Applied and Environmental Microbiology, Apr. 2011, p. 2785–2787 Vol. 77, No. 8
0099-2240/11/$12.00 doi:10.1128/AEM.02493-10

Screening for Antimicrobial Resistance Genes and Virulence Factors via Genome Sequencing▽†

Mads Bennedsen, Birgitte Stuer-Lauridsen, Morten Danielsen, and Eric Johansen*

Innovation, Chr. Hansen A/S, 10-12 Bøge Allé, DK-2970 Hørsholm, Denmark

Received 22 October 2010/Accepted 8 February 2011

Second-generation genome sequencing and alignment of the resulting reads to *in silico* genomes containing antimicrobial resistance and virulence factor genes were used to screen for undesirable genes in 28 strains which could be used in human nutrition. No virulence factor genes were detected, while several isolates contained antimicrobial resistance genes.

A large variety of bacteria are intentionally added to the food supply. These include starter cultures for production of fermented foods and probiotics in food and dietary supplements. Since these bacteria are typically viable when consumed, considerable characterization is required to ensure the absence of undesirable properties (11, 13).

European Food Safety Authority (EFSA) guidelines specify the presence of transmissible antimicrobial resistance genes and virulence factors as undesirable (1, 3). More than 375 types of antimicrobial resistance genes, encoding resistance to nearly 250 antimicrobials, are described in the Antibiotic Resistance Genes Database (ARDB) (7). The Virulence Factor Database (VFDB) (15) contains the sequences of 2,353 genes, representing 408 virulence factors and 24 pathogenicity islands.

Construction of *in silico* genomes. Antimicrobial resistance gene sequences were downloaded from GenBank and imported into Genomic Workbench 3 (CLC bio, Aarhus, Denmark) to create an *in silico* genome containing >250 concatenated gene sequences (see Table S1 in the supplemental material). The primary sources were ARDB (7) and an online overview of tetracycline and macrolide-lincosamide-streptogramin B (MLS) resistance genes (http://faculty.washington.edu/marilynr/), supplemented with our knowledge of antibiotic resistance genes in species relevant to food, previously described genes in Gram-positive bacteria, and EFSA recommendations regarding particularly undesirable antimicrobial resistances (2).

All 2,353 DNA sequences from VFDB (15) were downloaded into Excel, imported into Genomic Workbench, and converted into an artificial genome of virulence factors (see Table S2 in the supplemental material).

RNA polymerase B subunit (*rpoB*) genes were included as positive controls. The nucleotide sequences of both *in silico* genomes are included in the supplemental material.

Bacterial strains are listed in Table 1; the prefix CHCC designates strains from the Chr. Hansen Culture Collection isolated from food and other natural sources. *Bifidobacterium animalis* subsp. *lactis* IPLAIC4 has been described previously (5). Taxonomic designations are based on 16S rRNA gene sequences.

Preparation of genomic DNA and sequencing. Bacteria were cultured overnight in M17 (Oxoid, Cambridge, United Kingdom) or MRS broth (Difco, BD, Franklin Lakes, NJ) and harvested by centrifugation (5 min at 4,000 × *g*). Total genomic DNA was purified using the DNeasy blood and tissue kit (Qiagen Inc., Valencia, CA) according to the manufacturer's instructions for Gram-positive bacteria, with one modification: purified DNA was eluted in water and concentrated to 200 ng/μl by vacuum drying. DNA was sequenced using Illumina GSII 38-bp single reads or 38-bp paired-end reads by Source BioScience (Nottingham, United Kingdom). Results were 2 to 15 million raw sequence reads in Fastq format, corresponding to 20- to 300-fold genome coverage. The sequences of *Escherichia coli* strains MG1655 and TW03542 were from the NCBI Sequence Read Archive (14).

Genome assembly and analysis. The *in silico* genomes were used as scaffolds to assemble the 38-bp sequence reads into contiguous sequences. Assembly occurs only when there is significant similarity, and a gene is detected only when many overlapping fragments can be assembled. When a gene is present, it is possible to assemble the complete gene with a depth of coverage comparable to that of *rpoB* from the same species. Variations in the efficiency of DNA sequencing and genome size result in different depths of coverage. Results are considered reliable if the depth of coverage is >20. *rpoB* is chromosomal in all species studied and was used to determine the relative copy number for any gene detected. Partial genes and chimeric genes can also be identified. When a gene is absent, no assembly to that part of the *in silico* genome occurs.

