Supplemental methods

Bacterial and plant material

M. extorquens strain DSM13060 [Genbank:AGJK00000000] was originally isolated from the buds of mature *P. sylvestris* trees growing in a natural stand in Oulu, Northern Finland (65°0´ N; 25°30´ E) (1). For the gDNA isolation and inoculant preparation, bacteria were grown by shaking at 28°C in M9 Minimal Salts Medium supplemented with 18.5 mM sodium succinate and 120 mM methanol (2). Pine seeds were heat-treated for 72h at +55°C, incubated in sterile water overnight at RT, and surface sterilized with 2.5% calcium hypochlorite for 15 min. After sterilization, the seeds were rinsed three times with sterile water and germinated on soaked sterile vermiculite in glass jars. The seeds were allowed to germinate for 5 days in the growth chamber at 24± 3°C at 16/8 h photoperiod. The GFP-tagged *M. extorquens* DSM13060 strain (13061) was used for the colonization studies (3). The bacterial culture was diluted with sterile water to the density of 2.5 x 10^7 CFU/mL and 100 µL of solution was pipetted on each germinating pine seed for the inoculation.

Isolation of genomic DNA

Bacteria were grown to the late logarithmic phase and cells were harvested by centrifugation at 4000 x g for 5 min at 4°C. The bacterial pellet was ground in liquid nitrogen with mortar and pestle to facilitate the lysis and increase the yield. The genomic DNA was isolated according to the Joint Genome Institute (JGI) standard protocols for bacterial DNA isolation by using CTAB (4) with minor modifications. Two additional

extraction steps with chloroform:isoamyl alcohol and phenol:chloroform:isoamyl alcohol were done to increase the DNA quality. The quantity and quality of DNA was assessed against molecular weight standards to meet the JGIs specifications in 1% (w/v) agarose gel electrophoresis in 1 x Tris-acetate-EDTA (TAE) buffer containing ethidium bromide (0.5 μg/ml).

Genome sequencing

The draft genome of *M. extorquens* DSM13060 was generated at the JGI using a combination of Illumina (5) and 454 technologies (6). For this, genome we constructed and sequenced an Illumina GAii shotgun library which generated 65,149,632 reads totaling 4,951.4 Mb and a paired end 454 library with an average insert size of 15 kb which generated 505,906 reads totaling 113.9 Mb of 454 data. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/. The initial draft assembly contained 734 contigs in 5 scaffolds. The 454 paired end data were assembled together with Newbler, version 2.6 (7). The Newbler consensus sequences were computationally shredded into 2 kb overlapping fake reads (shreds). Illumina sequencing data was assembled with VELVET, version 1.1.05 (8), and the consensus sequence were computationally shredded into 1.5 kb overlapping fake reads (shreds). We integrated the 454 Newbler consensus shreds, the Illumina VELVET consensus shreds and the read pairs in the 454 paired end library using parallel phrap, version SPS - 4.24 (High Performance Software, LLC). The software Consed (9) was used in the following finishing process. Illumina data was used to correct potential base errors

and increase consensus quality using the software Polisher developed at JGI (Alla Lapidus, unpublished). Possible mis-assemblies were corrected using gapResolution (Cliff Han, unpublished), Dupfinisher (10) (C. Han and P. Chain, International Conference on Bioinformatics & Computational Biology. CSREA Press, Las Vegas, Nevada, USA, 2006), or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR (J-F Cheng, unpublished) primer walks. A total of 725 reactions were used to close additional gaps. The final draft DSM 13060 genome consists of 12 scaffolds (240 contigs) totaling ~6.7 Mb.

Genome annotation

Automatic annotation was performed using the JGI-Oak Ridge National Laboratory annotation pipeline. Additional automatic and manual sequence annotations and comparative genome analysis was performed using the Integrated Microbial Genomes system (11), as well as in-house Perl/BioPerl/Perl Tk scripts. If we assume co-linearity with the genome of strain AM1, the five largest scaffolds together add up 5.51 Mb and align to the 5.51 Mb AM1 chromosome. The remaining large scaffold (1194411 bp) were very similar in size and gene content to the AM1 megaplasmid, and is therefore likely to be a DSM13060 megaplasmid. Apart from the analyses concerning the megaplasmid our genome comparisons do not rely on synteny. The remaining 6 contigs were between 5-16 kb long, and did not encode AM1 orthologs. No small plasmids could be identified. The whole genome shotgun sequences have been deposited in GenBank under accession number AGJK00000000.

