A single *sym* plasmid type predominates across diverse chromosomal lineages of
Cupriavidus nodule symbionts

Matthew A. Parker

Department of Biological Sciences, State University of New York, Binghamton, NY 13902 USA

author for correspondence:

Matthew A. Parker, Department of Biological Sciences, State University of New York,
Binghamton, NY 13902 USA

phone: 607-777-6283

fax: 607-777-6521

e-mail: mparker@binghamton.edu

running title: *Cupriavidus sym* plasmid evolution

Abstract

Cupriavidus nodule symbionts from *Mimosa* host legumes indigenous to five locations around the Caribbean region were analyzed by sequencing portions of five chromosomal housekeeping loci and five *sym* plasmid loci in 80 isolates. Nodule symbionts did not form a single clade separated from non-symbiotic reference strains of *Cupriavidus* and *Ralstonia*, implying that either convergent losses or independent gains of the trait of legume symbiosis have taken place. Chromosomal genes exhibited significantly higher nucleotide polymorphism and haplotype diversity than *sym* plasmid loci. A single derived *sym* plasmid haplotype (A1) was found to predominate in four of the populations, and was shared by multiple housekeeping gene clades. This suggests that one *sym* plasmid variant has recently spread geographically and has been acquired by diverse chromosomal lineages within the region. Inoculation of two *Mimosa* host species indicated that strains carrying the predominant A1 haplotype ranked either first or second among the five major *sym* plasmid haplotype groups with respect to plant growth enhancement. Symbiotic outcomes also varied greatly among chromosomally diverse strains that all shared the A1 haplotype. Thus, chromosomal as well as *sym* plasmid variants likely contribute to differential interactions with *Mimosa* host species.

key words: β -rhizobia, horizontal gene transfer, phylogeography, symbiotic specificity

1 Introduction

2 Two genera of Betaproteobacteria, *Burkholderia* and *Cupriavidus*, are currently known to
 3 include strains that are nitrogen-fixing symbionts of legume plants [24]. For *Burkholderia*,
 4 considerable progress has been made in understanding diversity, biogeography and host plant
 5 relationships [7,8,10,11,12,16,17,18,19,22,24,25]. However, much less is currently known about
 6 the distribution and relationships of symbiotic *Cupriavidus* bacteria. The natural hosts of
 7 *Cupriavidus* nodule symbionts found to date have been species of *Mimosa* [2,5] and the related
 8 genus *Parapiptadenia* [37] indigenous to the Americas. *Cupriavidus* symbionts have also been
 9 found in many parts of Asia associated with invasive weedy species of *Mimosa* [9,10,23,28,30],
 10 but available evidence suggests that these originated from the ancestral range of their host
 11 legumes in tropical America [3].

12 Based on nodules taken from legumes growing *in situ* in their native geographic range,
 13 *Cupriavidus* symbionts have been found in only four regions: Costa Rica [5], south Texas [2],
 14 Puerto Rico [3], and Uruguay [37]. Soil inoculation experiments using a trap host (*M. pudica*)
 15 that was naturally present at the sample site also detected *Cupriavidus* in French Guiana [32].
 16 Soil inoculation using two promiscuous legumes (*Phaseolus vulgaris*, *Leucaena leucocephala*)
 17 also recovered multiple *Cupriavidus* isolates from Minas Gerais, Brazil [15]. Sampling of
 18 diverse *Mimosa* species and related genera have failed to detect *Cupriavidus* in several regions of
 19 Brazil [7,8,17]. However, the Minas Gerais *Cupriavidus* isolates can effectively nodulate
 20 *Mimosa* [15], and possess *nodC* and *nifH* genes with affinities to *Cupriavidus* strain UYPR2.512
 21 from *Parapiptadenia* in Uruguay [37]. The results of da Silva *et al.* [15] therefore suggest that
 22 mimosoid legumes harboring *Cupriavidus* are likely to occur in Brazil.

