INTRODUCTION

Strain W23 holds an important place in the history of Bacillus subtilis genetics. The development of B. subtilis as a bacterial model system began over 60 years ago with the pioneering mutagenesis studies of Burkholder & Giles (1947). Strain 168, a tryptophan auxotrophic mutant isolated in those experiments, was subsequently found to be naturally competent for genetic transformation (Spizizen, 1958). This discovery quickly spawned a research community focused on exploring the genetics and development of this Gram-positive spore-former and on exploiting its biotechnological potential.

In addition to strain 168, a handful of other Burkholder and Giles mutants were also preserved (Spizizen, 1984). One of them, strain 23, was a threonine auxotroph that can readily revert to prototrophy (D. R. Zeigler, unpublished results). Strain W23, believed to be one such revertant, was widely viewed as a ‘wild-type’ counterpart to strain 168. For several years, W23 served as a ‘universal donor’ and 168 derivatives as ‘universal recipients’ for strain construction and genetic mapping (Anagnostopoulos & Spizizen, 1961; Nester & Lederberg, 1961). Evidence gradually accumulated, however, that W23 and 168 were fundamentally different in cell wall composition, resident prophages and genome content (Zeigler et al., 2008). It is now clear that 168 is closely related to – and probably a descendant of – the B. subtilis Marburg strain, which is stocked in the American Type Culture Collection (ATCC) as 6051T and in the National Collection of Industrial, Marine and Food Bacteria (NCIMB) collection as 3610T (Srivatsan et al., 2008; Zeigler et al., 2008). It is equally clear that W23 must have arisen from an independent isolate, although its exact origin has remained a mystery (Hemphill & Whiteley, 1975).

Strain W23 continues to have relevance for the B. subtilis research community. For example, its cell wall is constructed from ribitol teichoic acid polymers, rather than the glycerol teichoic acid found in 168 (Coley et al., 1978). In this way W23 more closely resembles the Gram-positive pathogen Staphylococcus aureus (Qian et al., 2006). Understanding the enzymology (Brown et al., 2010) and regulation (Minnig et al., 2005) of W23 wall synthesis remains an ongoing concern. W23 also remains a useful comparison to 168 for the study of B. subtilis taxonomy, ecology and evolution, since W23 is the best-characterized member of the subspecies spizizenii and 168 of subspecies subtilis (Nakamura et al., 1999). Finally, the study of W23 has relevance for biotechnology, since 168–W23 hybrid strains have been shown to have subtle but potentially significant differences in physiology and metabolism from those seen in 168 (Zeigler et al., 2008).

As a stimulus to further research, the complete genome sequence of B. subtilis W23 is reported here. The 4.0 Mb sequence of Bacillus subtilis subsp. spizizenii W23: insights into speciation within the B. subtilis complex and into the history of B. subtilis genetics

Daniel R. Zeigler

The genome sequence of Bacillus subtilis subsp. spizizenii W23 has been determined. The sequence strongly suggests that W23 is a direct descendant of B. subtilis ATCC 6633. W23 shares a 3.6 Mb core genome with the intensively studied model organism B. subtilis subsp. subtilis 168, and gene order within this core has been strongly conserved. Additionally, the W23 genome has 157 accessory (that is, non-core) genome segments that are not found in 168, while the 168 genome has 141 segments not found in W23. The distribution of sequences similar to these accessory segments among other genomes of the B. subtilis species complex shows that those sequences having entered into the phylogeny of the complex more recently tend to be larger and more AT-rich than those having entered earlier. A simple model can account for these observations, in which parasitic or symbiotic DNAs are transferred into the genome and then are reduced in size and modified in base composition during speciation.

Abbreviations: ATCC, American Type Culture Collection; HGT, horizontal gene transfer.

The GenBank/EMBL/DDBJ accession number for the genome sequence of Bacillus subtilis subsp. spizizenii W23 is CP002183.

