Multilocus Sequence Typing Scheme That Provides Both Species and Strain Differentiation for the *Burkholderia cepacia* Complex

Adam Baldwin,1 Eshwar Mahenthiralingam,2 Kathleen M. Thickett,3 David Honeybourne,3 Martin C. J. Maiden,4 John R. Govan,5 David P. Speert,6 John J. LiPuma,7 Peter Vandamme,8 and Chris G. Dowson1**

Department of Biological Sciences, Warwick University, Coventry CV4 7AL, United Kingdom;2 Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3TL, United Kingdom;3 Department of Respiratory Medicine, Birmingham Heartlands Hospital, Birmingham, United Kingdom;3 Peter Medawar Building for Pathogen Research and Department of Zoology, University of Oxford, Oxford, United Kingdom;4 Department of Medical Microbiology, University of Edinburgh Medical School, Teviot Place, Edinburgh, Scotland, United Kingdom;5 Department of Pediatrics, University of British Columbia, Vancouver, British Columbia, Canada;6 Department of Pediatrics and Communicable Diseases, University of Michigan Medical School, 1150 W. Medical Center Drive, 8323 MRSB III, Box 0646, Ann Arbor, Michigan 48109-0646;7 and Laboratorium voor Microbiologie, Universiteit Gent, Ledeganckstraat 35, B-9000 Ghent, Belgium8

Received 19 April 2005/Returned for modification 25 May 2005/Accepted 21 June 2005

A single multilocus sequence typing (MLST) scheme was developed for precise characterization of the opportunistic pathogens of *Burkholderia cepacia* complex (BCC), a group composed of at least nine closely related species. Seven conserved housekeeping genes were selected after a comparison of five *Burkholderia* species, and a collection of strains was subjected to nucleotide sequence analysis using a nested PCR amplification approach for each gene. MLST differentiated all nine current BCC species and identified 114 sequence types within a collection of 119 strains. No differentiation was found between strains recovered from environmental or clinical sources. The improved resolution in strain identification offered by MLST was able to identify previously characterized epidemic strain lineages and also demonstrated the presence of four novel potential species groups within the complex. There was also evidence for recombination having an important role in the recent evolution of individual BCC species. This highly transferable, validated, MLST scheme provides a new means to assist in species identification as well as unambiguous strain discrimination of the BCC by a single approach. It is also the first MLST scheme designed at the outset to incorporate multiple species and should facilitate global epidemiological investigations of the BCC.

The *Burkholderia cepacia* complex (BCC) is a closely related group of gram-negative bacteria found in many niches of both natural and clinical environments. Their classification has undergone considerable taxonomic changes over the last two decades (6, 35), and the group is now known to encompass at least nine distinct species whose laboratory identification can often prove difficult. Members of the BCC are opportunistic pathogens, capable of causing disease in plants, invertebrates, animals, and humans (3, 8, 16, 29). They can be particularly devastating, highly virulent, cystic fibrosis (CF) pathogens (20) that are also able to cause nosocomial infections among other groups of debilitated patients (14, 18). Due to the high intrinsic resistance of the BCC to antibiotics and antimicrobial compounds, all of these infections can prove very difficult to treat and may be fatal (1). All nine species have been found to possess strains capable of causing colonization in CF patients (4, 6, 33). The genetic diversity of the BCC is such that multiple diagnostic tests are necessary for accurate characterization, and difficulties with strain identification mean that misidentification can easily occur, with possible major implications for patient care (25, 26). There is also a need for surveillance of epidemic strains when outbreaks occur, and stringent infection control measures already exist as an unfortunate necessity to protect vulnerable members of the community (30).

Various molecular typing methods are currently utilized for the discrimination of the BCC. Techniques using the single locus of the *recA* gene, such as restriction fragment length polymorphisms (RFLP), are transferable but offer limited resolution and are primarily applied as a means of identification at the BCC species level (21). Other techniques used to discriminate beyond the species level include multilocus restriction typing, pulsed-field gel electrophoresis (PFGE), random amplified polymorphic DNA (RAPD), and BOX-PCR (5); how these methods are applied is dependent upon the organisms being investigated and the questions being addressed by the study. The PCR-based techniques, such as BOX-PCR and RAPD, are highly discriminatory but not always easily transferable between different laboratories. PFGE is also not always a transferable technique, requiring some degree of specialty both in equipment and in use. Multilocus restriction typing offers superior strain discrimination over single-locus RFLP by analyzing multiple genes, but these pattern-matching techniques based on gel banding have inherent variability and ambiguities (5).

A relatively new technique that is fast becoming the “gold

---

*Corresponding author. Mailing address: Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom. Phone: 44 (0)2476 523534. Fax: 44 (0)2476 523568. E-mail: c.g.dowson@warwick.ac.uk.*
standard” of bacterial typing methods is multilocus sequence typing (MLST) (24). It has been successfully applied to many clinically problematic species, several of which are prominent as respiratory pathogens, such as Streptococcus pneumoniae (11), Pseudomonas aeruginosa (10), and Haemophilus influenzae (27), and even to the highly pathogenic Burkholderia pseudomallei and Burkholderia mallei species, which are closely related to the BCC (12). Previously MLST schemes have been optimized to type a single species, whereas here we report the development of a single robust MLST scheme and database that encompasses all of the nine known BCC species, enabling improved identification of this complex group at both the species and strain levels within a single approach.

### MATERIALS AND METHODS

#### Bacterial strains.

BCC strains were obtained from the Belgium Co-ordinated Collection of Micro-organisms LMG Bacteria collection, Cardiff University collection (23), the U.S. B. cepacia Research Laboratory and Repository (19), and representatives of the published strain panels (7, 23). They covered a time period of the last 16 years from different continents. Culture and genomic DNA extraction were performed as described previously (22, 23). In addition, all isolates were genetically typed by either RAPD (22) or PFGE (21) prior to inclusion in the study to avoid unnecessary duplication of isolates of the same genotype. Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich Company Ltd., Poole, Dorset, United Kingdom.