23 Since prior surveys of *Cupriavidus* have been based on few loci, expanded sampling of

both chromosomal housekeeping loci and genes on the symbiotic (*sym*) plasmid [1] is needed to understand the diversity and relationships of *Cupriavidus* populations in the Americas. In particular, three recent analyses of *Burkholderia* symbionts from South America have found that tree topologies were strongly congruent for *sym* plasmid loci and housekeeping loci, suggesting a long-term preponderance of vertical inheritance for symbiotic genes [7,8,32]. The first goal of this study was to analyze whether *Cupriavidus sym* plasmid inheritance was also primarily vertical. The second main goal was to understand how housekeeping and *sym* plasmid variants may contribute to differential interactions with host legumes. To address this issue, I performed inoculation experiments with two of the *Mimosa* species from which isolates were sampled.

Materials and methods

Isolate sampling

The study expands on an earlier project where 32 American *Cupriavidus* strains were characterized at five loci [3]. I analyzed additional isolates from four sites sampled previously (south Texas, northeastern Puerto Rico, and two sites 60 km apart in Costa Rica (Manuel Antonio [MA] and Tres Piedras [TP])). I also added a fifth sampling location on the Caribbean island of Guadeloupe (Table S1). *Mimosa asperata* was the only *Cupriavidus* host present in the Texas site [2], and this plant did not occur at any of the other sampling locations. *M. pudica* was the main host legume sampled in the other four sites, although a small number of isolates were also obtained from *M. pigra* and *M. diplotricha* (Table S1). Nodules were surface disinfected in 100% ethanol and 3.2% sodium hypochlorite, and a single isolate was purified from each nodule. A total of 80 isolates was analyzed (16 per population).

DNA sequencing and phylogenetic analyses

Genomic DNA was obtained by heating cells at 95⁰ C for 5 minutes in a lysis buffer containing 1% Triton X-100, followed by chloroform extraction. Portions of five *sym* plasmid loci (*nifH* [473 bp], *nodA* [488 bp], *nodC* [797 bp], *fixA* [676 bp], *phaZ* [encoding poly(3-hydroxybutyrate) depolymerase; 524 bp]) and five chromosomal (housekeeping) loci (*recA* [701 bp], *rpoB* [661 bp], 23S rRNA [390 bp], *gyrB* [730 bp], *glnD* [encoding uridylyltransferase; 742 bp]) were sequenced (see Table S2 for primers). For *phaZ*, three isolates failed to yield any PCR product when tested with two alternative forward primers and three alternative reverse primers in multiple combinations, suggesting an apparent gene deletion. Apart from *phaZ* in these three isolates, a complete ten-locus data set was obtained for the 80 *Cupriavidus* isolates.

DnaSP v5 [29] was used to calculate nucleotide diversity (the mean number of nucleotide differences per site) and haplotype diversity (the probability that two randomly chosen strains will carry different alleles). Separate phylogenetic analyses were performed on four concatenated housekeeping and four concatenated *sym* plasmid loci. For the *sym* plasmid tree analysis, the *phaZ* locus was excluded because homologs could not be identified in some *Burkholderia* strains. For the housekeeping loci, draft genome sequences for some outgroups lacked 23S rRNA sequence data, so this locus was also excluded from the housekeeping tree analysis. Variation in *phaZ* was highly congruent with other *sym* plasmid loci (see Results), and similar congruence was evident for 23S rRNA and other housekeeping loci. Consequently, exclusion of these two loci had little or no effect on tree topologies or branch support. For the four concatenated *sym* plasmid loci, the NeighborNet algorithm in SplitsTree [26] indicated that there was virtually no reticulation in the phylogenetic network. Thus, there appeared to be little phylogenetic uncertainty associated with recombination, justifying tree analysis on concatenated

data. Trees were inferred using MrBayes [35] with data partitioned by codon position and separate estimates of rates and nucleotide composition for each codon position and locus. A HKY substitution model was used. Some alternative models were tested but all yielded very similar trees. Analyses were run for 3 million generations, sampling every 400 generations, with the last 1000 samples saved for tree analysis. Replicate runs yielded completely identical trees.