Gene origin inference

Blastp was used to compare the The DSM 13060 proteome to the nr database downloaded from Genbank on July 10, 2013 (nr blast search; only hits with E<e-10 were included). For the top 10 hits of each protein, a downloaded NCBI taxonomy database was used to count the number of hits in each of the following taxonomic categories: non-extorquens *Methylobacterium* (m), non-*Methylobacterium Rhizobiales* (r), non-*Rhizobiales* Alphaproteobacteria, and non-Alphaproteobacteria (n), excluding hits classified as "unclassified sequences", and "other sequences". Next, Blastp was used to compare the The DSM 13060 proteome to and to a local database of the proteomes of *M. extorquens* PA1, *M. extorquens* BJ001, *M. extorquens* CM4, *M. extorquens* AM1, and *M. extorquens* DM4 (extorquens blast search; only hits with percent identity >80 and E<e-30 were included). The number of hits in each of the following taxonomic categories were counted: any *M. extorquens* strain other than DSM 13060 (e), *M. extorquens* AM1 (am1). The criteria used to identify potentially horizontally transferred genes and other genes of interest in the DSM13060 genome are showed in Table S1 in the supplemental material, along with the numbers of genes in the different categories and their distribution across the two replicons. The categories are non-overlapping, e.g. 2a does not include genes in 1a. The remaining 3518 genes (not found in any of these categories) are presumed to be part of an Alphaproteobacterial core genome.

Utilization of 1-aminocyclopropane-1-carboxylic acid (ACC) as a sole nitrogen source

Enzymatic activity of putative ACC deaminase of *M. extorquens* DSM13060 was assayed according to Poonguzhali et al. (12). Briefly, *M. extorquens* DSM13060 was streaked onto ammonium mineral salts agar media (AMS) supplemented with 3 mM ACC (Sigma-Aldrich, St. Louis, Missouri, USA) instead of NH_4Cl as a sole nitrogen source. Plates were incubated at +28 °C for 10 days and bacterial growth was compared to a *M. extorquens* DSM13060 control grown on AMS media without any nitrogen source.

Screening of cellulose hydrolyzing enzymes

Ability of *M. extorquens* DSM13060 to hydrolyse cellulose was assayed according to Kasana et al. (13). *M. extorquens* DSM13060 culture was grown under shaking in M9 minimal media at +28 °C for 48h. Ten microlitres of exponential phase growth culture was pipetted to the center of carboxymethyl cellulose (CMC) agar plates. The plates were incubated at +28 °C for 10 days. Development of halo zone around the colony was considered as positive for production of cellulose hydrolyzing enzymes.

Promoter-reporter constructs

In order to study the activity of putative ACC deaminase gene (*acdS*) (IMG gene accession 2507325054), cobalamin synthase (*cobS*) (IMG gene accession 2507327870) and bacteriophytochrome (*bphP*) (IMG gene accession 2507328599), reporter constructs were created for *M. extorquens* DSM13060. The promoter regions of the genes were amplified from genomic DNA of *M. extorquens* DSM13060 using Phusion High-Fidelity DNA

polymerase (Thermo Scientific) and primers with restrictions sites for *Mlu*I and *Nde*I enzymes (FastDigest; Thermo Scientific) included in the overhangs, respectively. Similarly, the *mCherry* gene was amplified from the plasmid pMP7604 with restrictions sites for the *Nde*I and *Kpn*I enzymes (14, 15). The *Nde*I restriction enzyme was used to create a compatible restriction site between each promoter (3’) and the mCherry gene (5’). After restriction of promoter-reporter fragments, the ligated constructs were cloned into pJET 1.2 plasmid and propagated in *E.coli* DH5α (16). Plasmids were digested with *Mlu*I and *Kpn*I restriction enzymes, and the constructs were ligated into the respective sites of pME6031 plasmid (17), to create the plasmids pMExt054, pMExt870 and pMExt599 carrying the constructs *acdS::mCherry, cobs::mCherry and bphP::mCherry*, respectively. The primer sequences, plasmids and bacterial strains used in this study are described in Table S2.