#### Gene locus amplification.

Amplification primers were designed using available genome sequence data for the three BCC strains (Burkholderia cenocepacia strain J2315 [http://www.sanger.ac.uk/Projects/B_cenocepacia/], B. cepacia strain ATCC 17760 [http://www.tigr.org/msc/blasthome.html], and Burkholderia vietnamiensis strain G4 [http://www.tigr.org/mtc/blasthome.html]) (28). The genes selected for MLST were atpD, gfb, gyrB, recA, lepA, and phaC, as shown in Table 1. The genes gfb and lepA have been previously utilized for the MLST scheme developed for B. pseudomallei and B. mallei (29), though we have used a different portion of each gene.

For each locus, primers were designed to have a similar melting temperature (Tm) and were found to be successfully amplified by PCR over a wide range of annealing temperature conditions (50 to 60°C) for a diverse panel of BCC strains. Reaction conditions for all the primers were as follows: initial denaturation at 94°C for 2 min; 30 cycles of denaturation at 94°C for 1 min, primer annealing at 58°C for 1 min, and extension at 72°C for 2 min; followed by a final extension step of 72°C for 5 min. Each 50-μL amplification reaction mixture comprised ~10 ng chromosomal DNA, 20 pmol forward and reverse primer, and 1× PCR buffer (Qiagen, Crawley, United Kingdom) containing 1.5 mM MgCl₂, 0.8 mM deoxynucleotide triphosphates, and 1.25 U Taq (Qiagen). The amplification product was then purified using MinElute UF plates (Qiagen) following the manufacturer's protocol before being used in a sequencing reaction.

#### Multilocus sequence typing.

Internal nested primers were designed for sequencing in the same manner as the amplification primers (Table 1). Using these primers, nucleotide sequences were determined at least once on each DNA strand with the BigDye Terminator Ready reaction mix, version 3.1 (PE Systems, Foster City, Calif.) under standard sequencing conditions according to the manufacturer's protocol. Unincorporated dye terminators were removed by precipitation with 95% alcohol. The reaction products were separated and detected on an ABI PRISM genetic analyzer 3100 (PE Biosystems) using a standard sequencing module with a performance-optimized polymer and 5-cm array. The sequences from both strands of a given locus of the same isolate were aligned, trimmed to the desired length (Table 2), and edited using SeqMan II (DNAStar software).

#### Phylogenetic analysis.

To construct gene trees of the concatenated sequences (2,773 bp) for each isolate, the Jukes-Cantor neighbor-joining method was used (MEGA version 3; http://www.megasoftware.net). The significance of branching within the trees was evaluated by bootstrap analysis of 1,000 computer-generated trees. To calculate the index of association for the different BCC species, the LIAN program (version 3.1) (http://adenine.biz.fh-weihenstephan.de/lian/) was used. The software program START (http://www.mlst.net) (15) was used for all analyses unless otherwise stated.

### RESULTS

Selection of gene loci and chromosomal mapping. Several criteria were used in the selection of all potential loci. Genes included were those encoding putative housekeeping products necessary for biological roles in DNA repair, replication, and amino acid biosynthesis. Genes that were either located near or implicated as being putative virulence factors and mobile elements were avoided, since these may come under greater selective evolutionary pressures than other genes. The selected loci were distributed as much as possible across the chromosome to ensure that each locus was genetically unlinked. Each gene was also required to be approximately 500 bp in length to facilitate the design of universal nested primers for each locus, preferably in conserved flanking regions around a variable central core.

Development of the MLST scheme for the BCC. Of the loci chosen for the MLST scheme, atpD, gfb, gyrB, recA, lepA, and phaC were located on the largest BCC chromosome, which

---

**Table 1. Oligonucleotide nested primer sequences for the amplification and sequencing of the seven loci from genes in the BCC**

| Gene (gene label) | Putative gene product | Chromosome location (bp) | Gene size (bp) | Locus primer (5’→3’*) |
|------------------|-----------------------|-------------------------|----------------|----------------------|
| atpD (BCAL0036)  | ATP synthase β chain | Chr.1: 38673–400867     | 1,395          | GATCGTACAGTGCACTGG |
|                  |                       |                         |                | CTGCCGACCATCTAGTACC |
|                  |                       |                         |                | CGTCGAGCTCAAGAGCTC |
|                  |                       |                         |                | GGGACACCTTCTGACAGGA |
|                  |                       |                         |                | CAGACACCTGACGCAAGA |
|                  |                       |                         |                | GAGACACGGCTCTGAGTC |
|                  |                       |                         |                | GATAGCAAGAAGGCT GCC |
|                  |                       |                         |                | ACTCTCTTGCTCCATGCCC |
|                  |                       |                         |                | GCAACGCTAAAGTCATAA |
|                  |                       |                         |                | GCCGAGCCAAGTGACGG |
|                  |                       |                         |                | CAGTCATCGAAGACGTGA |
|                  |                       |                         |                | CGGATCGAAGAAGGGCTC |
|                  |                       |                         |                | AGGCGCCGAGAAAGGAC |
|                  |                       |                         |                | ATGGGCGGCTCTGAGGGT |
| gfb (BCAL0289)   | Glutamate synthase subunit | Chr.1: 31771–322474   | 4,704          | GCCGACGCTCAAGAGCTC |
|                  |                       |                         |                | GGGACACCTTCTGACAGGA |
|                  |                       |                         |                | CAGACACCTGACGCAAGA |
|                  |                       |                         |                | GAGACACGGCTCTGAGTC |
|                  |                       |                         |                | GATAGCAAGAAGGCT GCC |
|                  |                       |                         |                | ACTCTCTTGCTCCATGCCC |
|                  |                       |                         |                | GCAACGCTAAAGTCATAA |
|                  |                       |                         |                | GCCGAGCCAAGTGACGG |
|                  |                       |                         |                | CAGTCATCGAAGACGTGA |
|                  |                       |                         |                | CGGATCGAAGAAGGGCTC |
| gyrB (BCAL0421)  | DNA gyrase B         | Chr.1: 463355–465829    | 2,475          | CGACAACTCGATCGACGA |
|                  |                       |                         |                | ATCGTGATGCGGACGCTG |
|                  |                       |                         |                | GGTGTTAGCTGGCTGCC |
|                  |                       |                         |                | TAGATGCAAGAAGGCT GCC |
|                  |                       |                         |                | ACTCTCTTGCTCCATGCCC |
|                  |                       |                         |                | GCAACGCTAAAGTCATAA |
|                  |                       |                         |                | GCCGAGCCAAGTGACGG |
|                  |                       |                         |                | CAGTCATCGAAGACGTGA |
|                  |                       |                         |                | CGGATCGAAGAAGGGCTC |
| recA (BCAL0953)  | Recombinase A        | Chr.1: 1089631–1091424  | 1,794          | GAGACGGCAAGGTCTACA |
|                  |                       |                         |                | GGCATCAAGAAGAAGGCA |
|                  |                       |                         |                | GGCATCAAGAAGAAGGCA |
|                  |                       |                         |                | GGCATCAAGAAGAAGGCA |