For the four concatenated housekeeping loci, NeighborNet analysis revealed that significant reticulation was present in the phylogenetic network, especially with respect to placement of certain outgroup strains in relation to the 80-isolate *Cupriavidus* sample. I therefore used ClonalFrame 1.1 [14] to infer relationships in the presence of recombination events that may alter genealogy in different portions of the alignment. Default settings were used (50,000 burnin generations followed by tree sampling every 100 generations for another 50,000 generations). In practice, the trees inferred by ClonalFrame were virtually the same as the Bayesian tree for these data, although there were minor differences in the confidence assigned to some clades.

For the housekeeping gene analysis, all nodule-symbiotic *Cupriavidus* strains with completed or draft genomes were included as reference strains: *C. taiwanensis* LMG19424 (from Taiwan, NC_010528), *C. taiwanensis* STM 6018 (French Guiana, AXAK00000000), *C. taiwanensis* STM 6070 (New Caledonia, AQUR00000000), and *Cupriavidus* sp. UYPR2.512 (Uruguay, ARBE00000000). Eight other strains of *Cupriavidus* and the closely related genus *Ralstonia* with completed or draft genome sequences were also included: *C. necator* N-1 (NC_015726), *C. necator* A5-1 (JQLI01000001), *C. metallidurans* CH34 (NC_007973), *Cupriavidus* sp. SK-4 (JFJW01000001), *R. eutropha* H16 (NC_008313), *R. eutropha* JMP134 (CP000090), *R. pickettii* DTP0602 (NC_022513), and *R. solanacearum* FQY4 (NC_020799;

used as the outgroup). For the *sym* plasmid tree analysis, prior work has indicated that *Cupriavidus* symbiotic genes are apparently derived from nodule-symbiotic *Burkholderia* [2,10]. Consequently, reference strains for this analysis included several symbiotic *Burkholderia* strains with completed or draft genomes in addition to the four nodule-symbiotic *Cupriavidus* strains noted above: *B. phymatum* STM 815^T (French Guiana, NC_010622), *B. tuberum* STM 678^T (South Africa, JGI 2512047030), *Burkholderia* sp. CCGE1002 (Mexico, CP002016), *B. symbiotica* JPY366 (Brazil, AZXJ000000000), *B. mimosarum* LMG23256^T (Taiwan, AXAN000000000), *B. sprentiae* WSM5005^T (South Africa, AXBN000000000), and *B. nodosa* DSM21604^T (Brazil, JAF000000000). *Bradyrhizobium diazoefficiens* USDA110^T (NC_004463) was used as the outgroup for this analysis.

Symbiotic phenotypes

Seeds of *M. pudica* (Stokes Seed Company) and of *M. pigra* var. *pigra* (from garden-grown plants descended from a Florida population [13]) were scarified and surface-sterilized in concentrated sulfuric acid and then planted in a mixture of potting soil and sand known to be free of nodule bacteria. The variety represented by the *M. pudica* seeds is not known. Three *M. pudica* varieties recognized in Barneby [4] are diagnosed by floral traits, but experiments were terminated before any flowering took place. Germinated seedlings were transferred to 240 cm³ pots and inoculated with approximately 10⁹ cells of a particular isolate grown in yeast-mannitol broth (n=11 seedlings of each *Mimosa* species per isolate). Eleven uninoculated control plants were also grown simultaneously along with the inoculated plants, and precautions were taken to avoid contamination [40]. Control plants were found to be completely free of nodulation, suggesting that bacterial contamination was negligible. Plants were fertilized with a nitrogen-

free nutrient solution every 10 days. Plant positions were re-randomized about the room weekly. At harvest, 42 days after inoculation, total dry biomass was recorded.

