Multilocus sequence analysis of phytopathogenic species of the genus *Streptomyces*

David P. Labeda

National Center for Agricultural Utilization Research, USDA-ARS, Peoria, IL 61604, USA

The identification and classification of species within the genus *Streptomyces* is difficult because there are presently 576 species with validly published names and this number increases every year. The value of multilocus sequence analysis applied to the systematics of *Streptomyces* species has been well demonstrated in several recently published papers. In this study the sequence fragments of four housekeeping genes, *atpD*, *recA*, *rpoB* and *trpB*, were determined for the type strains of 10 known phytopathogenic species of the genus *Streptomyces*, including *Streptomyces scabiei*, *Streptomyces acidiscabies*, *Streptomyces europeiscabiei*, *Streptomyces luridiscabies*, *Streptomyces niveiscabies*, *Streptomyces puniciscabies*, *Streptomyces reticuliscabies*, *Streptomyces stelliscabies*, *Streptomyces turgidiscabies* and *Streptomyces ipomoeae*, as well as six uncharacterized phytopathogenic *Streptomyces* isolates. The type strains of 52 other species, including 19 species observed to be phylogenetically closely related to these, based on 16S rRNA gene sequence analysis, were also included in the study. Phylogenetic analysis of single gene alignments and a concatenated four-gene alignment demonstrated that the phytopathogenic species are taxonomically distinct from each other in spite of high 16S rRNA gene sequence similarities and provided a tool for the identification of unknown putative phytopathogenic *Streptomyces* strains at the species level.
of the housekeeping genes \textit{atpD}, \textit{recA}, \textit{rpoB} and \textit{trpB} were determined following the methods of Guo et al. (2008). It had been observed that two copies of the \textit{gyrB} operon (one degenerate) appeared to be present in the genome sequence of \textit{Streptomyces scabiei} RL87.22 available on the BacMap (Stothard et al., 2005) website (http://wishart.biology.ualberta.ca/BacMap/) at the time this study was initiated and the quality of the alignments of the sequence data for this gene from the previous studies of Guo et al. (2008), Rong et al. (2009) and Rong & Huang (2010) appeared as if more than a single copy had been amplified and sequenced. It was decided, therefore, that this gene locus would be omitted from the present study and only those genes certain to be present as single copies would be analysed. The gene sequences of these gene loci had been determined in several published studies (Guo et al. 2008; Rong et al., 2009; Rong & Huang, 2010). The sequences of 36 strains were obtained from the public databases to be included in the present analysis. The sequences were aligned and edited using BioEdit (Hall, 1999) and the unique allele sequences for each gene locus, including those from the published sequences and those determined in the present study, were loaded into a local implementation of the \textit{Streptomyces} MLST database (http://pubmlst.org/streptomyces/) utilizing the mlstdbNET application of Jolley et al. (2004). The sequence data were exported as single gene alignments or a concatenated four-gene alignment for subsequent analysis using the Tamura–Nei distance model (Tamura & Nei, 1993), using the neighbour-joining (Saitou & Nei, 1987) and maximum-parsimony (Felsenstein, 1993) algorithms with 1000 bootstrap replications (Felsenstein, 1985) in version 4 of the \textsc{MEGA} software package (Tamura et al., 2007), as well as the maximum-likelihood algorithm (Felsenstein, 1993) in the BioEdit program.

Phylogenetic relationships of the phytopathogenic species of the genus \textit{Streptomyces} and neighbouring taxa based on 16S rRNA gene sequences can be seen in Fig. 1a and b. Because of the highly conserved nature of this gene within the genus \textit{Streptomyces}, there is some question regarding the limitations it has for resolution and use in species discrimination. It is of note that the majority of the phytopathogenic species are observed within the so-called \textit{Streptomyces diastatochromogenes} 16S rRNA gene phylogenetic branch, or very near to it, as can be seen in Fig. 1a; this is in agreement with the traditional grouping of potato scab-producing strains by plant pathologists. \textit{Streptomyces acidiscabies}, \textit{Streptomyces niveiscabiei}, and \textit{Streptomyces punicaeabiei} are phylogenetically distant from the \textit{S. diastatochromogenes} clade, as can be seen in Fig. 1b. \textit{Streptomyces luridiscabiei} has recently been proposed as a later synonym of \textit{Streptomyces microflavus} (Rong & Huang, 2010) based on a multilocus sequence study and therefore its distant phylogenetic position with the \textit{Streptomyces microflavus} clade in the large \textit{Streptomyces} tree is not shown.

