## PS5-1 The *Sinorhizobium medicae* WSM419 genome sequencing project

**Wayne Reeve**[1], Patrick Chain[2], Ravi Tiwari[1], Graham O'Hara[1], John Howieson[1], and Lambert Bräu[1]

[1]Centre for Rhizobium Studies, Murdoch University, Murdoch, WA 6150, Australia; [2]Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

*Sinorhizobium medicae* is capable of fixing nitrogen with *Medicago arabica*, *M. murex*, *M. polymorpha*, *M. truncatula*, and *M. sativa*, the last two of which are also hosts for *Sinorhizobium meliloti* Sm1021. *S. medicae* WSM419 is saprophytically competent in moderately acidic soils (>pH 4.9) that are challenging to other sinorhizobia; a feature that enabled pasture production to be extended in southern Australia by a further 1 million ha. We now report on the complete genome sequence of *S. medicae* WSM419. For the sequencing strategy, a shotgun assembly approach was adopted using four libraries; one of which was constructed in a functional genomics vector (pTH1522) [1]. Double-ended plasmid sequencing reactions were then performed at the US Joint Genome Institute. Approximately 92,100 sequencing reads were assembled, producing an average of 12.9-fold coverage across the genome. Processing of sequence traces, base calling, assessment of data quality and assembly were performed with the PHRED/PHRAP/CONSED package. The initial assembly consisted of 30 contigs with at least 20 reads per contig. Gaps in the sequence were closed by primer walking on gap-spanning library clones or genomic DNA-amplified PCR products. Sequence finishing and polishing added 638 reads. Automated gene prediction was completed by assessing congruence of gene call results from Critica, Generation, and Glimmer, and by comparing the translations to GenBank's nonredundant database. Analysis of the genome (6 817 576 bp) reveals a multipartite structure consisting of a chromosome (3 781 904 bp) and three plasmids (pSMed01, 1 570 951 bp; pSmed02, 1 245 408 bp; and pSMed03, 219 313 bp) with a GC content 61.15%. In total, 6523 protein encoding ORFs could be identified of which 4646 (70.53%) could be assigned a putative function. This presentation will give an update on the current state of comparative analyses and future research directions.

This project has been made entirely possible by funding from the Joint Genome Institute Community Sequencing Program (Department of Energy; USA), the Australian Research Council, and Murdoch University.

[1] Cowie *et al*. (2006). *Appl. Environ. Microbiol*. 72:7156-7167.