Two supplementary tables and two supplementary figures are available with the online version of this paper.
Phylogenetic trees were generated by DNAMLK (Felsenstein, 1989) or (CP000813), Bacillus mojavensis ATCC 14580 (AE017333), and Bacillus subtilis (NCBI AP011541), the latter of which includes B. subtilis 168 (AL009126.3) and its close relatives. The identified accessory segments from W23 or 168 were then aligned individually with the 12 genome sequences using the BLASTn algorithm. Two criteria were applied to sequence matches in related genomes to identify orthologues likely to have arisen from the same horizontal transfer event. First, the DNA sequence alignment of the segment with the candidate orthologue was required to have an E value < 10^{-2} and a length covering at least 60% of the segment. Second, the segment and orthologue were also required to share at least one homologous flanking sequence (a match of at least 200 nt within 2 kb upstream or downstream) and the alignments of these flanking sequences were again required to have an E value < 10^{-2}.

Optical mapping. An Ncol fragment optical map was prepared for the W23 genome at OpGen Technologies according to methods described previously (Reslewic et al., 2005).

**RESULTS AND DISCUSSION**

The B. subtilis subsp. spizizenii W23 genome assembly presented here has a G+C content of 43.89 mol% and a length of 4,027,676 nt, a value consistent with estimates based on physical methods (Kavenoff, 1972). Gene prediction algorithms uncovered 4062 genes encoding proteins and 107 encoding structural RNAs. Several previously characterized features were recognized in the sequence, including defective prophage PBSZ (Seaman et al., 1964), the tryptophan operon region (Zeigler et al., 2008), the ribitol teichoic acid gene cluster (Lazarevic et al., 2002) and the biosynthetic gene clusters for the peptide antibiotics rhizocin (Borisova et al., 2010) and mycosubtilin (Duitman et al., 1999).

**rDNA operons within the W23 genome**

Although ten rRNA operons have been annotated in the strain 168 genome, only eight were found in W23. In strain 168, five of the operons are located in tandem within two clusters, rrf–rrnW and rnl–rrnH–rrnG, which have proved difficult to resolve using short-read sequencing technology (Barbe et al., 2009). In the present study, repeated attempts to clone inter-operon spacer regions yielded only a single fragment corresponding to a tRNA gene cluster located between rnr1 and rnrH in strain 168; spacers corresponding to the rrf–rrnW or rnrH–rrnG inter-operon sequences were never found (results not shown). To confirm the number and arrangement of the rRNA operons in strain W23, an ordered Ncol restriction map of its entire genome was created using optical mapping technology (Latreille et al., 2007; Reslewic et al., 2005). The overall sequence assembly for the genome was strongly confirmed, as there were no significant differences in size or order between restriction fragments predicted by in silico analysis and those observed by optical mapping (see Supplementary Fig. S1, available with the online version of this paper). The existence of a single rRNA operon in place of rrf–rrnW and two in place of rnr1–rrnH–rrnG was also consistent with the optical map. The genome sequence predicted that the tandem operon pair...
should be contained in adjacent NcoI fragments of 8.6 kb and 9.8 kb; the observed restriction fragments in this region closely matched the prediction (Fig. 1). A similar correspondence was observed between the predicted 21.1 kb fragment containing a single rnl-like operon and the observed 20.3 kb fragment. In neither case is there room in the optical map data for the additional 4–5 kb required for an rRNA operon and its flanking sequences. These results are consistent with a published report that strains from subspecies spizizenii have eight to ten rRNA operons, based on restriction fragment length polymorphisms seen in Southern blots (Shaver et al., 2002). It is interesting to note that the number of rRNA operons is reduced from ten to nine in some laboratory strains derived from 168, apparently due to homologous recombination between operons in either the rnl–rnnW or rnl–rnnH–rnnG cluster, with no obvious deleterious effects for the strain (Widom et al., 1988). It is possible that similar deletion events may have occurred in the W23 lineage.