Two inoculation experiments were performed. In the first experiment, plants were inoculated with two *Cupriavidus* isolates from each of the five main lineages detected in phylogenetic analysis of *sym* plasmid loci (see Results). In a second experiment, plants were inoculated with eleven isolates that all had the same *sym* plasmid type, but which belonged to diverse chromosomal (housekeeping gene) lineages. To enable comparison of the two *Mimosa* host species, analysis of variance was done on relative biomass (log-transformed to improve homoscedasticity). Let C_{pud} represent the mean mass of uninoculated *M. pudica* plants, and X_i be the mass of an individual *M. pudica* plant inoculated with a particular bacterial strain. The response variable analyzed was $(X_i - C_{\text{pud}})/C_{\text{pud}}$. Biomass of *M. pigra* plants was scaled similarly. On this scale, inoculated plants of either species that grew twice as much as uninoculated controls would have a relative biomass of 100%. In the first experiment, plant species and *Cupriavidus* symbiotic lineage were both considered to be fixed effects, and variation among replicate isolates within a symbiotic lineage was considered a nested random effect. In the second experiment, all bacterial isolates shared the same symbiotic lineage, resulting a simple 2-way mixed model ANOVA (plant species[fixed] x isolate[random]).

GenBank accession numbers for the 637 new sequence records obtained in this study, and for the 160 other existing sequence records utilized, are provided in Table S3.

Results and Discussion

Nucleotide diversity

Cupriavidus isolates had more than two-fold higher nucleotide diversity at chromosomal

housekeeping loci compared to *sym* plasmid loci (Table 1). Nucleotide diversity values did not overlap for the two groups, and a Mann-Whitney test indicated that the difference between housekeeping and *sym* plasmid loci was significant ($P = 0.012$). Haplotype diversity was also twice as high for housekeeping genes relative to *sym* plasmid loci (Table 1). These patterns suggest that the *Cupriavidus sym* plasmids have been subject to diversity-purging forces (such as a selective sweep [21]) that have not similarly affected chromosomal housekeeping genes.

The prevalence of private alleles (single-nucleotide variants [SNVs] found in only a single population) provided further evidence that housekeeping and *sym* plasmid loci have been affected by disparate forces of evolution. In the absence of selection, genetic drift and migration will affect all loci in the same way, and private allele frequency is expected to show a simple inverse relationship to the number of migrants entering a population per generation [36]. In the *Cupriavidus* data, there were more than twice as many private SNVs for *sym* plasmid loci even though the total number of mutations was far higher for housekeeping genes (Table 2). Allele frequency for private SNVs was also significantly higher for *sym* plasmid loci. These outcomes could be generated by spatially heterogeneous selection (e.g., if a certain *sym* plasmid variant adapted to one environment was selected against upon migration to a different population). The data suggest that selection on one or more *sym* plasmid loci may have resulted in hitchhiking effects throughout the *sym* plasmid.

sym plasmid relationships

All *Cupriavidus sym* plasmid sequence variants clustered into a strongly supported clade that formed a sister group to *Burkholderia* strains from Central and South America (Fig. 1; note that *B. mimosarum* strain LMG23256 isolated in Taiwan is likely of South American origin [12],

because its host legume (*M. pigra*) is an invasive weed from the Neotropics). The two *Burkholderia* taxa from South Africa (*B. tuberum* STM 678, *B. sprentiae* WSM5005) formed a lineage that was well separated from the remaining strains. The 80 *Cupriavidus* isolates formed a monophyletic group with five distinct sub-clades (A, B, C, D, E). The relatively short branches within this group suggest that divergence among these five sub-clades happened recently relative to lineage splitting events in other parts of the tree. All branches in the tree were highly supported (posterior probability [PP] = 1.0), except for a single short branch uniting *Cupriavidus* group C together with A and D (PP=0.81). With respect to host origin, group A strains came from three *Mimosa* species (*M. pudica*, *M. pigra*, *M. diplotricha*), and each of the other groups was sampled from just a single species (B: *M. pudica*, C and D: *M. asperata*, E: *M. pigra*).

