11 June 2012

This document describes the data used for:

Epstein B, Branca A, Mudge J, Arvind KB, Briskine R, Farmer A, Sugawara M, Young ND, Sadowsky MJ, Tiffin P. 2012. Population genomics of the facultatively mutualistic bacteria *Sinorhizobium meliloti* and *S. medicae*. PLoS Genetics (in press).

SNP and other variant data

32 *S. meliloti* and 12 *S. medicae* strains were sequenced using Illumina paired end reads (SRP009881) and aligned to the reference genomes: either *S. meliloti* 1021 or *S. medicae* WSM419. A base was called as different from the reference genome if it was covered by ≥ 10 unique reads, ≥ 70% of the reads at the site called that base, and there were < 500 unique reads covering that site. Sites that met the same coverage criteria, but with ≤ 30% of the reads calling a variant were assumed to be the same as the reference genome. Sites that did not meet these requirements were treated as ambiguous ("N").

SNP Files

The SNP files contain a list of differences between the reference genomes and the resequenced strains.

There are 2 files - one for each species. For each position at which a SNP was called there is a row in the file. The columns for the SNP files are:

Position	position of the SNP (in the reference genome).
Replicon	replicon the SNP was called on (Chr, pSymA, pSymB, pSmed01, pSmed02)

An additional column for each strain, labeled by the name of the strain.