The origin of strains W23 and 168

Unexpectedly, analysis of the W23 genome sequence strongly suggests an answer to a longstanding puzzle about the strain’s origins: it is apparently a direct descendant of B. subtilis strain ATCC 6633. This strain was isolated around 1911–1912 in the laboratory of Karl F Kellerman, who had developed a method for selecting cellulose-degrading bacteria from environmental samples (Kellerman & McBeth, 1912). Kellerman’s colleague, Nathan R. Smith, maintained the isolate for many years as strain NRS-231 (Smith et al., 1952). In 1944, the isolation of the lantibiotic subtilin from NRS-231 was reported (Jansen & Hirschmann, 1944). At that time Smith deposited the strain in the ATCC under accession number 6633. Fortuitously, a group investigating the biosynthesis of the antifungal peptide rizocidin recently released a draft genome sequence for ATCC 6633, assembled into 37 contigs comprising 3979 kb (Borisova et al., 2010). The finished genome sequence of W23 is 99.995% identical with the draft sequence of strain 6633, differing in only 216 nt. This relatively small number of polymorphisms can be attributed to the different sequencing and assembly methodologies employed by the two projects, along with an exposure of the W23 parent to a round of mutagenesis to create strain 23, followed by years of subsequent cultivation and maintenance in the laboratory. This observation fills a gap in the early history of B. subtilis genetics. The exact details may never be known, but the following scenario seems plausible: Burkholder and Giles probably received at least two strains from the ATCC in the late 1940s. ATCC 6051T would have been the parent of strain 168 and its siblings, while ATCC 6633 would have been the parent of strain 23, from which W23 was derived.

Comparison of the W23 and 168 genomes

W23 and 168 represent divergent lineages within B. subtilis (Nakamura et al., 1999); the last common ancestor of their core genomes would have been very closely related to the most recent progenitor of the entire B. subtilis species. Comparison of their genome sequences could provide valuable insights into the changes that have accompanied speciation within this bacterial group. Orthologous sequences shared by the strains would presumably have been inherited vertically from their common ancestor, with minor sequence differences arising from a slow, continuous process of random mutation subject to natural selection. Sequences unique to one of the strains would have arisen either by genome expansion through horizontal transfer or reduction through deletion. For simplicity, the following discussion will adopt the ‘core genome’ and ‘accessory genome’ terminology to distinguish the shared (core) from non-shared (accessory) content.

The W23 and 168 genomes share a 3.6 Mb core genome that averages 92.4% nucleotide sequence identity (results not shown). Gene order has been remarkably conserved within this shared content, on both the local and global levels. No major rearrangements are apparent in the genome alignment, and synteny appears to be perfectly conserved, the only exception being a segment specifying spermidine resistance that has the gene order bltR–bltD in 168 and bltD–bltR–blt in W23. In addition to the shared content, the W23 genome has 386 kb of unique sequence not found in 168, while the 168 genome has 567 kb not found in W23. Much of this accessory genome content appears to have originated with the transfer of mobile genetic elements, including plages, plasmids, transposons and insertion sequences. Within the 168 genome, the fully functional Spβ prophage accounts for 138 kb of unique content; the ICEB1 mobile element and the phage-like skin element account for an additional 20 kb and 48 kb, respectively (Auchtung et al., 2005; Lazarevic et al., 1998; Takemaru et al., 1995). The W23 genome likewise has been shaped by mobile elements. A 39 kb prophage of...
unexplored function, named here \( \phi W23 \), is integrated into a tRNA gene, trnSL-Arg1, located near the replication terminus. Ten copies of a 1.3 kb insertion sequence, IS\( Bu1 \), are also present within the W23 genome. In summary, analysis of the unique content uncovered 157 accessory genome segments dispersed among 129 discrete sites within the W23 genome that distinguish it from strain 168 (see Supplementary Table S1). Similarly, there are 141 segments dispersed among 106 sites in the 168 genome that distinguish it from strain W23 (see Supplementary Table S2). These 298 accessory segments can be understood to represent instances of either genome expansion through horizontal gene transfer (HGT) or reduction through deletion. Fig. 2 shows the sizes and locations of the accessory genome segments for each strain relative to their similar collinear core genome.