Isolates in three of the groups displayed single-nucleotide variants (designated by a number following the letter). One sequence variant (type A1) predominated in four of the five *Cupriavidus* populations, and occurred in 59% of the isolates. This variant was also found in the *C. taiwanensis* type strain LMG19424 [1] as well as *C. taiwanensis* STM 6018 and STM 6070. Variants A2(3 isolates) and A3(3 isolates) each had a distinct nucleotide substitution in the *nodA* locus, and variant A4(5 isolates) had a single substitution in *nodC*. *Cupriavidus* samples from Texas shared no haplotypes in common with any of the other locations. Two closely related Texas haplotypes, D1 and D2, differed by one *nifH* nucleotide substitution. The other two Texas haplotypes (C1 and C2) differed by one *nodC* substitution. The remaining two sequence types (B, E) were each restricted to a single Caribbean location.

Sequence variation at one additional *sym* plasmid locus (*phaZ*) closely matched the divergence pattern in Fig. 1. Within each of the five groups (A, B, C, D, E), all isolates shared a single *phaZ* sequence type, and the structure of divergence among groups was broadly congruent

with the tree for other *sym* plasmid loci (Fig. S1).

Divergence at chromosomal housekeeping loci

ClonalFrame analysis partitioned the 80 *Cupriavidus* isolates into seven highly supported clades (Fig. 2; see Fig. S2 for the tree inferred by Bayesian analysis, which matched Fig. 2 in almost all respects). This analysis inferred that some gene segments had recombined in certain groups. For example, the group of four strains that included *Cupriavidus* UYPR2.512 was most similar to neighboring groups (clade 1, *R. pickettii* DTP0602, *C. necator* A5-1) for three genes, but had a portion of the *glnD* locus that was more similar to clade 3. However, the estimated ratio of probabilities that an individual nucleotide had been altered by recombination vs. by point mutation (*r/m*) was only 0.5 (95% credibility region, 0.3 - 0.8), which would place these strains in the lowest quartile of the distribution of *r/m* values found among bacteria and archaea [39].

Divergent *Cupriavidus* clades co-occurred within all five sampling locations. The fact that all seven *Cupriavidus* clades were distributed across two or more regions implies that considerable migration has taken place during the diversification of these bacteria.

The position of three non-symbiotic outgroup taxa on the tree suggested an early divergence of the remaining strains into two large lineages, only one of which was exclusively composed of symbiotic strains (clades 2-7). *Cupriavidus* clade 1 belonged to a strongly supported group with the South American *Cupriavidus* symbiotic strain UYPR2.512 and five non-symbiotic strains: *Ralstonia pickettii* DTP0602 (from Japan [33]), *C. necator* N-1 and A5-1 (from Pennsylvania and Washington State, USA [6,38]), *Cupriavidus* sp. SK-4 (from Indiana, USA [27]), and *R. eutropha* H16 (from Germany [34]). Genome BLAST searches indicated that none of these non-symbiotic strains had any of the *sym* plasmid genes characterized in the other

isolates (*nodA*, *nodC*, *nifH*, *fixA*, *phaZ*). The fact that symbiotic *Cupriavidus* strains did not form an exclusive monophyletic lineage implies that either convergent gains or convergent losses of legume nodule symbiosis have occurred. If acquisition of legume symbiosis was a singular event, then Fig. 2 implies that the five non-symbiotic strains related to clade 1 are all descended from symbiotic ancestors. Loss of plasmid-encoded symbiotic genes would not be an improbable occurrence, although the fact that these strains are ecologically quite different, and were found on three separate continents, would suggest a long time scale for their diversification.

Alternatively, the affinities of *Cupriavidus* clade 1 and UYPR2.512 to various non-symbiotic *Cupriavidus* strains, and their strong divergence from other symbiotic *Cupriavidus* (Fig. 2) raise the possibility that these groups gained the trait of legume nodule symbiosis independently from the main lineage of symbiotic *Cupriavidus* (clades 2-7). If the gain of symbiotic genes in *Cupriavidus* was a singular event with strict vertical transmission subsequently, then the *sym* gene tree should parallel the housekeeping gene tree topology. This was not the case. *Sym* genes from UYPR2.512 and *Cupriavidus* clade 1 (types C1, C2, E) did not form a sister group to the *sym* genes of all remaining strains (Fig. 1). Further sampling of both symbiotic and non-symbiotic *Cupriavidus* lineages will be important to resolve whether the trait of legume symbiosis may have been gained more than once in separate *Cupriavidus* clades.

