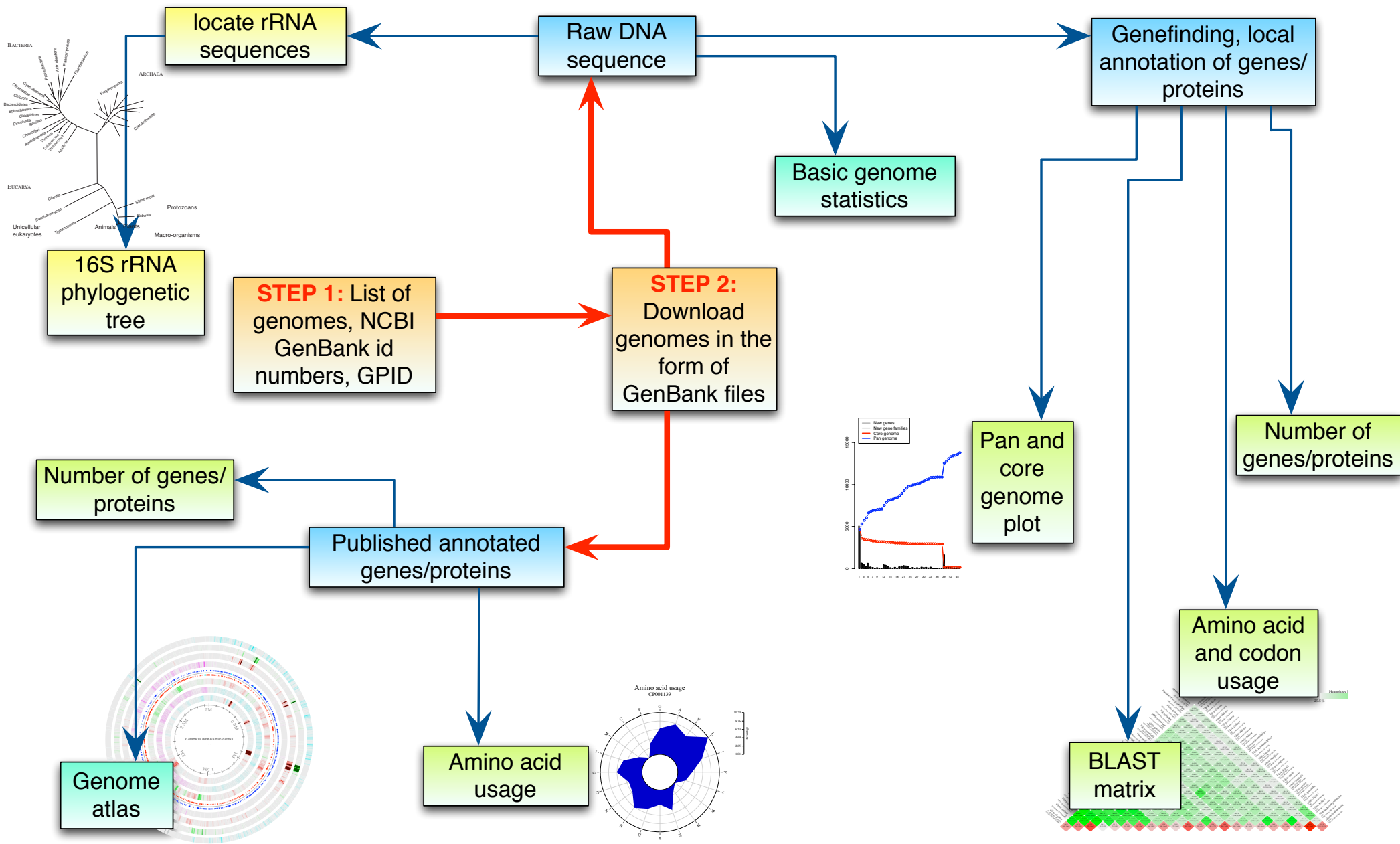
31 August, 2011

<http://www.cbs.dtu.dk/dtucourse/genomics27104.php>

September	October	November
<p>Journal Clubs</p> <p>Learn tools to compare genomes</p> <p>Lectures on textbook & scientific communication</p>	<p>Posters</p> <p>Posters due Wed. 26 Okt., 2011</p> <p>Poster Party Thurs. 27 Okt., 2011</p>	<p>Papers</p> <p>Papers due 17 Nov., 2011</p> <p>Referee papers</p> <p>Final, revised Papers due 1 Dec., 2011</p>

What is expected?

- **Journal clubs** - convince me you've read and understood the article. Good critique. Follow template for my journal club (tomorrow, 1 September).
- **Posters** - Display figures made in class
- **Papers** - Tell story. Introduction, Methods, Results, Discussion, References.