Alphaproteobacterial pfams excluded from the pfam analysis

pfam00115, pfam00033, pfam02790, pfam01059, pfam00361, pfam00361, pfam00146, pfam00662, pfam00119, pfam00507, pfam00116, pfam00499, pfam00510, pfam00163, pfam13631, pfam00032

References for Supplemental methods

1. **Pirttila, A. M., H. Laukkanen, H. Pospiech, R. Myllyla, and A. Hohtola.** 2000. Detection of intracellular bacteria in the buds of Scotch pine (*Pinus sylvestris* L.) by in situ hybridization. Appl. Environ. Microbiol. **66:**3073-3077.

2. **Sambrook, J., and D.W. Russell.** 2001. Molecular Cloning: A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

3. **Pohjanen, J., J. J. Koskimaki, S. Sutela, P. Ardanov, M. Suorsa, K. Niemi, T. Sarjala, H. Haggman, and A. M. Pirttila.** 2014. Interaction with ectomycorrhizal fungi and endophytic *Methylobacterium* affects nutrient uptake and growth of pine seedlings in vitro. Tree Physiol. **34:**993-1005.

4. **Wilson, K.** 2001. Preparation of genomic DNA from bacteria. Curr. Protoc. Mol. Biol. **Chapter 2:**Unit 2.4.

5. **Bennett, S.** 2004. Solexa ltd. Pharmacogenomics. **5:**433-438.

6. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. Chen, and Z. Chen.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature. **437:**376-380.

7. **Miller, J. R., S. Koren, and G. Sutton.** 2010. Assembly algorithms for next-generation sequencing data. Genomics. **95:**315-327.

8. **Zerbino, D. R., and E. Birney.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18:**821-829.

9. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186-194.

10. **Han, C., and P. Chain.** 2006. Finishing repetitive regions automatically with Dupfinisher, p. 142-147. *In* H. R. Arabnia and H. Valafar (ed.), Proceedings of the 2006 International Conference on Bioinformatics & Computational Biology. CSREA Press, Las Vegas, NV.

11. **Markowitz, V. M., I. M. Chen, K. Chu, E. Szeto, K. Palaniappan, M. Pillay, A. Ratner, J. Huang, I. Pagani, S. Tringe, M. Huntemann, K. Billis, N. Varghese, K. Tennessen, K. Mavromatis, A. Pati, N. N. Ivanova, and N. C. Kyrpides.** 2014. IMG/M 4 version of the integrated metagenome comparative analysis system. Nucleic Acids Res. **42:**D568-73.

12. **Poonguzhali, S., M. Madhaiyan, and T. Sa.** 2006. Cultivation-dependent characterization of rhizobacterial communities from field grown Chinese cabbage *Brassica campestris* ssp *pekinensis* and screening of traits for potential plant growth promotion. Plant Soil. **286:**167-180.

13. **Kasana, R. C., R. Salwan, H. Dhar, S. Dutt, and A. Gulati.** 2008. A rapid and easy method for the detection of microbial cellulases on agar plates using Gram's iodine. Curr. Microbiol. **57:**503-507.

14. **Shaner, N. C., R. E. Campbell, P. A. Steinbach, B. N. Giepmans, A. E. Palmer, and R. Y. Tsien.** 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. Nat. Biotechnol. **22:**1567-1572.

15. **Lagendijk, E. L., S. Validov, G. E. Lamers, S. De Weert, and G. V. Bloemberg.** 2010. Genetic tools for tagging Gram-negative bacteria with mCherry for visualization in vitro and in natural habitats, biofilm and pathogenicity studies. FEMS Microbiol. Lett. **305:**81-90.

16. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41:**459-472.

17. **Heeb, S., Y. Itoh, T. Nishijyo, U. Schnider, C. Keel, J. Wade, U. Walsh, F. O'Gara, and D. Haas.** 2000. Small, stable shuttle vectors based on the minimal pVS1 replicon for use in gram-negative, plant-associated bacteria. Mol. Plant-Microbe Interact. **13:**232-237.