Sequence assembly was done as "Reference assembly" using Genomic Workbench 3, with up to 4 mismatches allowed per 38 bp. The specific parameters used are as follows: fast, ungapped reference assembly; mismatch cost, 2; and limit, 8. In the case of paired-end data, the parameters were as follows: minimum read distance, 180 bp, and maximum read distance, 600 bp.

Detection of antimicrobial resistance genes and virulence factors. Figure 1 indicates the detection of *tet*(W) in *B. animalis* subsp. *lactis* IPLAIC4 and *tet*(S) in *Lactococcus lactis* CHCC6005. *tet*(S) in CHCC6005 has a coverage 12-fold higher

* Corresponding author. Mailing address: Innovation, Chr. Hansen A/S, 10-12 Bøge Allé, DK-2970 Hørsholm, Denmark. Phone: (45) 45 74 74 74. Fax: (45) 45 74 89 91. E-mail: dkejo@chr-hansen.com.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 18 February 2011.

TABLE 1. List of strains analyzed and results of *in silico* analyses

Species or subspecies	Strain name	Antimicrobial resistance gene detected	No. of virulence factor genes detected
Lactobacillus delbrueckii subsp. *bulgaricus*	CHCC8942	None	None
Lactobacillus delbrueckii subsp. *bulgaricus*	CHCC769	None	None
Lactobacillus delbrueckii subsp. *bulgaricus*	CHCC5213	None	None
Streptococcus thermophilus	CHCC2136	None	None
Streptococcus thermophilus	CHCC2222	None	None
Streptococcus thermophilus	CHCC3047	None	None
Lactobacillus fermentum	CHCC10568	None	None
Lactobacillus johnsonii	CHCC5774	None	None
Lactobacillus paracasei	CHCC3136	None	None
Lactobacillus paracasei	CHCC2115	None	None
Lactobacillus paracasei	CHCC10665	None	None
Lactobacillus acidophilus	CHCC3777	None	None
Lactobacillus rhamnosus	CHCC5366	None	None
Lactobacillus rhamnosus	CHCC3402	None	None
Lactobacillus rhamnosus	CHCC5150	None	None
Lactobacillus reuteri	CHCC12039	None	None
Lactobacillus plantarum	CHCC2365	None	None
Lactobacillus plantarum	CHCC10668	None	None
Lactobacillus plantarum	CHCC10672	None	None
Bifidobacterium longum	CHCC2182	None	None
Bifidobacterium longum	CHCC8879	None	None
Bifidobacterium longum subsp. *infantis*	CHCC2228	None	None
Bifidobacterium animalis subsp. *lactis*	ATCC27536	*tet*(W)	None
Bifidobacterium animalis subsp. *lactis*	CHCC13471	*tet*(W)	None
Bifidobacterium animalis subsp. *lactis*	IPLAIC4	*tet*(W)	None
Lactococcus lactis	CHCC6005	*tet*(S)	None
Lactococcus lactis	CHCC2350	None	None
Lactococcus lactis	CHCC1182	None	None
Escherichia coli	MG1655	*ampC*	63
Escherichia coli	TW03542	*ampC*	12

than the corresponding *rpoB* gene and is inferred to be plasmid borne. All other antimicrobial resistance genes detected (Table 1) have the same coverage as the relevant *rpoB* gene and are inferred to be chromosomal.