**Fig. 2.** Comparison of the *B. subtilis* W23 and 168 genomes. The collinear core genome that is conserved in both strains is indicated by grey rectangles, while accessory genome sequences that distinguish the strains are indicated by black rectangles. Specific accessory genome segments mentioned in the text and other large features are labelled. ‘RM’ denotes a segment encoding restriction–modification enzymes.
Tracking the acquisition and processing of horizontally transferred DNA during speciation in the *B. subtilis* complex

Several recently described bacterial taxa appear to be closely related to *B. subtilis* based on morphology, biochemical tests and 16S rRNA gene sequence comparisons. This phylogenetic cluster has informally been termed the *B. subtilis* species complex (Rooney et al., 2009). *B. licheniformis* and *B. pumilus* are often grouped with *B. subtilis* (Borriss et al., 2010), and in this analysis will be considered to define the phylogenetic boundaries of the complex. Several genome sequences from the *B. subtilis* complex are now substantially complete. Fig. 3 presents a phylogenetic tree, constructed by aligning the concatenated full-length DNA sequences from eight housekeeping genes (see text) were aligned with CLUSTAL W to generate a distance matrix, which was rendered as a bootstrapped maximum-likelihood tree by the PHYLIP DNAMLK application. Strain abbreviations: B sub W23, B sub TU-B-10 and B sub DV1-B-1 are *B. subtilis* subsp. *spizizenii*; B sub 168, BEST195 and RO-NN-1 are *B. subtilis* subsp. *subtilis*; B moj is *B. mojavensis*; B atr is *B. atrophaeus*; B amy FZB42 is *B. amylophilus* subsp. *plantarum*; B amy DSM7 is *B. amylophilus* subsp. *amylophilus*; B lic is *B. licheniformis*; B pum is *B. pumilus*. *B. meg DSM319 is Bacillus megaterium* (outgroup). For comparison, an unrooted tree constructed with the PhyML application is presented in Supplementary Fig. S2.

A detailed analysis tracing the history of the W23-like segments in the *B. subtilis* complex is presented in Supplementary Table S1 and a summary in Table 1. A similar analysis of the 168-like segments is presented in Supplementary Table S2 and Table 2. Tables 1 and 2 categorize the accessory genome segments based on when they appear to have entered the phylogeny of the *B. subtilis* complex. The top row in each table enumerates segments that are truly strain-specific. Subsequent rows grow in phylogenetic depth until the bottom rows are reached, which encompass the entire complex. The distribution of these accessory segments is sometimes patchy, as can been observed by reading across the rows in the Supplementary Tables, and this observation suggests their complete loss in some lines of descent. Chromosome reduction through deletion has been viewed as a more parsimonious explanation than multiple rounds of HGT to explain similar patchiness in other phylogenies (Koonin & Wolf, 2008). For most accessory segments, the DNA sequence identity values within a given genome (see Supplementary Tables S1 and S2) fall within the range expected from housekeeping gene comparisons, such as those used to produce the phylogram in Fig. 3. This observation supports the notion that each segment has co-evolved with the genomes after a single ancestral HGT event.

Tables 1 and 2 can be read, then, as records of HGT events, with comparatively recent ones near the top and more ancient ones near the bottom. Two trends are immediately obvious: the putative HGT units decrease in size and increase in G+C content as one moves from the top rows to the bottom. If one posits that the nature of HGT has remained fairly constant during the speciation of the *B. subtilis* complex, it follows that genome evolution in this group has involved reducing the size and adjusting base composition of transferred DNA. Both processes could be explained if horizontally transferred material is primarily parasitic or symbiotic DNA. Phages, plasmids and insertion sequences have consistently been found to have an AT nucleotide bias compared with their hosts, over a very large strains to species (Zeigler, 2003). The clustering of these organisms into species and subspecies is strongly supported by this analysis.
2038