*With gene number and location of gene within the genome of the B. cenocepacia strain J2315. Chr.1, largest chromosome; Chr.2, second-largest chromosome in the J2315 strain genome.
  a BCR1.
  b BCR2.
  c BCR2 (21).
TABLE 2. Analysis of the seven MLST loci in the BCC strains sampled

| Gene | Size (bp) of fragment analyzed | No. of alleles | No. of polymorphic sites | Proportion of polymorphic sites (%) | Mean G+C content (%) | \(d_{\text{SN}}/d_{\text{S}}\) |
|------|--------------------------------|---------------|--------------------------|-------------------------------------|----------------------|-------------------|
| atpD | 443                            | 70            | 58                       | 13.1                               | 62.2                 | 0.109             |
| gIB  | 400                            | 88            | 101                      | 25.3                               | 67.6                 | 0.098             |
| gyrB | 454                            | 92            | 170                      | 37.4                               | 62.6                 | 0.126             |
| recA | 393                            | 78            | 105                      | 26.7                               | 67.9                 | 0.049             |
| lepA | 397                            | 79            | 130                      | 32.7                               | 65.4                 | 0.118             |
| phaC | 385                            | 71            | 84                       | 21.8                               | 60.9                 | 0.040             |
| trpB | 301                            | 79            | 84                       | 27.9                               | 69.5                 | 0.068             |
| Mean | 396                            | 79.6          | 105                      | 26.5                               | 65.2                 | 0.087             |

appears to contain the majority of housekeeping genes. The remaining seventh locus (trpB) was chosen from the second-largest chromosome to ensure that the MLST scheme encompassed some of the diversity within other chromosomes of the multireplicon BCC. The chromosomal locations of all these loci were confirmed by bioinformatics analysis of the J2315 genome sequence (NC_004503 [http://www.sanger.ac.uk/Projects/B_cenocepacia/]). The putative gene products, gene sizes, and location within the J2315 genome are shown in Table 1.

Allelic variation. Since MLST uses multiple loci in its analysis, a greater degree of variation and therefore better resolution for typing BCC members and for inferring evolutionary and epidemiological relatedness can be obtained than with a single locus alone.

Novel sequence information for all seven loci was obtained from a collection of 119 BCC strains. To assess the performance of the MLST scheme, BCC strains were selected to be representative of the species and genetic diversity of the complex (evaluated in previous molecular epidemiological studies [7, 19, 23]) and were also distributed both temporally and geographically in terms of their isolation. The collection also comprised 74 isolates of clinical origin and 45 isolates recovered from environmental sources (Table 3). In silico sequence data were also obtained for all the loci from B. pseudomallei strain K96243 (NC_002930 [http://www.sanger.ac.uk/Projects/B_mallei]) [28], ”Burkholderia” strain SAR-1, a metagenome from the Sargasso Sea (NS_000028 [http://www.ncbi.nlm.nih.gov/genomes/static/es.html] [34a]), and Burkholderia xenovorans, strain LB400 (NZ_AAAT0000000 [http://genome.jgi-psf.org/finished_microbes/bfur/bfuru.home.html]). The latter strain sequence data were used to root the data set.

The mean allele length was 396 bp for the scheme and ranged between 301 bp (trpB) and 454 bp (gyrB) (Table 2). All alleles within a particular locus were found to be of identical length for all BCC strains and the non-BCC Burkholderia species examined, with the only exception being B. xenovorans strain LB400 at the atpD locus, where an in-frame deletion of 24 bp was detected. Nucleotide sequence diversity was found to be extensive at all seven loci, as shown in Table 2. The proportion of variable sites varied from 13.1% (atpD) to 37.4% (gyrB), which extended over the whole section of the sequenced allele. The polymorphic sites within the phaC locus are shown as an example of the allelic diversity observed (Fig. 1).

Allele variation is not necessarily equally likely at every nucleotide of each locus. If a locus does not have a role affected by selective pressure (such as antibiotic exposure), then nucleotide substitutions would frequently not be expected to change the amino acid sequence (synonymous), since changes are likely to be eliminated by purifying selection. By calculating the \(d_{\text{SN}}/d_{\text{S}}\) ratio (nonsynonymous substitutions to synonymous substitutions), the degree of selection operating on each locus can be estimated. The \(d_{\text{SN}}/d_{\text{S}}\) ratio for all seven loci within BCC strains was found to be significantly less than 1 (Table 2), indicating that no strong positive selective pressure was present at any of the loci selected, validating their suitability for inclusion in the BCC MLST scheme.