Phylogenetic relationships of the individual housekeeping genes for this set of strains (Figs 2–5) and the tree constructed from the concatenated sequence alignment of four housekeeping genes (2111 bp) (Fig. 6) largely agreed with those inferred from analysis of 16S rRNA gene sequences (Fig. 1a, 1b), but bootstrap support for these groupings was significantly higher. The topology of the phylogenetic trees constructed using the maximum-parsimony and
Fig. 2. Phylogenetic tree calculated from partial sequences of the atpD gene (496 bp) using the Tamura–Nei evolutionary distance method and the neighbour-joining algorithm. Bootstrap values $>60\%$ (based on 1000 resampled datasets) are given at nodes. Bar, 0.02 substitutions per nucleotide position.

Fig. 3. Phylogenetic tree calculated from partial sequences of the recA gene (504 bp) using the Tamura–Nei evolutionary distance method and the neighbour-joining algorithm. Bootstrap values $>60\%$ (based on 1000 resampled datasets) are given at nodes. Bar, 0.02 substitutions per nucleotide position.
maximum-likelihood algorithms was quite similar. It was noted, however, that *S. diastatocromogenes* NRRL B-1698<sup>T</sup> was phylogenetically distant from the scab-producing taxa in all of the single gene trees as well as in the concatenated four-gene tree, while the plant pathogenic species were distributed into eight distinct clades in the latter tree (see Fig. 6). *Streptomyces bottropensis* NRRL ISP-5262<sup>T</sup>, *S. europaeascabiei* NRRL B-24443<sup>T</sup>, *S. scabiei* NRRL B-16523<sup>T</sup> and *S. stelliscabiei* NRRL B-24447<sup>T</sup> were members of a statistically well-supported clade in the phylogenetic trees calculated from *recA* gene sequences (Fig. 3) and *rpoB* gene sequences (Fig. 4) and were part of a larger cluster that includes clades 1, 2 and 3 in the tree (Fig. 6) calculated from the four-gene concatenated alignment. All of the species, however, could be differentiated from each other based on their concatenated four-gene phylogeny (Fig. 6) and the differences observed in their individual housekeeping gene alleles (Supplementary Table S2).

Bouchek-Mechiche et al. (2006) proposed that *S. turgidiscabiei* and *S. reticuliscabiei* belong to the same genomic species but suggested that the names of these species be conserved because the type of scab disease produced on potatoes by each is morphologically distinct and combining the names would have caused confusion within the plant pathology community. In the present study it was observed that *S. turgidiscabiei* and *S. reticuliscabiei* formed a two-membered clade distinct from the *S. scabiei* cluster in the four-gene tree (clade 6, Fig. 6), but they could also be clearly differentiated from one another based on differences in their *atpD* and *trpB* gene sequences (Figs 2 and 5). This was reflected in the fact that they shared the same alleles for both the *recA* and *rpoB* genes but contained different alleles for the *atpD* and *trpB* genes. This observation supported the proposal of Bouchek-Mechiche et al. (2006) to maintain them as separate species.

The other phytopathogenic species were also phylogenetically distant from the *S. scabiei* cluster in the concatenated four-gene phylogenetic tree, with *S. acidiscabiei* NRRL B-16524<sup>T</sup> and *S. niveiscabiei* NRRL B-24457<sup>T</sup> forming one well-supported clade (clade 5, Fig. 6) and *S. ipomoeae* NRRL B-12321<sup>T</sup> and *S. puniceiscabiei* NRRL B-24456<sup>T</sup> forming another (clade 7, Fig. 6). The members of each of these clades, however, appear to represent distinct species since the members of each clade do not share common alleles for any of the individual housekeeping genes. The housekeeping gene sequences of *S. luridiscabiei* NRRL B-24555<sup>T</sup> were used as a positive control for comparison against those reported by Rong & Huang (2010); not surprisingly, they were found to agree since they represent the same strain. In their recently published study Rong & Huang (2010) proposed that *S. luridiscabiei* LMG 21390<sup>T</sup> represents a later synonym of *S. microflavus*, but results from the present study indicate that they are likely to represent related but distinct species because *S. luridiscabiei* LMG 21390<sup>T</sup> (=NRRL B-24555<sup>T</sup>) and *S. microflavus* CGMCC 4.1428<sup>T</sup> share no alleles in common for any of the four housekeeping genes.