Bacterial genera

- for course 27104 groups Autumn, 2011

Group 01 *Acinetobacter*

Group 02 *Salmonella*

Group 03 *Lactobacillus*

Group 04 *Bifidobacterium*

Group 05 *Pseudomonas*

Group 06 *Bacteroides*

Group 07 *Brucella*

Group 08 *Campylobacter*

Group 09 *Escherichia*

Group 11 *Vibrio*

Group 12 *Burkholderia*

Group 13 *Staphylococcus*

Group 14 *Bacillus*

Organism	GPID	status
Group 01 Acinetobacter		
<i>Acetobacter pasteurianus</i> IFO 3283-01	31129	complete
<i>Acetobacter pasteurianus</i> IFO 3283-01-42C	31141	complete
<i>Acetobacter pasteurianus</i> IFO 3283-03	31131	complete
<i>Acetobacter pasteurianus</i> IFO 3283-07	31133	complete
<i>Acetobacter pasteurianus</i> IFO 3283-12	32203	complete
<i>Acetobacter pasteurianus</i> IFO 3283-22	31135	complete
<i>Acetobacter pasteurianus</i> IFO 3283-26	31137	complete
<i>Acetobacter pasteurianus</i> IFO 3283-32	31139	complete
<i>Acetobacter acetii</i> NBRC 14818	52649	assembly
<i>Acetobacter pomorum</i> DM001	60787	assembly
<i>Acetobacter tropicalis</i> NBRC 101654	46891	assembly

<http://www.pnas.org/content/108/33/13758.abstract?etoc>

Published online before print August 8, 2011, doi: 10.1073/pnas.1104404108
 PNAS August 16, 2011 vol. 108 no. 33 13758-13763

Genome-wide recombination drives diversification of epidemic strains of *Acinetobacter baumannii*

Evan S. Snitkin, Adrian M. Zelazny, Clemente I. Montero, Frida Stock, Lilia Mijares, NISC Comparative Sequence Program, Patrick R. Murray, and Julie A. Segre

Abstract

Acinetobacter baumannii is an emerging human pathogen and a significant cause of nosocomial infections among hospital patients worldwide. The enormous increase in multidrug resistance among hospital isolates and the recent emergence of pan-drug-resistant strains underscores the urgency to understand how *A. baumannii* evolves in hospital environments. To this end, we undertook a genomic study of a polyclonal outbreak of multidrug-resistant *A. baumannii* at the research-based National Institutes of Health Clinical Center. Comparing the complete genome sequences of the three dominant outbreak strain types enabled us to conclude that, despite all belonging to the same epidemic lineage, the three strains diverged before their arrival at the National Institutes of Health. The simultaneous presence of three divergent strains from this lineage supports its increasing prevalence in international hospitals and suggests an ongoing adaptation to the hospital environment. Further genomic comparisons uncovered that much of the diversification that occurred since the divergence of the three outbreak strains was mediated by homologous recombination across 20% of their genomes. Inspection of recombinant regions revealed that several regions were associated with either the loss or swapping out of genes encoding proteins that are exposed to the cell surface or that synthesize cell-surface molecules. Extending our analysis to a larger set of international clinical isolates revealed a previously unappreciated ability of *A. baumannii* to vary surface molecules through horizontal gene transfer, with subsequent intraspecies dissemination by homologous recombination. These findings have immediate implications in surveillance, prevention, and treatment of *A. baumannii* infections.

Organism	GPID	status
Group 02 Salmonella		
Salmonella enterica subsp. enterica serovar Agona str. SL483	20063	complete
Salmonella enterica subsp. enterica serovar Choleraesuis str. SC-B67	9618	complete
Salmonella enterica subsp. enterica serovar Dublin str. CT_02021853	19467	complete
Salmonella enterica subsp. enterica serovar Enteritidis str. P125109	30687	complete
Salmonella enterica subsp. enterica serovar Gallinarum str. 287/91	30689	complete
Salmonella enterica subsp. enterica serovar Heidelberg str. SL476	20045	complete
Salmonella enterica subsp. enterica serovar Newport str. SL254	18747	complete
Salmonella enterica subsp. enterica serovar Paratyphi A str. AKU_12601	30943	complete
Salmonella enterica subsp. enterica serovar Paratyphi C strain RKS4594	20993	complete
Salmonella enterica subsp. enterica serovar Schwarzengrund str. CVM19633	19459	complete
Salmonella enterica subsp. enterica serovar Typhi str. CT18	236	complete
Salmonella enterica subsp. enterica serovar Typhi str. Ty2	371	complete
Salmonella enterica subsp. enterica serovar Typhimurium str. 14028S	33067	complete
Salmonella enterica subsp. enterica serovar Typhimurium str. LT2	241	complete
Salmonella enterica subsp. enterica serovar Typhimurium str. SL1344	50407	complete
Salmonella enterica subsp. enterica serovar Typhimurium str. ST4/74	56087	complete
Salmonella enterica subsp. enterica serovar Typhimurium str. T000240	45951	complete
Salmonella enterica subsp. enterica serovar Typhimurium str. UK-1	63211	complete
Salmonella enterica subsp. arizonae serovar 62:z4,z23:-- str. RSK2980	13030	complete
Salmonella bongori NCTC 12419	351	complete

Microb Ecol. 2011 Jun 4.

The *Salmonella enterica* Pan-genome.

Jacobsen A, Hendriksen RS, Aaresturp FM, Ussery DW, Friis C.

Abstract

Salmonella enterica is divided into four subspecies containing a large number of different serovars, several of which are important zoonotic pathogens and some show a high degree of host specificity or host preference. We compare 45 sequenced *S. enterica* genomes that are publicly available (22 complete and 23 draft genome sequences). Of these, 35 were found to be of sufficiently good quality to allow a detailed analysis, along with two *Escherichia coli* strains (K-12 substr. DH10B and the avian pathogenic *E. coli* (APEC O1) strain). All genomes were subjected to standardized gene finding, and the core and pan-genome of *Salmonella* were estimated to be around 2,800 and 10,000 gene families, respectively. The constructed pan-genomic dendrograms suggest that gene content is often, but not uniformly correlated to serotype. Any given *Salmonella* strain has a large stable core, whilst there is an abundance of accessory genes, including the *Salmonella* pathogenicity islands (SPIs), transposable elements, phages, and plasmid DNA. We visualize conservation in the genomes in relation to chromosomal location and DNA structural features and find that variation in gene content is localized in a selection of variable genomic regions or islands. These include the SPIs but also encompass phage insertion sites and transposable elements. The islands were typically well conserved in several, but not all, isolates—a difference which may have implications in, e.g., host specificity.