Regardless of how this issue is resolved, the distribution of symbiotic gene variants across housekeeping-gene clades (Fig. 2) clearly indicated that there has been a complex history of *sym* plasmid transfer. For the 80-isolate *Cupriavidus* sample, the earliest diverging haplotype in the *sym* gene tree (type B; Fig. 1) occurred in two distinct clades (5, 7) that were not basally positioned in the housekeeping gene tree. Strains that were closely related for chromosomal markers sometimes carried divergent *sym* plasmid haplotypes (e.g., various strains in clades 1, 5,

6, 7). A single common *sym* gene haplotype (A1) occurred in six of the seven chromosomal clades. Thus, numerous *Cupriavidus* strains that were divergent for other genetic markers all shared a uniform plasmid type. The fact that type A1 occurred in a derived position on the *sym* gene tree (Fig. 1) suggests that this is a recently-evolved *sym* plasmid variant that has spread across multiple chromosomal lineages and geographic regions.

Symbiotic phenotypes

Growth varied greatly when plants were exposed to *Cupriavidus* strains having different *sym* plasmid haplotypes (Fig. 3). All ten bacterial isolates developed nodules on both hosts, although there was more than a two-fold range in mean nodule numbers per plant (Table S4). Yet despite forming nodules, strains with one symbiotic gene variant (type C1) provided no growth benefit to either of the host species. For the other four *Cupriavidus sym* plasmid types, relative benefit to the two *Mimosa* species was highly unequal. Growth of *Mimosa pudica* hosts was significantly boosted by bacterial isolates harboring all variant types except C1. By contrast, *Mimosa pigra* plants only experienced enhanced growth with two *sym* plasmid types (A1, E). This is consistent with several prior studies indicating that *M. pudica* associates with *Cupriavidus* symbionts more frequently than does *M. pigra* [5,10,11], with *M. pigra* apparently being more adapted to symbiosis with strains of *Burkholderia* [19,20].

The most beneficial *Cupriavidus sym* plasmid type differed for *M. pudica* and *M. pigra* (A1 vs. E, respectively). Thus, the local abundance of strains carrying different *sym* plasmids may be a factor affecting ecological success of these *Mimosa* species within particular habitats. For *Cupriavidus* strains with *sym* plasmid types C1 and D1, no seeds of their original host (*M. asperata*) were available to enable testing of symbiotic phenotypes. Thus, it remains an open

question whether the very low symbiotic benefits conferred by types C1 and D1 on *M. pigra* and *M. pudica* would also be evident when these strains nodulated their original *M. asperata* hosts.

Melkonian *et al.* [31] observed that nodulation competitiveness of *C. taiwanensis* relative to *Burkholderia phymatum* differed across *M. pudica* varieties or geographic populations. This suggests that other aspects of symbiotic interaction, such as patterns of differential plant growth with *Cupriavidus* lineages (Fig. 3), may also show intraspecific variability in *M. pudica*. An important problem for future research will be characterize patterns of genetic variation in *M. pudica* symbiotic phenotypes, within and across populations, and how this affects the utilization of *Cupriavidus* lineages in natural environments.

Cupriavidus sym plasmid haplotype groups were nonrandomly distributed across chromosomal lineages (Fig. 2). Thus, it is not possible to attribute these different phenotypes to symbiotic plasmid variation *per se*. The strains may have differed at chromosomal loci as well as *sym* plasmid loci affecting symbiotic performance. To study the effects of chromosomal background, eleven chromosomally diverse strains that all carried the same A1 *sym* gene haplotype were analyzed in a second inoculation experiment. These strains varied widely in symbiotic behavior toward two *Mimosa* host species (Fig. 4). There was a 4 to 6-fold range among strains in plant biomass gain, and a 2.5-fold range in relative benefit to the two *Mimosa* hosts. However, no clear relationship was apparent between chromosomal lineage and symbiotic phenotype. For example, separate strains from both clades 2 and 4 (Fig. 2) were among the best and also among the worst symbionts for the two host species. More broadly, if symbiotic benefit is considered as a character to be mapped on the housekeeping gene tree (Fig. 2), there does not appear to be any phylogenetic pattern involving a change toward either increased or decreased benefit across the *Cupriavidus* tree as a whole.