**Table 1.** Distribution of W23-like segments in accessory genomes from the *B. subtilis* species complex

<table>
<thead>
<tr>
<th>Genome distribution*</th>
<th>&lt;1 kb</th>
<th>1–4 kb</th>
<th>&gt;4 kb</th>
<th>Mol% G + C</th>
<th>Representative coding potential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique to W23</td>
<td>10</td>
<td>24</td>
<td>8</td>
<td>37.4</td>
<td>ISBslI insertion sites; rhizotisin synthesis cluster; φW23 prophage; other phage-like elements; ribitol teichoic and teichuronic acid cluster; short-chain dehydrogenases, aldo-keto reductases</td>
</tr>
<tr>
<td>W23 and TU-B-10</td>
<td>12</td>
<td>9</td>
<td>2</td>
<td>39.6</td>
<td>Subtilin synthesis cluster; glycosyltransferases; nitrogen reductase</td>
</tr>
<tr>
<td>W23 and DV-1-B</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>38.6</td>
<td>Short-chain dehydrogenase; CAAX amino terminal protease; glycerolphosphoryl diester phosphodiesterase</td>
</tr>
<tr>
<td><em>spizizenii</em> subspecies</td>
<td>7</td>
<td>10</td>
<td>0</td>
<td>41.3</td>
<td>Branched-chain amino acid transport; Major Facilitator Superfamily efflux transporter; chromate transporter; Na⁺/H⁺ antiporter; bacterial acid phosphatase; biofilm formation, necrosis, acetoin reductase</td>
</tr>
<tr>
<td><em>subtilis</em> species</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>39.3</td>
<td>Repeated non-coding sequence; spore coat polysaccharide biosynthesis protein; GNAT family acetyltransferase; HIT-family hydrolase; carbohydrate transport and metabolism</td>
</tr>
<tr>
<td>B sub + B moj</td>
<td>4</td>
<td>15</td>
<td>1</td>
<td>40.2</td>
<td>Urea ABC transporter; ABC multidrug resistance systems; divalent metal ion transporter; type I restriction–modification system; EamA-family permease; TerC-family transporter</td>
</tr>
<tr>
<td>B sub + B moj + B atr + B amy</td>
<td>9</td>
<td>10</td>
<td>2</td>
<td>40.9</td>
<td>Mycosubtilin biosynthesis cluster; antibiotic transport protein; β-1-4-glucosyltransferase (membrane synthesis); ABC transporters; endo-1,4-β-xylanase; short-chain dehydrogenase/reductase</td>
</tr>
<tr>
<td>Entire species complex</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>42.4</td>
<td>Glycoside hydrolase; β-glucoside transporter; glycosyltransferases; Na⁺/phosphate symporter; fatty acid desaturase; cell wall anchor domain-containing protein; efflux permease; cytochrome P450</td>
</tr>
</tbody>
</table>

*Taxonomic abbreviations are as in Fig. 3.

range of bacterial taxa (Rocha & Danchin, 2002). It has also been observed that bacteria inactivate resident prophages primarily by large deletions rather than by the accumulation of point mutations (Casjens, 2003). Consistent with this view, the upper rows of both tables contain the strain-specific viable phage SPβ and other large phage-like elements of unknown function. Of broader phylogenetic distribution are the *skin* element, a defective phage that has been co-opted to function in the regulation of sporulation (Takemaru *et al*., 1995), and an integrating conjugative element (ICE) that has no obvious ability to harm the host during its life cycle (Wozniak & Waldor, 2010). The lowest rows in the tables, representing the accessory genome segments with the broadest distribution and a presumably more ancient entry into the phylogeny, are primarily filled by genes encoding transporters, permeases and enzymes as well as proteins of unknown function. These base frequencies of these genes have been ameliorated to a large extent to more closely match the characteristic frequency for *B. subtilis*. These processes – the incorporation of HGT material with anomalous base composition into the genome and its gradual amelioration – have been well documented for other bacteria (Lawrence & Ochman, 1997, 1998).