Antimicrobial sensitivity testing was done using Etest strips (AB Biodisk, Solna, Sweden) as described previously (6). *B. animalis* subsp. *lactis* strains CHCC13471, ATCC 27536, and IPLAIC4 show low-level resistance to tetracycline, which has a MIC of 16 μg/ml for these strains. *L. lactis* CHCC6005 shows high-level resistance to tetracycline (MIC > 256 μg/ml). Gueimonde et al. (5) showed by molecular techniques that IPLAIC4 indeed contains *tet*(W). The presence of *tet*(S) on a plasmid in CHCC6005 was confirmed by Southern hybridization and DNA sequencing (12). None of the MICs of the antimicrobials tested (ampicillin, streptomycin, kanamycin, gentamicin, chloramphenicol, tetracycline, erythromycin, quinupristin-dalfopristin, vancomycin, trimethoprim, ciprofloxacin, linezolid, and rifampin) for the other strains were above the breakpoints described previously for the various species (4, 6, 10).

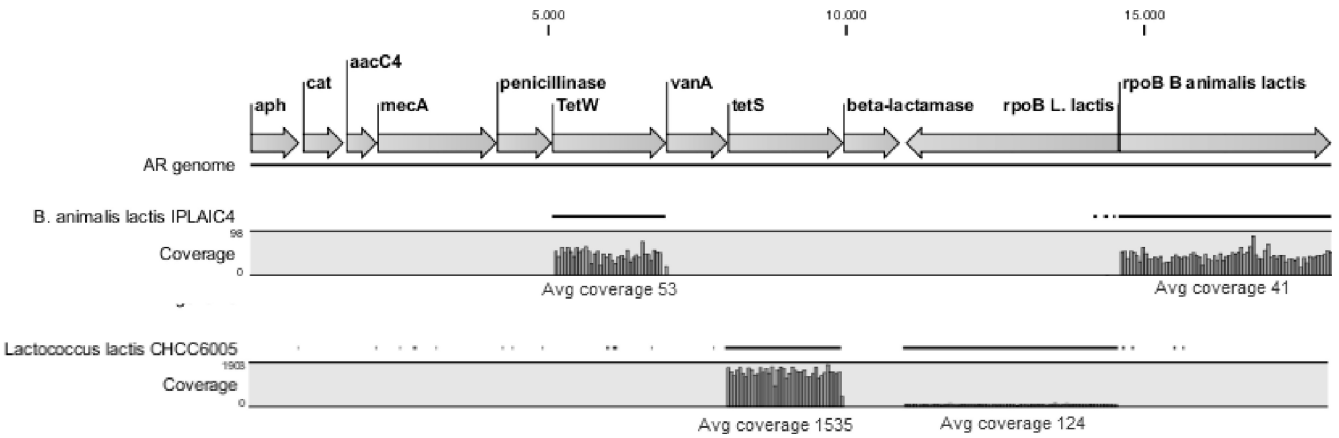

FIG. 1. Reference assembly of DNA sequence reads against the *in silico* antibiotic resistance (AR) genome. The vertical scales for the two strains differ. Only a subset of genes in the *in silico* genome is shown.

Only the two *E. coli* strains contain any of the virulence factor genes (Table 1).

Advantages of the screening method. In contrast to phenotypic methods, this screening method is independent of growth conditions. Detection of newly discovered genes does not require laboratory work; a simple *in silico* analysis is sufficient. False-positive results will not occur, as a complete gene cannot be assembled, at a depth of coverage similar to that for *rpoB*, if the gene is absent from the strain. False-negative results are also unlikely, especially for well-known genes, of which many variants are included in the *in silico* genome. This type of analysis can be used for the detection of any gene of interest.

Screening is done without gap filling, generation of a complete circular genome sequence, or annotation. While some sequences may be missing, it is unlikely that, with a depth of coverage of >20, any gene will be completely absent from the sequence data. Assembly of the sequence reads for *Lactobacillus johnsonii* CHCC5774 to the published genome sequence of *Lactobacillus johnsonii* NCC 533 (9) reveals many single-nucleotide differences but only four small deletions (<200 bp). Thus, use of raw sequencing reads is unlikely to lead to undesirable genes escaping detection.

Safety considerations. We have tested 28 strains for the presence of >250 antimicrobial resistance genes and >400 toxin and virulence factor genes. *L. lactis* CHCC6005 carries the *tet*(S) gene on a medium-copy-number plasmid, of which this strain should be cured before use. All three *B. animalis* subsp. *lactis* strains contain *tet*(W). This determinant is widespread in *B. animalis* (5); no naturally occurring *B. animalis* subsp. *lactis* strain lacking *tet*(W) has been described. Transfer of *tet*(W) from *B. animalis* subsp. *lactis* to other bacteria has never been demonstrated (5, 8); thus, *tet*(W) is not considered to be transmissible.