PMID: 21643699

Organism	GPID	status
Group 03 Lactobacillus		
Lactobacillus acidophilus 30SC	53145	complete
Lactobacillus amylovorus GRL 1112	42073	complete
Lactobacillus brevis ATCC 367	404	complete
Lactobacillus buchneri NRRL B-30929	29003	complete
Lactobacillus casei ATCC 334	402	complete
Lactobacillus casei BL23	30359	complete
Lactobacillus casei str. Zhang	28537	complete
Lactobacillus crispatus ST1	46813	complete
Lactobacillus delbrueckii subsp. bulgaricus ATCC 11842	16871	complete
Lactobacillus fermentum IFO 3956	18979	complete
Lactobacillus gasseri ATCC 33323	84	complete
Lactobacillus helveticus DPC 4571	17811	complete
Lactobacillus johnsonii FI9785	36575	complete
Lactobacillus kefiranofaciens ZW3	66525	complete
Lactobacillus plantarum JDM1	32969	complete
Lactobacillus reuteri DSM 20016	15766	complete
Lactobacillus rhamnosus GG	32195	complete
Lactobacillus sakei subsp. sakei 23K	13435	complete
Lactobacillus salivarius UCC118	13280	complete
Lactococcus lactis subsp. cremoris MG1363	18797	complete

Genes Nutr. 2011 Aug;6(3):319-40. Epub 2010 Oct 28.

Group-specific comparison of four lactobacilli isolated from human sources using differential blast analysis.

Altermann E, Klaenhammer TR.

Abstract

Lactic acid bacteria (LAB) have been used in fermentation processes for centuries. More recent applications including the use of LAB as probiotics have significantly increased industrial interest. Here we present a comparative genomic analysis of four completely sequenced Lactobacillus strains, isolated from the human gastrointestinal tract, versus 25 lactic acid bacterial genomes present in the public database at the time of analysis. Lactobacillus acidophilus NCFM, Lactobacillus johnsonii NCC533, Lactobacillus gasseri ATCC33323, and Lactobacillus plantarum WCFS1 are all considered probiotic and widely used in industrial applications. Using Differential Blast Analysis (DBA), each genome was compared to the respective remaining three other Lactobacillus and 25 other LAB genomes. DBA highlighted strain-specific genes that were not represented in any other LAB used in this analysis and also identified group-specific genes shared within lactobacilli. Initial comparative analyses highlighted a significant number of genes involved in cell adhesion, stress responses, DNA repair and modification, and metabolic capabilities. Furthermore, the range of the recently identified potential autonomous units (PAUs) was broadened significantly, indicating the possibility of distinct families within this genetic element. Based on in silico results obtained for the model organism L. acidophilus NCFM, DBA proved to be a valuable tool to identify new key genetic regions for functional genomics and also suggested re-classification of previously annotated genes.

PMID: 21484153

Organism	GPID	status
Group 04 Bifidobacterium		
Bifidobacterium adolescentis ATCC 15703	16321	complete
Bifidobacterium animalis subsp. lactis AD011	19423	complete
Bifidobacterium animalis subsp. lactis BB-12	42883	complete
Bifidobacterium animalis subsp. lactis BI-04	32897	complete
Bifidobacterium animalis subsp. lactis DSM 10140	32893	complete
Bifidobacterium animalis subsp. lactis V9	32515	complete
Bifidobacterium bifidum PRL2010	42863	complete
Bifidobacterium bifidum S17	51963	complete
Bifidobacterium breve ACS-071-V-Sch8b	51077	complete
Bifidobacterium breve UCC2003	13487	complete
Bifidobacterium dentium Bd1	17583	complete
Bifidobacterium longum DJO10A	18773	complete
Bifidobacterium longum NCC2705	328	complete
Bifidobacterium longum subsp. infantis 157F	32051	complete
Bifidobacterium longum subsp. infantis ATCC 15697	17189	complete
Bifidobacterium longum subsp. longum BBMN68	53143	complete
Bifidobacterium longum subsp. longum F8	45963	complete
Bifidobacterium longum subsp. longum JCM 1217	32047	complete
Bifidobacterium longum subsp. longum JDM301	47579	complete
Bifidobacterium longum subsp. longum KACC 91563	66401	complete

Res Microbiol. 2011 Sep;162(7):664-70.

Analysis of infant isolates of *Bifidobacterium breve* by comparative genome hybridization indicates the existence of new subspecies with marked infant specificity.

Boesten R, Schuren F, Wind RD, Knol J, de Vos WM.