These observations suggest a simple model of genome evolution during speciation that is compatible with insights gleaned from comparative genomic studies within genus *Escherichia* (Touchon *et al*., 2009) and with more complex models that seek to explain bacterial evolution on a global scale (Gogarten & Townsend, 2005; Koonin & Wolf, 2008).

1. **HGT**, primarily through the infection of temperate phages and the acquisition of self-transmissible mobile elements, is frequent. (2) **Over time, any deleterious effects of these largely parasitic DNAs are mitigated and eventually eliminated by deletion events.** (3) **Retained sequences increase the fitness of the organism by expanding its metabolic capacity, increasing its antimicrobial arsenal, or protecting it from harmful environmental conditions.**

**Natural selection also tailors the retained sequences to match more closely the base composition of the host genome.** HGT therefore can lead to the emergence of new, ecologically distinct populations (Kopac & Cohan, 2011).

**Preliminary analysis of the genomes of strains BSn5 and gtP20b**

After the principal analyses were completed for this report, two additional genome sequences were released for novel *B. subtilis* strains, the endophyte BSn5 (Deng *et al*., 2011) and the marine sediment isolate gtP20b (Fan *et al*., 2011). Comparison of these sequences to the genomes of 168 and W23 with the BLAT application of the Genome-To-Genome Distance Calculator (Auch *et al*., 2010) shows that BSn5 is significantly more closely related to strain 168 (estimated hybridization 91.4 %) than to W23 (estimated 78.5 %), while gtP20b is about as similar to 168 and W23 (estimated 78.4 % and 80.8 %, respectively) as 168 and W23 are to each other (76.1 %). This analysis suggests that BSn5, like...
Table 2. Distribution of 168-like segments in accessory genomes from the B. subtilis species complex

<table>
<thead>
<tr>
<th>Genome distribution*</th>
<th>&lt;1 kb</th>
<th>1–4 kb</th>
<th>&gt;4 kb</th>
<th>Mol%G+C</th>
<th>Representative coding potential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique to B sub 168</td>
<td>3</td>
<td>11</td>
<td>10</td>
<td>35.0</td>
<td>SPb prophage; sporulation killing factor synthesis/export; Spβ killing factor operon; ComQXP competence proteins; BsuM restriction system; tetracycline antiporter; efflux transporter; sulfur-containing amino-acid ABC transporter; 24 kb phage-like element</td>
</tr>
<tr>
<td>B sub 168 and BEST195</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>36.8</td>
<td>Phage proteins; potassium/proton-divalent cation antiporter; branched-chain amino acid transporters; acetyltransferases; oxidoreductases; efflux transporter</td>
</tr>
<tr>
<td>B sub 168 and RO-NN-1 subtilis subspecies</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>37.6</td>
<td>Phage protein; N-acetyltransferase; biofilm forming exported protein</td>
</tr>
<tr>
<td>subtilis species</td>
<td>10</td>
<td>14</td>
<td>3</td>
<td>38.9</td>
<td>RNase inhibitor; cytochrome P450; streptomycin resistance; efflux transporter; phage proteins</td>
</tr>
<tr>
<td>B sub + B moj</td>
<td>3</td>
<td>9</td>
<td>2</td>
<td>41.3</td>
<td>ECF sigma factor Z; efflux transporters; spore germination proteins; skin element and other phage proteins; cell envelope stress genes</td>
</tr>
<tr>
<td>B sub + B moj + B atr + B any</td>
<td>9</td>
<td>21</td>
<td>2</td>
<td>41.4</td>
<td>Maltooligosaccharide utilization operon; NatR/NatK system; Na+ ABC efflux transporter; DNA-3-methyladenine glycosylase; DNA alkyltransferases; toxic compound efflux transporter; extracellular neutral protease B</td>
</tr>
<tr>
<td>Entire species complex</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>43.3</td>
<td>ICEB1 mobile element; lipopstatin synthesis clusterefflux transporters; spore-specific glycosylases; ferrous ion permease</td>
</tr>
</tbody>
</table>

*Taxonomic abbreviations are as in Fig. 3.