Assignment of allele and sequence types (ST). Each distinct sequence (allele sequence) at a particular locus was assigned a unique arbitrary number (allele type). The numbers of different alleles resolved from this BCC MLST scheme at each locus ranged from 70 (atpD) to 92 (gyrB). The mean number of allele types per locus was found to be 79.6, providing the potential for distinguishing \(>2.0 \times 10^{13}\) different genotypes within the BCC and also making it highly unlikely that identical STs would be obtained by chance.

After sequencing and assignment of allele types to all seven loci, each strain was then designated by a combination of seven numbers, called an allelic profile, in the order atpD, gIB, gyrB, recA, lepA, phaC, and trpB, which represented an ST for that particular strain (e.g., ST-1). Subsequent isolates with an identical allelic profile were assigned the same ST identifier and considered to be isogenic as they were indistinguishable at all seven loci.

For the development of the MLST scheme, it was important to use a diverse collection of strains to obtain primarily different STs (based upon known recA sequence or recA-RFLP profiles) to validate the scheme’s effectiveness across the whole of the BCC. A total of 114 STs were found for the 119 strains examined (Table 3); 114 were present only once, with ST-104 occurring twice and ST-28 occurring 5 times for 5 strains of the ET12 lineage (J2315, P1-1, LMG13307, LMG13316, and LMG13327).
TABLE 3. MLST analysis of the BCC strains showing their sources, geographic locations, and species

| Strain group or ST | Strain | Source | recA RFLP Location | Allelic profile |
|-------------------|--------|--------|--------------------|----------------|
| **B. cepacia, genomovar I** | **ATCC 17759** | ENV | E Trinidad | 1 1 1 1 2 1 1 |
| 1 | **LMG 17997** | NON | E Sweden | 2 2 2 2 1 2 2 |
| 2 | **BCC0464** | CF | E Italy | 2 54 43 37 38 62 48 |
| 3 | **BCC0116** | CF | E USA | 6 59 50 76 44 43 51 |
| 4 | **LMG 18821** | CF | E Australia | 4 3 40 3 3 3 53 |
| 5 | **AU0113** | CF | E USA | 109 49 3 | 3 40 53 |
| 6 | **BCC0240** | NON | E USA | 75 58 48 37 3 3 21 |
| 7 | **BCC0412** | ENV | E Italy | 6 59 50 76 44 43 51 |
| 8 | **BCC0227** | NON | E Canada | 91 93 96 103 42 1 21 |
| 9 | **ATCC 25416** | ENV | D USA | 5 4 44 4 4 4 48 |
| 10 | **ATCC 49709** | ENV | D USA | 24 94 125 53 41 94 102 |
| 11 | **BCC0218** | CF | D Australia | 72 53 46 73 34 38 51 |
| 12 | **BCC0394** | NON | D Japan | 78 55 45 37 40 37 50 |
| 13 | **IST431** | CF | AG Portugal | 6 52 3 5 5 5 3 |

**B. multivorans, genomovar II**

| Strain | Source | recA RFLP Location | Allelic profile |
|--------|--------|--------------------|----------------|
| 15 | **LMG 18825** | CF | F UK | 8 5 5 7 7 42 5 |
| 16 | **BCC0300** | CF | F France | 8 5 5 7 7 42 105 |
| 17 | **BCC0149** | CF | F USA | 97 50 4 79 37 63 55 |
| 18 | **C1662** | NON | F UK | 9 75 54 93 63 35 66 |
| 19 | **CS593** | CF | F Canada | 12 11 5 10 7 100 6 |
| 20 | **BCC0321** | ENVH | F UK | 12 50 52 78 37 35 54 |
| 21 | **ATCC 17616** | ENV | F USA | 13 78 100 94 92 96 6 |
| 22 | **BCC0317** | ENV | F Canada | 13 63 53 80 61 96 56 |
| 23 | **BCC0281** | CF | R USA | 7 111 4 6 6 12 4 |
| 24 | **BCC0866** | CF | R Canada | 80 61 97 11 64 96 104 |
| 25 | **AU0066** | CF | O USA | 10 60 4 77 37 35 5 |
| 26 | **BCC0497** | CF | O UK | 13 9 83 12 7 42 7 |
| 27 | **C1576** | CF | C UK | 13 7 6 10 8 42 6 |

**B. cenocepacia, genomovar III-A**

| Strain | Source | recA RFLP Location | Allelic profile |
|--------|--------|--------------------|----------------|
| 28 | **J2315** | CF | G UK | 15 11 9 14 11 6 12 |
| 29 | **LMG 13316** | CF | G UK | 15 11 9 14 11 6 12 |
| 30 | **LMG 13307** | CF | G UK | 15 11 9 14 11 6 12 |
| 31 | **LMG 13327** | CF | G UK | 15 11 9 14 11 6 12 |
| 32 | **BCC0711** | CF | G UK | 15 11 9 14 11 6 12 |
| 33 | **C5424** | CF | G Canada | 15 11 9 14 11 6 12 |
| 34 | **K50-2** | CF | G Canada | 21 11 13 14 11 6 12 |
| 35 | **BC-1** | CF | G Canada | 15 11 9 14 11 6 12 |
| 36 | **BC7** | CF | G Canada | 15 11 9 14 11 6 79 |
| 37 | **POPR8** | ENV | G Mexico | 16 11 10 14 11 6 79 |
| 38 | **BCC0560** | CF | G Canada | 16 11 10 95 11 6 79 |

**B. cenocepacia, genomovar III-B**

| Strain | Source | recA RFLP Location | Allelic profile |
|--------|--------|--------------------|----------------|
| 39 | **J415** | CF | H UK | 17 107 119 15 93 6 13 |
| 40 | **C1394** | CF | H UK | 17 13 12 17 66 6 11 |
| 41 | **ATCC 17765** | NON | H UK | 23 65 86 20 69 8 14 |
| 42 | **M36** | ENV | H USA | 17 65 57 15 69 8 14 |
| 43 | **CEP0511** | CF | I Australia | 16 108 121 49 94 41 9 |
| 44 | **PC184** | CF | J' USA | 17 15 85 19 68 41 13 |
| 45 | **BCC0491** | CF | J Canada | 17 64 56 39 79 41 13 |
| 46 | **AU0787** | CF | AQ USA | 17 97 104 58 80 76 60 |
| 47 | **IST452** | CF | AN Portugal | 67 98 59 68 47 6 19 |