Abstract

A total of 20 *Bifidobacterium* strains were isolated from fecal samples of 4 breast- and bottle-fed infants and all were characterized as *Bifidobacterium breve* based on 16S rRNA gene sequence and metabolic analysis. These isolates were further characterized and compared to the type strains of *B. breve* and 7 other *Bifidobacterium* spp. by comparative genome hybridization. For this purpose, we constructed and used a DNA-based microarray containing over 2000 randomly cloned DNA fragments from *B. breve* type strain LMG13208. This molecular analysis revealed a high degree of genomic variation between the isolated strains and allowed the vast majority to be grouped into 4 clusters. One cluster contained a single isolate that was virtually indistinguishable from the *B. breve* type strain. The 3 other clusters included 19 *B. breve* strains that differed considerably from all type strains. Remarkably, each of the 4 clusters included strains that were isolated from a single infant, indicating that a niche adaptation may contribute to variation within the *B. breve* species. Based on genomic hybridization data, the new *B. breve* isolates were estimated to contain approximately 60-90% of the genes of the *B. breve* type strain, attesting to the existence of various subspecies within the species *B. breve*. Further bioinformatic analysis identified several hundred diagnostic clones specific to the genomic clustering of the *B. breve* isolates. Molecular analysis of representatives of these revealed that annotated genes from the conserved *B. breve* core encoded mainly housekeeping functions, while the strain-specific genes were predicted to code for functions related to life style, such as carbohydrate metabolism and transport. This is compatible with genetic adaptation of the strains to their niche, a combination of infants and diet.

PMID: 21726634

Organism	GPID	status
Group 05 Pseudomonas		
<i>Pseudomonas aeruginosa</i> LESB58	31101	complete
<i>Pseudomonas aeruginosa</i> PA7	16720	complete
<i>Pseudomonas aeruginosa</i> PAO1	331	complete
<i>Pseudomonas brassicacearum</i> subsp. <i>brassicacearum</i> NFM421	63495	complete
<i>Pseudomonas entomophila</i> L48	16800	complete
<i>Pseudomonas fluorescens</i> Pf-5	327	complete
<i>Pseudomonas fluorescens</i> Pf0-1	12	complete
<i>Pseudomonas fulva</i> 12-X	49675	complete
<i>Pseudomonas mendocina</i> NK-01	64797	complete
<i>Pseudomonas mendocina</i> ymp	17457	complete
<i>Pseudomonas putida</i> F1	13909	complete
<i>Pseudomonas putida</i> GB-1	17629	complete
<i>Pseudomonas putida</i> KT2440	267	complete
<i>Pseudomonas putida</i> S16	67881	complete
<i>Pseudomonas putida</i> W619	17053	complete
<i>Pseudomonas stutzeri</i> A1501	16817	complete
<i>Pseudomonas stutzeri</i> ATCC 17588 = LMG 11199	68131	complete
<i>Pseudomonas syringae</i> pv. <i>phaseolicola</i> 1448A	12416	complete
<i>Pseudomonas syringae</i> pv. <i>syringae</i> B728a	323	complete
<i>Pseudomonas syringae</i> pv. <i>tomato</i> str. DC3000	359	complete

Microb Ecol. 2011 Aug 3. [Epub ahead of print]

Defining the *Pseudomonas* Genus: Where Do We Draw the Line with *Azotobacter*?

Ozen AI, Ussery DW.

Abstract

The genus *Pseudomonas* has gone through many taxonomic revisions over the past 100 years, going from a very large and diverse group of bacteria to a smaller, more refined and ordered list having specific properties. The relationship of the *Pseudomonas* genus to *Azotobacter vinelandii* is examined using three genomic sequence-based methods. First, using 16S rRNA trees, it is shown that *A. vinelandii* groups within the *Pseudomonas* close to *Pseudomonas aeruginosa*. Genomes from other related organisms (*Acinetobacter*, *Psychrobacter*, and *Cellvibrio*) are outside the *Pseudomonas* cluster. Second, pan genome family trees based on conserved gene families also show *A. vinelandii* to be more closely related to *Pseudomonas* than other related organisms. Third, exhaustive BLAST comparisons demonstrate that the fraction of shared genes between *A. vinelandii* and *Pseudomonas* genomes is similar to that of *Pseudomonas* species with each other. The results of these different methods point to a high similarity between *A. vinelandii* and the *Pseudomonas* genus, suggesting that *Azotobacter* might actually be a *Pseudomonas*.

PMID: 21811795

Organism	GPID	status
Group 06 Bacteroides		
Bacteroides fragilis 638R	50405	complete
Bacteroides fragilis NCTC 9343	46	complete
Bacteroides fragilis YCH46	13067	complete
Bacteroides helcogenes P 36-108	41913	complete
Bacteroides salanitronis DSM 18170	40055	complete
Bacteroides thetaiotaomicron VPI-5482	399	complete
Bacteroides vulgatus ATCC 8482	13378	complete
Bacteroides xylanisolvens XB1A	39177	complete
Bacteroides caccae ATCC 43185	18163	assembly
Bacteroides cellulosilyticus DSM 14838	30027	assembly
Bacteroides clarus YIT 12056	48509	assembly
Bacteroides coprocola DSM 17136	20521	assembly
Bacteroides dorei DSM 17855	27831	assembly
Bacteroides eggerthii 1_2_48FAA	40009	assembly
Bacteroides finegoldii DSM 17565	27823	assembly
Bacteroides fragilis 3_1_12	32433	assembly
Bacteroides intestinalis DSM 17393	20523	assembly
Bacteroides ovatus 3_8_47FAA	40011	assembly
Bacteroides pectinophilus ATCC 43243	27825	assembly
Bacteroides plebeius DSM 17135	27829	assembly

Microb Ecol. 2011 Apr;61(3):473-85.

A closer look at *Bacteroides*: phylogenetic relationship and genomic implications of a life in the human gut.

Karlsson FH, Ussery DW, Nielsen J, Nookaew I.