Conclusions. We show here that second-generation genome sequencing can be used to screen strains for unwanted genetic content and provide a conceptual framework for querying any collection of genes assembled into an *in silico* genome. This screening supports, but does not replace, normal safety assessment of new strains.

Masoumeh Taremi, Jannie Schnabl, and Karin Schlichter are thanked for technical assistance.

REFERENCES

1. **European Food Safety Authority.** 2007. Opinion of the Scientific Committee on a request from EFSA on the introduction of a Qualified Presumption of Safety (QPS) approach for assessment of selected microorganisms referred to EFSA. EFSA J. **587:**1–16. doi:10.2903/j.efsa.2007.587.
2. **European Food Safety Authority.** 2008. Technical guidance prepared by the Panel on Additives and Products or Substances Used in Animal Feed (FEEDAP) on the update of the criteria used in the assessment of bacterial resistance to antimicrobials of human or veterinary importance. EFSA J. **732:**1–15. doi:10.2903/j.efsa.2008.732.
3. **European Food Safety Authority Panel on Biological Hazards.** 2009. Scientific opinion on the maintenance of the list of QPS microorganisms intentionally added to food or feed (2009 update). EFSA J. **7:**1431–1523. doi: 10.2903/j.efsa.2009.1431.
4. **Florez, A., et al.** 2008. Resistance-susceptibility profiles of *Lactococcus lactis* and *Streptococcus thermophilus* strains to eight antibiotics and proposition of new cut-offs. Int. J. Probiotics Prebiotics **3:**249–256.
5. **Gueimonde, M., et al.** 2010. Genetic basis of tetracycline resistance in *Bifidobacterium animalis* subsp. *lactis.* Appl. Environ. Microbiol. **76:**3364–3369.
6. **Korhonen, J., et al.** 2008. Antimicrobial susceptibility and proposed microbiological cut-off values of lactobacilli by phenotypic determination. Int. J. Probiotics Prebiotics **3:**257–268.
7. **Liu, B., and M. Pop.** 2009. ARDB—Antibiotic Resistance Genes Database. Nucleic Acids Res. **37:**D443–D447.
8. **Masco, L., K. Van Hoorde, E. De Brandt, J. Swings, and G. Huys.** 2006. Antimicrobial susceptibility of *Bifidobacterium* strains from humans, animals and probiotic products. J. Antimicrob. Chemother. **58:**85–94.
9. **Pridmore, R., et al.** 2004. The genome sequence of the probiotic intestinal bacterium *Lactobacillus johnsonii* NCC 533. Proc. Natl. Acad. Sci. U. S. A. **101:**2512–2517.
10. **Saarela, M., et al.** 2008. Susceptibility of bifidobacteria originating from different sources to tetracycline, erythromycin, streptomycin and vancomycin. Int. J. Probiotics Prebiotics **3:**269–270.
11. **Sanders, M.** 2009. How do we know when something called "probiotic" is really a probiotic? A guideline for consumers and health care professionals. Functional Food Rev. **1:**3–12.
12. **Scanlon, S.** 2007. Characterisation of tetracycline resistance in *Lactococcus lactis* subsp. *lactis* CHCC6005 and *Streptococcus thermophilus* CHCC9087. M.S. thesis. University of Copenhagen, Copenhagen, Denmark.
13. **Vankerckhoven, V., et al.** 2008. Biosafety assessment of probiotics used for human consumption: recommendations from the EU-PROSAFE project. Trends Food Sci. Technol. **19:**102–114.
14. **Wheeler, D., et al.** 2008. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. **36:**D13–D21.
15. **Yang, J., L. Chen, L. Sun, J. Yu, and Q. Jin.** 2008. VFDB 2008 release: an enhanced web-based resource for comparative pathogenomics. Nucleic Acids Res. **36:**D539–D542.