From comparisons using the EzTaxon website (Chun et al., 2007) *Streptomyces* strains NRRL B-24083, NRRL B-24084, NRRL B-24085, NRRL B-24089 and NRRL B-24090 exhibited 16S rRNA gene sequence similarity values >99% as well as DNA relatedness values of 49, 50, 36, 96 and 98%, respectively, with respect to *S. scabiei* ATCC 49173<sup>T</sup> (=NRRL B-16523<sup>T</sup>) as determined previously in the study of Bukhalid et al. (2002). Strain NRRL B-24079 exhibited 99.7% 16S rRNA gene sequence similarity to the type strain of *S. scabiei* but genomic DNA relatedness was not determined. Analysis of the phylogenetic position of these strains based on the concatenated four-gene alignment (Fig. 6) and the specific alleles present for each housekeeping gene made it possible to identify them as members of some of the recognized phytopathogenic species. Strains NRRL B-24089 and NRRL B-24090 were identified as members of the species *S. scabiei*, as seen in Fig. 6, clade 2, having identical alleles for the *atpD*, *recA* and *rpoB* genes. Similarly, strain NRRL B-24084 was identified as a member of *S. stelliscabiei* (Fig. 6, clade 1) containing identical alleles for all of the housekeeping genes. Strain NRRL B-24079, which was phylogenetically the most distant member of clade 1, shared only the allele for the *rpoB* gene with the other two strains and therefore may represent a new species. Strain NRRL B-24083 was identified as a member of the species *S. europaeascabiei* (Fig. 6, clade 3), the alleles for the *atpD*, *rpoB* and *trpB* genes being identical to those of the type strain, NRRL B-24443<sup>T</sup>. The identity of NRRL B-24085 was unclear at the time of writing, therefore, phylogenetic analyses of the concatenated four-gene alignment of all *Streptomyces* type strains may be required in order to determine if this strain belongs to another species with a validly published name.

Ritacco & Eveleigh (2008) suggested that *Streptomyces ederensis* Wallhäuser et al. 1966 represents a later synonym of *Streptomyces phaeochromogenes* based on 16S rRNA gene sequence and partial *rpoB* gene sequence comparisons. However, the results of phylogenetic analyses of the concatenated four-gene dataset from the present study (see Fig. 6) and the observation that *S. phaeochromogenes* NRRL B-1248<sup>T</sup> and *S. ederensis* NRRL B-8146<sup>T</sup> did not contain identical alleles for any of the housekeeping genes sequenced support the continued recognition of these species as distinct. The 15 other species included in this study were also determined to be distinct based on the concatenated four-gene phylogenetic tree (Fig. 6) and novel allele contents of the housekeeping genes sequenced in the present study.

The published studies of Guo et al. (2008), Rong et al. (2009) and Rong & Huang (2010) demonstrated the value of multilocus sequencing and phylogenetic analyses for the taxonomy and identification of members of the genus *Streptomyces*. This study extended their work to phytopathogenic species, and those that are phylogenetically related to these species based on 16S rRNA gene sequences, within the genus *Streptomyces* and confirmed
Fig. 4. Phylogenetic tree calculated from partial sequences of the rpoB gene (540 bp) using the Tamura–Nei evolutionary distance method and the neighbour-joining algorithm. Bootstrap values >60% (based on 1000 resampled datasets) are given at nodes. Bar, 0.02 substitutions per nucleotide position.

Fig. 5. Phylogenetic tree calculated from partial sequences of the trpB gene (571 bp) using the Tamura–Nei evolutionary distance method and the neighbour-joining algorithm. Bootstrap values >60% (based on 1000 resampled datasets) are given at nodes. Bar, 0.02 substitutions per nucleotide position.
that these species represent multiple distinct taxa. The study also demonstrated the utility of multilocus sequence analysis for the rapid and accurate identification of unknown strains isolated from diseased plant material, without the need to perform DNA–DNA hybridization experiments for definitive confirmation of identity. The data in this study appear to suggest that the four gene loci sequenced and analysed represent a minimum set of gene sequences that can prove useful for the identification of Streptomyces isolates. In the present study, identical alleles in at least three of the four housekeeping genes appears to strongly support identity at the species-level; however, this does not discount the fact that sequencing fewer genes in unknown strains might also yield useful results.

**Acknowledgements**

The able technical assistance of E. N. Hoekstra in the isolation of genomic DNA and determination of the sequences of the various housekeeping genes is gratefully acknowledged. Names are necessary
to report factually on available data, however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by USDA implies no approval of the product to the exclusion of others that may also be suitable.

References