Abstract

The human gut is extremely densely inhabited by bacteria mainly from two phyla, *Bacteroidetes* and Firmicutes, and there is a great interest in analyzing whole-genome sequences for these species because of their relation to human health and disease. Here, we do whole-genome comparison of 105 *Bacteroidetes/Chlorobi* genomes to elucidate their phylogenetic relationship and to gain insight into what is separating the gut living *Bacteroides* and *Parabacteroides* genera from other *Bacteroidetes/Chlorobi* species. A comprehensive analysis shows that *Bacteroides* species have a higher number of extracytoplasmic function σ factors (ECF σ factors) and two component systems for extracellular signal transduction compared to other *Bacteroidetes/Chlorobi* species. A whole-genome phylogenetic analysis shows a very little difference between the *Parabacteroides* and *Bacteroides* genera. Further analysis shows that *Bacteroides* and *Parabacteroides* species share a large common core of 1,085 protein families. Genome atlases illustrate that there are few and only small unique areas on the chromosomes of four *Bacteroides/Parabacteroides* genomes. Functional classification to clusters of orthologous groups show that *Bacteroides* species are enriched in carbohydrate transport and metabolism proteins. Classification of proteins in KEGG metabolic pathways gives a detailed view of the genome's metabolic capabilities that can be linked to its habitat. *Bacteroides pectinophilus* and *Bacteroides capillosus* do not cluster together with other *Bacteroides* species, based on analysis of 16S rRNA sequence, whole-genome protein families and functional content, 16S rRNA sequences of the two species suggest that they belong to the Firmicutes phylum. We have presented a more detailed and precise description of the phylogenetic relationships of members of the *Bacteroidetes/Chlorobi* phylum by whole genome comparison. Gut living *Bacteroides* have an enriched set of glycan, vitamin, and cofactor enzymes important for diet digestion.

PMID: 2122211

Organism	GPID	status
Group 07 Brucella		
<i>Brucella abortus</i> S19	18999	complete
<i>Brucella abortus</i> bv. 1 str. 9-941	9619	complete
<i>Brucella canis</i> ATCC 23365	20243	complete
<i>Brucella melitensis</i> ATCC 23457	30561	complete
<i>Brucella melitensis</i> M28	61505	complete
<i>Brucella melitensis</i> M5-90	42891	complete
<i>Brucella melitensis</i> biovar Abortus 2308	16203	complete
<i>Brucella melitensis</i> bv. 1 str. 16M	180	complete
<i>Brucella microti</i> CCM 4915	32233	complete
<i>Brucella ovis</i> ATCC 25840	12514	complete
<i>Brucella pinnipedialis</i> B2/94	41867	complete
<i>Brucella suis</i> 1330	320	complete
<i>Brucella suis</i> ATCC 23445	20371	complete
<i>Brucella abortus</i> NCTC 8038	34743	assembly
<i>Brucella ceti</i> B1/94	33573	assembly
<i>Brucella ceti</i> M13/05/1	33043	assembly
<i>Brucella melitensis</i> bv. 1 str. 16M	34747	assembly
<i>Brucella neotomae</i> 5K33	33567	assembly
<i>Brucella pinnipedialis</i> B2/94	33039	assembly
<i>Brucella pinnipedialis</i> M163/99/10	33037	assembly

BMC Evol Biol. 2010 Aug 13;10:249.

Genomic comparisons of *Brucella* spp. and closely related bacteria using base compositional and proteome based methods.

Bohlin J, Snipen L, Cloeckert A, Lagesen K, Ussery D, Kristoffersen AB, Godfroid J.

BACKGROUND:

Classification of bacteria within the genus *Brucella* has been difficult due in part to considerable genomic homogeneity between the different species and biovars, in spite of clear differences in phenotypes. Therefore, many different methods have been used to assess *Brucella* taxonomy. In the current work, we examine 32 sequenced genomes from genus *Brucella* representing the six classical species, as well as more recently described species, using bioinformatical methods. Comparisons were made at the level of genomic DNA using oligonucleotide based methods (Markov chain based genomic signatures, genomic codon and amino acid frequencies based comparisons) and proteomes (all-against-all BLAST protein comparisons and pan-genomic analyses).

RESULTS:

We found that the oligonucleotide based methods gave different results compared to that of the proteome based methods. Differences were also found between the oligonucleotide based methods used. Whilst the Markov chain based genomic signatures grouped the different species in genus *Brucella* according to host preference, the codon and amino acid frequencies based methods reflected small differences between the *Brucella* species. Only minor differences could be detected between all genera included in this study using the codon and amino acid frequencies based methods. Proteome comparisons were found to be in strong accordance with current *Brucella* taxonomy indicating a remarkable association between gene gain or loss on one hand and mutations in marker genes on the other. The proteome based methods found greater similarity between *Brucella* species and *Ochrobactrum* species than between species within genus *Agrobacterium* compared to each other. In other words, proteome comparisons of species within genus *Agrobacterium* were found to be more diverse than proteome comparisons between species in genus *Brucella* and genus *Ochrobactrum*. Pan-genomic analyses indicated that uptake of DNA from outside genus *Brucella* appears to be limited.

CONCLUSIONS:

While both the proteome based methods and the Markov chain based genomic signatures were able to reflect environmental diversity between the different species and strains of genus *Brucella*, the genomic codon and amino acid frequencies based comparisons were not found adequate for such comparisons. The proteome comparison based phylogenies of the species in genus *Brucella* showed a surprising consistency with current *Brucella* taxonomy.