**B. cenocepacia, genomovar III-C**

| Strain | Source | recA RFLP Location | Allelic profile |
|--------|--------|--------------------|----------------|
| 48 | **LMG 19230** | ENV | H2 France | 65 49 41 47 33 36 44 |
| 49 | **LMG 19238** | ENV | H2 Australia | 62 112 92 47 99 36 87 |

**B. cenocepacia, genomovar III-D**

| Strain | Source | recA RFLP Location | Allelic profile |
|--------|--------|--------------------|----------------|
| 50 | **BCC0458** | CF | U Italy | 55 39 32 39 24 30 38 |

**B. cenocepacia, genomovar III-E**

| Strain | Source | recA RFLP Location | Allelic profile |
|--------|--------|--------------------|----------------|
| 51 | **BCC0276** | ENV | V USA | 53 147 33 38 23 96 46 |
| 52 | **BCC0610** | ENV | V USA | 87 38 73 38 56 8 47 |
| 53 | **BCC0517** | ENV | V USA | 53 40 33 38 23 30 46 |

**B. stabilis, genomovar IV**

| Strain | Source | recA RFLP Location | Allelic profile |
|--------|--------|--------------------|----------------|
| 54 | **LMG 14294** | CF | J Belgium | 26 18 14 21 70 10 16 |
| 55 | **ATCC 35254** | ENVH | J USA | 26 18 42 21 70 10 16 |
| 56 | **BCC0248** | CF | J New Zealand | 25 18 42 21 70 10 16 |
| 57 | **BCC0717** | CF | J UK | 25 69 61 109 70 10 62 |
| 58 | **ATCC 27515** | NON | J UK | 25 68 60 109 70 43 61 |
| 59 | **BCC0418** | ENV | J Italy | 25 70 42 109 70 43 62 |

Continued on facing page
### TABLE 3—Continued

| Strain group or ST | Strain | Source | recA | RFLP Location | Allelic profile atpD, gfpB, gyrB, recA, lepA, phaC, tpeB |
|-------------------|--------|--------|------|---------------|---------------------------------------------------|
| **B. vietnamiensis, genomovar V** | | | | | |
| 56 AU0109 | CF | A | USA | 27 100 16 22 35 44 63 | |
| 57 FC0441* | NON | A | Canada | 29 19 17 22 12 11 80 | |
| 58 PC259* | CF | A | USA | 28 19 16 23 35 11 17 | |
| 59 BCC0136 | CF | A | Canada | 27 99 15 96 36 56 17 | |
| 60 G4 | ENV | A | USA | 27 20 16 48 12 11 17 | |
| 61 BCC0042 | ENVH | AK | USA | 27 20 16 23 35 56 17 | |
| 62 BCC0268 | ENV | A | New Zealand | 27 101 22 36 11 17 | |
| 63 CRE-7 | ENV | A | USA | 27 19 15 22 36 11 17 | |
| 64 BCC0104 | NON | A | Canada | 27 19 15 22 48 11 17 | |
| 65 LMG 19290* | ENV | B | Vietnam | 27 19 23 22 36 11 17 | |
| 66 BCC0128 | CF | B | Canada | 27 19 107 22 48 56 17 | |
| 67 BCC0581 | CF | B | Canada | 27 103 16 23 49 56 63 | |
| 68 BCC0151 | CF | AK | USA | 27 102 15 23 35 11 63 | |
| 69 BCC0124 | CF | AK | USA | 27 19 87 23 12 56 81 | |
| **B. dolosa, genomovar VI** | | | | | |
| 70 AU0746* | CF | Q | USA | 31 22 19 25 71 13 22 | |
| 71 LMG 19468 | CF | Q | USA | 30 21 127 24 72 13 20 | |
| 72 LMG 18945* | CF | Q | USA | 30 21 18 24 72 13 20 | |
| **B. ambifaria, genomovar VII** | | | | | |
| 73 M54* | ENV | L | USA | 32 23 20 26 13 57 23 | |
| 74 ATCC 53266* | ENV | L | USA | 38 25 23 28 16 16 25 | |
| 75 ATCC 53267 | ENV | L | USA | 38 25 23 28 16 16 25 | |
| 76 BCC0118 | CF | N | USA | 35 25 123 98 82 79 49 | |
| 77 AMMD* | ENV | N | USA | 35 25 123 98 103 59 49 | |
| 78 LMG 19467* | CF | N | USA | 39 29 24 29 17 17 26 | |
| 79 Ra3* | ENV | N | USA | 33 86 21 50 14 58 82 | |
| 80 AU1366 | CF | N | USA | 36 27 22 27 15 15 24 | |
| 81 BCC0250* | CF | N | USA | 36 26 89 99 104 90 99 | |
| 82 BCC0410 | ENV | N | USA | 36 72 64 84 62 66 49 | |
| 83 AU1293* | CF | N | USA | 36 27 22 27 15 15 24 | |
| 84 BCC0399 | ENV | N | Italy | 37 26 90 51 15 48 100 | |
| **B. anthina, genomovar VIII** | | | | | |
| 85 LMG 20983* | CF | T | UK | 40 30 25 30 18 18 27 | |
| 86 LMG 20987* | ENV | T | USA | 41 31 26 31 19 19 28 | |
| 87 LMG 20982* | ENVH | T | USA | 42 32 91 32 73 50 29 | |
| 88 R-11761 | ENV | AS | USA | 43 33 27 104 21 21 106 | |
| 89 LMG 16670* | ENV | AS | USA | 43 33 27 33 20 21 30 | |
| 90 AU1293* | CF | AS | USA | 90 33 27 33 21 21 30 | |
| 91 BI1 | ENV | AH | USA | 44 34 124 52 95 91 83 | |
| **B. pyrrocinia, genomovar IX** | | | | | |
| 92 LMG 21823* | ENV | AR | USA | 51 90 94 36 77 26 35 | |
| 93 AU2419* | CF | Sel3 | USA | 50 37 93 35 76 25 34 | |
| 94 R-13543 | ENV | P | USA | 49 89 30 91 107 24 33 | |
| 95 ATCC 39277* | ENV | P | USA | 46 36 116 107 97 92 85 | |
| **BCC group K** | | | | | |
| 96 IST410 | CF | K | Portugal | 89 82 80 71 60 73 74 | |
| 97 CEP0964 | CF | K | Australia | 89 83 81 71 39 54 75 | |
| 98 BI | ENV | K | USA | 89 84 82 65 45 55 77 | |
| 99 ATCC 17460 | ENV | K | Trinidad | 18 85 47 90 90 74 95 | |
| 100 CEP1056 | CF | K | Canada | 89 114 114 66 91 98 96 | |
| 101 ATCC 17760 | ENV | K | Trinidad | 63 46 38 44 30 33 42 | |
| 102 SAR-1 | ENV | | Sargasso Sea | 64 80 76 89 105 97 70 | |
| 103 BCC0335 | CF | K | Canada | 68 51 78 65 58 71 71 | |
| **B. cepacia complex 1** | | | | | |
| 104 R-11767 | CF | W | UK | 96 118 130 112 106 99 110 | |
| 105 R-11768 | CF | W | UK | 96 118 130 112 106 99 110 | |
| **B. cepacia complex 2** | | | | | |
| 106 BCC0110 | CF | H2 | Canada | 57 41 77 40 88 31 40 | |
| 107 BCC0329 | NON | H2 | Canada | 57 41 51 40 26 31 40 | |
| 108 R-9912 | CF | H2 | Canada | 92 81 39 40 89 83 41 | |
| **B. cepacia complex 3** | | | | | |
| 109 LMG 14939 | CF | J2 | Belgium | 60 43 36 42 27 72 72 | |
| 110 BCC0049 | NON | J2 | Germany | 60 43 36 42 27 61 72 | |
| 111 J2543 | ENV | J2 | UK | 88 113 112 62 81 84 108 | |
| **B. cepacia complex 4** | | | | | |
| 112 T21 | ENV | AA | USA | 48 28 29 54 22 23 32 | |
| 113 LMG 21824 | CF | AA | USA | 52 91 31 55 76 27 36 | |
| 114 BCC0124 | ENV | AU | USA | 86 77 52 72 55 69 31 | |