168, could be assigned to subspecies subtilis, while gtP20b, although labelled as subspecies spizizenii on unspecified grounds (Fan et al., 2011), more probably represents a novel subspecies. The availability of these additional sequences affords an opportunity to assess the reasonableness of the B. subtilis ‘core genome’ identified by alignment of the sequences for 168 and W23. BLAST alignments (not shown) reveal that BSn5 has orthologous sequences matching all but 7.9 kb of the proposed 3630 kb core genome. The average sequence identity shared by BSn5 and 168 throughout this core is 99.3%, indicating that nearly all of the genome differences between these strains fall in their ‘accessory’ portions. Although gtP20b is much more distantly related to both 168 and W23, it nevertheless includes all but 68.9 kb of the core genome proposed based on their alignment (not shown). The sequence identity shared in the remaining core sequences for gpT20b with 168 (93.1%) and W23 (94.3%) is similar to the level of sequence identity shared by W23 and 168 (92.4%) over the core. In summary, while the inclusion of two additional genome sequences allows a slight refinement of the assignments, the identification of ‘core’ and ‘accessory’ sequences is a robust concept that could prove useful for analysing B. subtilis genomes.

W23-derived sequence islands in B. subtilis legacy strains

Many laboratory strains of B. subtilis are descended from legacy strains constructed nearly 50 years ago, before W23 fell into disfavour as a source of transforming DNA. Because the trpC2 allele of strain 168 is non-reverting (Albertini & Galizzi, 1999), most Trp+ laboratory strains are mosaics that derive a portion of their genomes from W23. The phenomenon of congression allows multiple fragments of DNA to be simultaneously incorporated into the 168 genome when the donor DNA is at a saturating concentration (Cutting & Vander Horn, 1990). Indeed, some Trp+ B. subtilis legacy strains, such as PY79, have multiple ‘islands’ of W23 DNA in their genomes, some of them unlinked to the trpC locus (Zeigler et al., 2008). One motive of the present research, then, was to enable the B. subtilis research community to recognize the presence of W23 DNA in sequence data derived from other strains, including those that are phenotypically Trp-. A BLAST search of the NCBI sequence database with individual reading frames from the W23 genome uncovered one such example.

Much of what is known about the biology of defective phage PBSX comes from studies with B. subtilis RB1479. This strain, a mutant derived from the legacy strain 168TT (trpC2 thyA1 thyB1), has two relevant phenotypes. First, it is temperature-inducible for PBSX due to an allele termed xin-1479. Second, unlike wild-type PBSX, lysates from RB1479 fail to kill W23 cells, a trait assigned to the allele xki-1479, which was shown to be genetically linked to the xin locus (Buxton, 1976). Later, the PBSX repressor gene xre from RB1479 was characterized in detail (Wood et al., 1990). The sequence (GenBank accession M36477) was only 94% identical to that of the database genome.
sequence for strain 168, but Wood and colleagues, interpreted these differences as arising from mutagenesis and suggested one single nucleotide polymorphism to be a candidate for the xin-1479 lesion. Based on the W23 genome sequence, however, it is now possible to identify the xre gene of RB1479 as 100% identical to the xre gene of the W23 defective prophage, PBSZ. The parent of RB1479, 168TT, has both 168 and W23 in its lineage (Farmer & Rothman, 1965). It is probable, then, that a W23 sequence island within the PBSX prophage was inadvertently introduced during strain construction. This island may also be responsible for the Xki phenotype, since PBSZ does not kill W23. This example illustrates how genomic mosaicism in B. subtilis legacy strains can confound analysis of genetic and physiological data. It is hoped that the completed W23 genome sequence may help detect and solve other similar puzzles in the literature.

ACKNOWLEDGEMENTS

The BGSC is funded in part by a grant from the National Sciences Foundation (0234214). I thank Frederick M Cohan, William W Metcalf and John B Perkins for fruitful discussions about the project and this paper.

REFERENCES


Bacillus subtilis subsp. spizizenii W23 genome sequence


Edited by: D. W. Ussery