PMID: 20707916

Organism	GPID	status
Group 08 Campylobacter		
Campylobacter concisus 13826	17159	complete
Campylobacter curvus 525.92	17161	complete
Campylobacter fetus subsp. fetus 82-40	16293	complete
Campylobacter hominis ATCC BAA-381	20083	complete
Campylobacter jejuni RM1221	303	complete
Campylobacter jejuni subsp. doylei 269.97	17163	complete
Campylobacter jejuni subsp. jejuni 81-176	16135	complete
Campylobacter jejuni subsp. jejuni 81116	17953	complete
Campylobacter jejuni subsp. jejuni IA3902	28907	complete
Campylobacter jejuni subsp. jejuni ICDCCJ07001	47949	complete
Campylobacter jejuni subsp. jejuni M1	38041	complete
Campylobacter jejuni subsp. jejuni NCTC 11168	8	complete
Campylobacter jejuni subsp. jejuni S3	45947	complete
Campylobacter lari RM2100	12517	complete
Campylobacter coli JV20	50513	assembly
Campylobacter coli RM2228	12516	assembly
Campylobacter fetus subsp. venerealis NCTC 10354	62521	assembly
Campylobacter fetus subsp. venerealis str. Azul-94	7	assembly
Campylobacter jejuni subsp. jejuni 1336	33777	assembly
Campylobacter jejuni subsp. jejuni 260.94	16229	assembly

PLoS One. 2010 Aug 26;5(8):e12253.

Genomic characterization of *Campylobacter jejuni* strain M1.

Friis C, Wassenaar TM, Javed MA, Snipen L, Lagesen K, Hallin PF, Newell DG, Toszeghy M, Ridley A, Manning G, Ussery DW.

Abstract

Campylobacter jejuni strain M1 (laboratory designation 99/308) is a rarely documented case of direct transmission of *C. jejuni* from chicken to a person, resulting in enteritis. We have sequenced the genome of *C. jejuni* strain M1, and compared this to 12 other *C. jejuni* sequenced genomes currently publicly available. Compared to these, M1 is closest to strain 81116. Based on the 13 genome sequences, we have identified the *C. jejuni* pan-genome, as well as the core genome, the auxiliary genes, and genes unique between strains M1 and 81116. The pan-genome contains 2,427 gene families, whilst the core genome comprised 1,295 gene families, or about two-thirds of the gene content of the average of the sequenced *C. jejuni* genomes. Various comparison and visualization tools were applied to the 13 *C. jejuni* genome sequences, including a species pan- and core genome plot, a BLAST Matrix and a BLAST Atlas. Trees based on 16S rRNA sequences and on the total gene families in each genome are presented. The findings are discussed in the background of the proven virulence potential of M1.

PMID: 20865039

Organism	GPID	status
Group 09 E.coli		
Escherichia coli ABU 83972	38725	complete
Escherichia coli APEC O1	16718	complete
Escherichia coli ATCC 8739	18083	complete
Escherichia coli B str. REL606	18281	complete
Escherichia coli BW2952	33775	complete
Escherichia coli CFT073	313	complete
Escherichia coli E24377A	13960	complete
Escherichia coli HS	13959	complete
Escherichia coli O157:H7 str. EC4115	27739	complete
Escherichia coli O157:H7 str. EDL933	259	complete
Escherichia coli O157:H7 str. TW14359	30045	complete
Escherichia coli O26:H11 str. 11368	32509	complete
Escherichia coli O127:H6 str. E2348/69	32571	complete
Escherichia coli UTI89	16259	complete
Escherichia coli str. K-12 substr. DH10B	20079	complete
Escherichia coli str. K-12 substr. MG1655	225	complete
Escherichia coli SE11	18057	complete
Escherichia coli IAI1	33373	complete
Escherichia coli BL21(DE3)	28965	complete
Escherichia coli DH1	30031	complete

Microb Ecol. 2010 Nov;60(4):708-20.

Comparison of 61 sequenced *Escherichia coli* genomes.

Lukjancenko O, Wassenaar TM, Ussery DW.

Abstract

Escherichia coli is an important component of the biosphere and is an ideal model for studies of processes involved in bacterial genome evolution. Sixty-one publicly available *E. coli* and *Shigella* spp. sequenced genomes are compared, using basic methods to produce phylogenetic and proteomics trees, and to identify the pan- and core genomes of this set of sequenced strains. A hierarchical clustering of variable genes allowed clear separation of the strains into clusters, including known pathotypes; clinically relevant serotypes can also be resolved in this way. In contrast, when in silico MLST was performed, many of the various strains appear jumbled and less well resolved. The predicted pan-genome comprises 15,741 gene families, and only 993 (6%) of the families are represented in every genome, comprising the core genome. The variable or 'accessory' genes thus make up more than 90% of the pan-genome and about 80% of a typical genome; some of these variable genes tend to be co-localized on genomic islands. The diversity within the species *E. coli*, and the overlap in gene content between this and related species, suggests a continuum rather than sharp species borders in this group of *Enterobacteriaceae*.

PMID: 20623278

Organism	GPID	status
Group 10 Chlorobi		
<i>Chlorobium chlorochromatii</i> CaD3	13921	complete
<i>Chlorobium limicola</i> DSM 245	12606	complete
<i>Chlorobium luteolum</i> DSM 273	13012	complete
<i>Chlorobium phaeobacteroides</i> BS1	12608	complete
<i>Chlorobium phaeobacteroides</i> DSM 266	12609	complete
<i>Chlorobium phaeovibrioides</i> DSM 265	12607	complete
<i>Chlorobium tepidum</i> TLS	302	complete
<i>Chlorobaculum parvum</i> NCIB 8327	29213	complete
<i>Prosthecochloris aestuarii</i> DSM 271	12749	complete
<i>Chloroherpeton thalassium</i> ATCC 35110	29215	complete
<i>Chlorobium ferrooxidans</i> DSM 13031	16644	assembly

Photosynth Res. 2010 Jun;104(2-3):137-52.