*a* An asterisk indicates the isolate is a panel strain. BCC group K, BCC group awaiting species designation (35); B. cepacia complex n, unidentified BCC groups; CF, isolated from a CF patient; NON, isolated from a non-CF patient; ENV, isolated from the environment; ENVH, isolated from a hospital environment; USA, United States; UK, United Kingdom.
Analysis of recombination among the BCC. Bacteria existing as clonal populations evolve diversity by the accumulation of point mutations, while nonclonal populations evolve through recombination within or between species. In this study identical alleles were found within species and between the different \textit{B. cenocepacia} subgroups but not between different BCC species (Table 3).

Evidence for clonal or recombining populations can be estimated by assessing the level of linkage between alleles at different loci around the chromosome. The index of association ($I_a$) (31) measures the extent of linkage. An $I_a$ not significantly greater than 0 after 1,000 computer randomizations would suggest that a single species population (monophyletic) is in linkage equilibrium (freely recombining), while a population with an $I_a$ significantly greater than 0 ($P < 0.001$) is considered to be in linkage disequilibrium (clonal). Since the BCC comprises many different species (polyphyletic), an $I_a$ value was not calculated for all 114 STs together; instead, each BCC species for which there were at least 10 different STs was examined.

Of the BCC species examined, \textit{B. vietnamiensis} exhibited the greatest evidence of recombination, with an $I_a$ value of $-0.067$ (14 STs), in contrast to \textit{Burkholderia ambifaria}, which exhibited the lowest $I_a$ value, 2.043 (12 STs). When the \textit{B. cenocepacia} subgroups III-A and III-B (16 STs) were combined, the $I_a$ value rose from a value of 0.374 for \textit{B. cenocepacia} III-B (10 STs) alone to 1.786, confirming that III-A and III-B are distinct subgroups of \textit{B. cenocepacia}. For \textit{B. cepacia} (14 STs) and \textit{Burkholderia multivorans} (13 STs), $I_a$ values of 0.431 and 0.852 were found, respectively. The number of STs for each species is low, and therefore, a much larger sample size is required for a more accurate comparison of mechanisms of evolution for each BCC species.

Relationships among the BCC and related species using concatenated nucleotide sequences. Comparisons of the topology of neighbor-joining trees for the nucleotide sequence of each individual locus (data not shown), including the \textit{trpB} locus located on the second-largest chromosome, revealed there was a high level of congruence between the trees at the interspecies level. The level of congruence within species varied from one species to another, since some species showed higher levels of congruence (e.g., \textit{Burkholderia stabilis} and \textit{B. ambifaria}) than others (e.g., \textit{B. vietnamiensis} and \textit{B. cenocepacia}).

In order to assess all the loci together in one tree, concatenated nucleotide sequences were used. Analysis of the allelic profiles by construction of an unweighted pair group method with arithmetic mean tree was found to be inappropriate due to the high level of variability between the alleles present at each loci.

Concatenated nucleotide sequences (2,773 bp) for the 114 BCC STs alongside sequences for strains of \textit{B. pseudomallei}, \textit{B. mallei}, and \textit{B. xenovorans} were analyzed using a neighbor-joining tree (Fig. 2), and the latter sequence was used to root the data. The BCC strains were fully resolved, falling into a distinctive broad cluster of strains, agreeing with the identification of all isolates as BCC isolates prior to this study. The BCC strains were clearly separated from \textit{B. xenovorans} and also to a much lesser extent from the \textit{B. pseudomallei} and \textit{B. mallei} strains (100% bootstraps). All of the known species of the BCC and most \textit{B. cenocepacia} subgroups were clearly distinguished with 100% bootstrap values.

All of the four known \textit{recA} lineages of \textit{B. cenocepacia} clustered into distinct groups (III-A, III-B, III-C, and III-D), each with high bootstrap values, along with a fifth \textit{B. cenocepacia} subgroup (III-E). \textit{B. cepacia} was also separated into sublineages: two clusters which had been observed previously, the type strain for \textit{B. cepacia} (21) and a group K cluster (\textit{B. cepacia} group K [36] with ST-96, -97, -98, -99, -100, -101, -102, and -103), with bootstrap values of 100% and 98%, respectively (Fig. 2). An additional four groups containing unidentified BCC isolates also appeared to fall outside of existing species clusters. The first of these, called \textit{B. cepacia} complex I, was composed of two strains of the same ST (ST-104), which formed a separate branch from \textit{B. vietnamiensis}. A second group of four unidentified BCC isolates (ST-105, -106, -107, and -108; \textit{B. cepacia} complex 2) clustered with 100% bootstraps from \textit{B. ambifaria}. A third group, \textit{B. cepacia} complex 3 (ST-109, -110, and -111), formed a cluster distinct from the other species with a 100% bootstrap. The fourth group of unidentified BCC isolates, \textit{B. cepacia} complex 4 (ST-112, -113, and -114), formed a separate branch in Fig. 2 from \textit{B. stabilis} and \textit{Burkholderia pyrocina} with 100% bootstraps.

Identification of epidemic CF strains. Epidemic CF strains previously implicated in patient-to-patient spread were also analyzed by MLST. The strain collection contained eight isolates identified as the ET12 transmissible lineage, which has spread within the Canadian and United Kingdom CF populations (20), and all were found to be part of a closely related
clonal complex. The five strains of this lineage isolated from United Kingdom CF patients were ST-28 and were identical at all seven MLST loci (isolates J2315, LMG 13316, LMG 13307, LMG 13327, and BCC0711) (Table 3). The three ET12 strains obtained from Canadian CF patients, C5424, K56-2, and BC7, were ST-29, -30, and -31, respectively, with each being single or double locus variants of the United Kingdom ET12 strains (Table 3). Unique sequence types were found for the other transmissible CF strains contained in the Bcc strain panel (7, 23) (Table 3): the Edinburgh outbreak B. multivorans strain (C1576) was ST-27, and the Manchester, United Kingdom (C1394), Sydney, Australia (CEP0511), and Cleveland, Ohio (PC184), epidemic B. cenocepacia strains were ST-35, -39, and -40, respectively. The strain (BCC0458) representative of the B. cenocepacia III-D isolates that have spread among CF patients in Italy (2) was ST-46. The transmissible BCC RFLP type K strain recovered from multiple Portuguese CF patients (9), represented by strain IST410, was found to be ST-96.

FIG. 2. Phylogenetic tree of concatenated nucleotide sequences from the seven loci, using the neighbor-joining method, Jukes-Cantor. Bootstrap values are shown for 1,000 replicates. Species names are given, followed by their former genomovar number, with the number of STs given in parentheses.

DISCUSSION

Schemes for the unequivocal typing and characterization of isolates are essential for epidemiological and evolutionary analysis of bacterial pathogens. Methodological differences in many genotyping techniques for the BCC reduce the efficacy of analyzing population genetics from one study to another. Strain typing based on the comparison of DNA sequence content rather than genome organization or restriction fragments is a more reliable and unambiguous indicator of strain identification, MLST is therefore highly appropriate for use on the BCC. In addition to accurate strain typing, the MLST method was shown to clearly differentiate all existing species in the closely related BCC. The ability to carry out both strain differentiation and species identification in a single approach represents a major advance that should greatly aid the clinical diagnosis of B. cepacia complex infection.

This BCC MLST scheme encompasses the most variable group of organisms thus far reported using this typing method in a single approach. MLST therefore meets a need for an easily transferable, precise, and reproducible typing tool for all species of the BCC. It is a simple tool that can offer a high level of strain identification without using polyphasic techniques. With environmental and clinical isolates still requiring comprehensive analysis, this study demonstrates that MLST could be used not only to resolve the BCC species but also to effectively disseminate the identity of these BCC isolates by this simple, widely used technique that is directly comparable via the Internet. It should provide a practical basis for multicenter collaborative analysis in a way not previously possible. Also, by comparison of patient information, isolate properties affecting disease prognosis might be better understood. Examination of the reference set of BCC strains assembled here to validate MLST has already shown that the approach can clearly identify epidemic CF strains and assist in the global infection control of these pathogens.

The ability to exchange genetic material is of growing clinical interest and concern. Recombination of even a single gene can have profound effects, including increased resistance to antimicrobials, vaccine immunity, and increased virulence. The low index of association values seen for some of the BCC species examined (B. cepacia, B. multivorans, B. cenocepacia III-B, and B. vietnamiensis) indicates that recombination has had an important role in their long-term evolution. These recombination events could be found among strains from different geographic locations (ST-15 and ST-16; ST-28 and ST-29) and may not be limited to just clinical or environmental isolates. MLST has also shown that strains of the ET12 lineage and other major transmissible strains constitute closely related clonal complexes, correlating to the minor variations seen in macrorestriction and RAPD analysis of these epidemic strain clusters (22).

Conversely, a high index of association values was obtained for B. ambifaria, implying that recombination has had less of a role in its evolution than in that of the other BCC species analyzed. Previous studies have reported B. stabilis to be a highly conserved population (34), and analysis of the six STs reported here so far concurs with this observation of clonality (Ia value of 2.417).