Comparative genomics of green sulfur bacteria.

Davenport C, Ussery DW, Tümmler B.

Abstract

Eleven completely sequenced Chlorobi genomes were compared in oligonucleotide usage, gene contents, and synteny. The green sulfur bacteria (GSB) are equipped with a core genome that sustains their anoxygenic phototrophic lifestyle by photosynthesis, sulfur oxidation, and CO₂ fixation. Whole-genome gene family and single gene sequence comparisons yielded similar phylogenetic trees of the sequenced chromosomes indicating a concerted vertical evolution of large gene sets. Chromosomal synteny of genes is not preserved in the phylum Chlorobi. The accessory genome is characterized by anomalous oligonucleotide usage and endows the strains with individual features for transport, secretion, cell wall, extracellular constituents, and a few elements of the biosynthetic apparatus. Giant genes are a peculiar feature of the genera *Chlorobium* and *Prosthecochloris*. The predicted proteins have a huge molecular weight of 10(6), and are probably instrumental for the bacteria to generate their own intimate (micro)environment.

PMID: 20099081

Organism	GPID	status
Group 11 Vibrio		
Vibrio cholerae Ban5	39751	complete
Vibrio cholerae Ind4	39753	complete
Vibrio cholerae LMA3894-4	61113	complete
Vibrio cholerae M66-2	32851	complete
Vibrio cholerae MJ-1236	33555	complete
Vibrio cholerae Mex1	39757	complete
Vibrio cholerae O1 biovar El Tor str. N16961	36	complete
Vibrio cholerae O395	15667	complete
Vibrio fischeri ES114	12986	complete
Vibrio fischeri MJ11	19393	complete
Vibrio fluvialis Ind1	39759	complete
Vibrio furnissii NCTC 11218	53247	complete
Vibrio harveyi ATCC BAA-1116	19857	complete
Vibrio parahaemolyticus RIMD 2210633	360	complete
Vibrio sp. Ex25	40507	complete
Vibrio splendidus LGP32	32815	complete
Vibrio vulnificus CMCP6	349	complete
Vibrio vulnificus MO6-24/O	59881	complete
Vibrio vulnificus YJ016	1430	complete
Vibrio anguillarum 775	51883	complete

Microb Ecol. 2010 Jan;59(1):1-13.

On the origins of a *Vibrio* species.

Vesth T, Wassenaar TM, Hallin PF, Snipen L, Lagesen K, Ussery DW.

Abstract

Thirty-two genome sequences of various *Vibrionaceae* members are compared, with emphasis on what makes *V. cholerae* unique. As few as 1,000 gene families are conserved across all the *Vibrionaceae* genomes analysed; this fraction roughly doubles for gene families conserved within the species *V. cholerae*. Of these, approximately 200 gene families that cluster on various locations of the genome are not found in other sequenced *Vibrionaceae*; these are possibly unique to the *V. cholerae* species. By comparing gene family content of the analysed genomes, the relatedness to a particular species is identified for two unsequenced genomes. Conversely, two genomes presumably belonging to the same species have suspiciously dissimilar gene family content. We are able to identify a number of genes that are conserved in, and unique to, *V. cholerae*. Some of these genes may be crucial to the niche adaptation of this species.

PMID: 19830476

Organism	GPID	status
Group 12 Burkholderia		
Burkholderia ambifaria AMMD	13490	complete
Burkholderia cenocepacia AU 1054	13919	complete
Burkholderia cenocepacia HI2424	13918	complete
Burkholderia cenocepacia J2315	339	complete
Burkholderia cenocepacia MC0-3	17929	complete
Burkholderia mallei ATCC 23344	171	complete
Burkholderia mallei NCTC 10229	13943	complete
Burkholderia mallei NCTC 10247	13946	complete
Burkholderia mallei SAVP1	13947	complete
Burkholderia pseudomallei 1106a	16182	complete
Burkholderia pseudomallei 1710b	13954	complete
Burkholderia pseudomallei 668	13953	complete
Burkholderia pseudomallei K96243	178	complete
Burkholderia gladioli BSR3	64503	complete
Burkholderia glumae BGR1	33901	complete
Burkholderia multivorans ATCC 17616	17407	complete
Burkholderia phymatum STM815	17409	complete
Burkholderia phytofirmans PsJN	17463	complete
Burkholderia rhizoxinica HKI 454	51915	complete
Burkholderia thailandensis E264	10774	complete

Genome Dyn. 2009;6:140-57. Epub 2009 Aug 19.

The genus *Burkholderia*: analysis of 56 genomic sequences.

Ussery DW, Kiil K, Lagesen K, Sicheritz-Pontén T, Bohlin J, Wassenaar TM.

Abstract

The genus *Burkholderia* consists of a number of very diverse species, both in terms of lifestyle (which varies from category B pathogens to apathogenic soil bacteria and plant colonizers) and their genetic contents. We have used 56 publicly available genomes to explore the genomic diversity within this genus, including genome sequences that are not completely finished, but are available from the NCBI database. Defining the pan- and core genomes of species results in insights in the conserved and variable fraction of genomes, and can verify (or question) historic, taxonomic groupings. We find only several hundred genes that are conserved across all *Burkholderia* genomes, whilst there are more than 40,000 gene families in the *Burkholderia* pan-genome. A BLAST matrix visualizes the fraction of conserved genes in pairwise comparisons. A BLAST atlas shows which genes are actually conserved in a number of genomes, located and visualized with reference to a chosen genome. Genomic islands are common in many *Burkholderia* genomes, and most of these can be readily visualized by DNA structural properties of the chromosome. Trees that are based on relatedness of gene family content yield different results depending on what genes are analyzed. Some of the differences can be explained by errors in incomplete genome sequences, but, as our data illustrate, the outcome of phylogenetic trees depends on the type of genes that are analyzed.