Certainly systems exist in the BCC to facilitate recombin-
tion, with an extensive presence of insertion sequences (17), phages (32), conjugative transfer genes, and genomic islands (35). Allele sharing has been found throughout several individual species in this study, which implies that recombination between different species could also be occurring. A larger collection of strains will need to be assessed by MLST to investigate species-to-species recombination further.

The MLST scheme reported here for the BCC provides a population scheme that is congruent with current species assignments for the BCC, allowing unambiguous identification at the species level. It also clearly resolves several unidentified groups of isolates, which should serve to support any future novel species or subgroup classification for them, thus providing a global platform from which important, high-level strain identification and epidemiological evaluation can be facilitated.

Future work will include a detailed comparison between MLST and other typing methods, such as PFGE and PCR fingerprinting-based techniques (5). A larger study will also be undertaken to further investigate recombination among the different BCC species, evaluate BCC clonal complexes, examine more unidentified BCC strains, and explore isolates from different ecological and epidemiological niches.

ACKNOWLEDGMENTS

This work was funded by the Wellcome Trust, grant number 072853 and in part by Heartlands Hospital, Birmingham.

We acknowledge the Sanger center for sequencing the J2315 B. cenocepacia genome and also Angela Marchbank (Cardiff University), Barry Curran, Helen Bird (Warwick University) and Lynne Richard- son (Oxford University). This publication made use of the

REFERENCES

1. Aaron, S. D., W. Ferris, D. A. Henry, D. P. Speert, and N. E. Macdonald. 2000. Multiple combination bactericidal antibiotic testing for patients with cystic fibrosis infected with Burkholderia cepacia. Am. J. Respir. Crit. Care Med. 161:1200–1201.
2. Agodi, A., E. Mahenthiralingam, M. Barchitta, V. Giannino, A. Sciacca, and S. Stefani. 2001. Burkholderia cepacia complex infection in Italian patients with cystic fibrosis: prevalence, epidemiology, and genomovar status. J. Clin. Microbiol. 39:2891–2896.
3. Berriatua, E., I. Zitzau, C. Miguel-Virto, P. Uribarren, R. Juste, S. Laevens, P. Vandamme, and J. R. Govan. 2001. Outbreak of subclinical mastitis in a flock of dairy sheep associated with Burkholderia cepacia complex infection. J. Clin. Microbiol. 39:990–994.
4. Coenye, T., E. Mahenthiralingam, D. Henry, J. J. LiPuma, S. Laevens, M. Gillis, D. P. Speert, and P. Vandamme. 2001. Burkholderia ambifaria sp. nov., a novel member of the Burkholderia cepacia complex including biocontrol and cystic fibrosis-related isolates. Int. J. Syst. Evol. Microbiol. 51:1481–1490.
5. Coenye, T., T. Spilker, A. Martin, and J. J. LiPuma. 2002. Comparative assessment of genotyping methods for epidemiologic study of Burkholderia cepacia genomovar III. J. Clin. Microbiol. 40:3300–3307.
6. Coenye, T., P. Vandamme, J. R. Govan, and J. J. LiPuma. 2001. Taxonomy and identification of the Burkholderia cepacia complex. J. Clin. Microbiol. 39:3427–3436.
7. Coenye, T., P. Vandamme, J. J. LiPuma, J. R. Govan, and E. Mahenthiralingam. 2003. Updated version of the Burkholderia cepacia complex experimental strain panel. J. Clin. Microbiol. 41:2797–2798.
8. Corey, M., and V. Farewell. 1996. Determinants of mortality from cystic fibrosis in Canada, 1970–1989. Am. J. Epidemiol. 143:1007–1017.
9. Cunha, M. V., J. H. Litao, E. Mahenthiralingam, P. Vandamme, L. Cito, C. Barreto, M. J. Salgado, and I. Sa-Corria. 2003. Molecular analysis of Burkholderia cepacia complex isolates from Portuguese cystic fibrosis cen- ter: a 7-year study. J. Clin. Microbiol. 41:4113–4120.
10. Curran, B., D. Jonas, H. Grundmann, T. Pitt, and C. G. Dowson. 2004.
31. Smith, J. M., N. H. Smith, M. O’Rourke, and B. G. Spratt. 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA 90:4384–4388.
32. Summer, E. J., C. F. Gonzalez, T. Carlisle, L. M. Mebane, A. M. Cass, C. G. Savva, J. LiPuma, and R. Young. 2004. Burkholderia cenocepacia phage BcepMu and a family of Mu-like phages encoding potential pathogenesis factors. J. Mol. Biol. 340:49–65.
33. Vandamme, P., B. Holmes, M. Vancanneyt, T. Coenye, B. Hoste, R. Coopman, H. Revets, S. Lauwers, M. Gillis, K. Kersters, and J. R. Govan. 1997. Occurrence of multiple genomovars of Burkholderia cepacia in cystic fibrosis patients and proposal of Burkholderia multivorans sp. nov. Int. J. Syst. Bacteriol. 47:1188–1200.
34. Vandamme, P., E. Mahenthiralingam, B. Holmes, T. Coenye, B. Hoste, P. De Vos, D. Henry, and D. P. Speert. 2000. Identification and population structure of Burkholderia stabilis sp. nov. (formerly Burkholderia cepacia genovar IV). J. Clin. Microbiol. 38:1042–1047.
34a. Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74.
35. Vermis, K., T. Coenye, J. J. LiPuma, E. Mahenthiralingam, H. J. Nelis, and P. Vandamme. 2004. Proposal to accommodate Burkholderia cepacia genomovar VI as Burkholderia dolosa sp. nov. Int. J. Syst. Evol. Microbiol. 54:689–691.
36. Vermis, K., T. Coenye, E. Mahenthiralingam, H. J. Nelis, and P. Vandamme. 2002. Evaluation of species-specific recA-based PCR tests for genomovar level identification within the Burkholderia cepacia complex. J. Med. Microbiol. 51:937–940.