PMID: 19696499

Organism	GPID	status
Group 13 Staphylococcus		
Staphylococcus aureus 04-02981	34809	complete
Staphylococcus aureus RF122	63	complete
Staphylococcus aureus subsp. aureus COL	238	complete
Staphylococcus aureus subsp. aureus ECT-R 2	52833	complete
Staphylococcus aureus subsp. aureus JH1	15758	complete
Staphylococcus aureus subsp. aureus JH9	15757	complete
Staphylococcus aureus subsp. aureus MRSA252	265	complete
Staphylococcus aureus subsp. aureus MW2	306	complete
Staphylococcus aureus subsp. aureus N315	264	complete
Staphylococcus aureus subsp. aureus NCTC 8325	237	complete
Staphylococcus aureus subsp. aureus ST398	29427	complete
Staphylococcus carnosus subsp. carnosus TM300	34811	complete
Staphylococcus epidermidis ATCC 12228	279	complete
Staphylococcus epidermidis RP62A	64	complete
Staphylococcus haemolyticus JCSC1435	12508	complete
Staphylococcus lugdunensis HKU09-01	42395	complete
Staphylococcus lugdunensis N920143	67127	complete
Staphylococcus pseudintermedius ED99	62991	complete
Staphylococcus pseudintermedius HKU10-03	61283	complete
Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305	15596	complete

FEMS Microbiol Lett. 2011 Sep;322(1):60-7. doi: 10.1111/j.1574-6968.2011.02339.x. Epub 2011 Jul 4.

Genome sequence of *Staphylococcus lugdunensis* N920143 allows identification of putative colonization and virulence factors.

Heilbronner S, Holden MT, van Tonder A, Geoghegan JA, Foster TJ, Parkhill J, Bentley SD.

Abstract

Staphylococcus lugdunensis is an opportunistic pathogen related to *Staphylococcus aureus* and *Staphylococcus epidermidis*. The genome sequence of *S. lugdunensis* strain N920143 has been compared with other *staphylococci*, and genes were identified that could promote survival of *S. lugdunensis* on human skin and pathogenesis of infections. *Staphylococcus lugdunensis* lacks virulence factors that characterize *S. aureus* and harbours a smaller number of genes encoding surface proteins. It is the only staphylococcal species other than *S. aureus* that possesses a locus encoding iron-regulated surface determinant (Isd) proteins involved in iron acquisition from haemoglobin.

PMID: 21682763

Organism	GPID	status
Group 14 Bacillus		
Bacillus anthracis str. Ames	309	complete
Bacillus anthracis str. CDC 684	31329	complete
Bacillus anthracis str. Sterne	10878	complete
Bacillus atrophaeus 1942	46075	complete
Bacillus cellulosilyticus DSM 2522	38423	complete
Bacillus cereus 03BB102	31307	complete
Bacillus cereus AH187	17715	complete
Bacillus cereus ATCC 10987	74	complete
Bacillus cereus B4264	17731	complete
Bacillus cereus G9842	17733	complete
Bacillus clausii KSM-K16	13291	complete
Bacillus halodurans C-125	235	complete
Bacillus licheniformis ATCC 14580	12388	complete
Bacillus megaterium DSM 319	42425	complete
Bacillus pseudofirmus OF4	28811	complete
Bacillus pumilus SAFR-032	20391	complete
Bacillus subtilis BSn5	61191	complete
Bacillus subtilis subsp. natto BEST195	38027	complete
Bacillus subtilis subsp. spizizenii str. W23	38713	complete
Bacillus thuringiensis BMB171	43631	complete

Genomics. 2011 Jul;98(1):26-39. doi: 10.1016/j.ygeno.2011.03.008.

Investigating the genome diversity of *B. cereus* and evolutionary aspects of *B. anthracis* emergence.

Papazisi L, Rasko DA, Ratnayake S, Bock GR, Remortel BG, Appalla L, Liu J, Dracheva T, Braisted JC, Shallom S, Jarrahi B, Snesrud E, Ahn S, Sun Q, Rilstone J, Okstad OA, Kolstø AB, Fleischmann RD, Peterson SN.

Abstract

Here we report the use of a multi-genome DNA microarray to investigate the genome diversity of *Bacillus cereus* group members and elucidate the events associated with the emergence of *Bacillus anthracis* the causative agent of anthrax—a lethal zoonotic disease. We initially performed directed genome sequencing of seven diverse *B. cereus* strains to identify novel sequences encoded in those genomes. The novel genes identified, combined with those publicly available, allowed the design of a "species" DNA microarray. Comparative genomic hybridization analyses of 41 strains indicate that substantial heterogeneity exists with respect to the genes comprising functional role categories. While the acquisition of the plasmid-encoded pathogenicity island (pXO1) and capsule genes (pXO2) represents a crucial landmark dictating the emergence of *B. anthracis*, the evolution of this species and its close relatives was associated with an overall shift in the fraction of genes devoted to energy metabolism, cellular processes, transport, as well as virulence.

PMID: 21447378