While Figs. 3 and 4 suggest that both chromosomal and *sym* plasmid loci may contribute to differential strain interactions with particular *Mimosa* host species, there is an important caveat. The A1 *sym* plasmid type was diagnosed by sequencing only five marker loci (2958 bp) out of the entire plasmid (557200 bp in *C. taiwanensis* LMG 19424 [1]). Thus, variation in symbiotic outcomes in the A1 inoculation experiment (Fig. 4) might be due to sequence heterogeneity within the A1 type at other *sym* plasmid loci. Plasmid transfer experiments would be needed to resolve this issue.

Conclusions

For symbiotic *Cupriavidus* strains from the Caribbean region, I have documented that *sym* plasmids are monophyletic and are differentiated into a limited variety of groups (A, B, C, D, E), which are non-randomly distributed across diverse chromosomal lineages. Substantial incongruence between the trees for *sym* loci and housekeeping loci, and the finding that one common *sym* plasmid type (A1) was shared by multiple chromosomal lineages, suggest that evolution of these populations has involved recurrent *sym* plasmid transfer. Strikingly different symbiotic outcomes across strains when interacting with two *Mimosa* species (Figs. 3, 4) suggest that host legumes will play an important role in shaping *Cupriavidus* genetic structure at both *sym* plasmid and chromosomal loci.

Acknowledgements

I am grateful to Ann Hirsch for permission to use unpublished sequence data, to Craig Barrett for providing strains, and to Alain Rousteau for generous assistance with field work in Guadeloupe. Financial support was provided by NSF grant MCB-0640246.

References

1. Amadou, C., Pascal, G., Mangenot, S., Glew, M., Bontemps, C., Capela, D., Carrere, S., Cruveiller, S., Dossat, C., Lajus, A., Marchetti, M., Poinso, V., Rouy, Z., Servin, B., Saad, M., Schenowitz, C., Barbe, V., Batut, J., Medigue, C., Masson-Boivin, C. (2008) Genome sequence of the beta-rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. *Genome Res.* 18, 1472-1483.
2. Andam, C.P., Mondo, S.J., Parker, M.A. (2007) Monophyly of *nodA* and *nifH* genes across Texas and Costa Rican populations of *Cupriavidus* nodule symbionts. *Appl. Environ. Microbiol.* 73, 4686-4690.
3. Andrus, A.D., Andam, C., Parker, M.A. (2012) American origin of *Cupriavidus* bacteria associated with invasive *Mimosa* legumes in the Philippines. *FEMS Microbiol. Ecol.* 80, 747-50.
4. Barneby, R.C. (1991) *Sensitivae censitae*: a description of the genus *Mimosa* Linnaeus (Mimosaceae) in the New World. New York Botanical Garden, New York.
5. Barrett, C.F., Parker, M.A. (2006) Coexistence of *Burkholderia*, *Cupriavidus*, and *Rhizobium* sp. Nodule Bacteria on two *Mimosa* spp. in Costa Rica. *Appl. Environ. Microbiol.* 72, 1198-1206.
6. Benzine, J., Shelobolina, E., Xiong, M.Y., Kennedy, D.W., McKinley, J.P., Lin, X., Roden, E.E. (2013) Fe-phylosilicate redox cycling organisms from a redox transition zone in Hanford 300 Area sediments. *Frontiers Microbiol.* 4, 388.
7. Bontemps, C., Elliott, G.N., Simon, M.F., Dos Reis Junior, F.B., Gross, E., Lawton, R.C., Neto, N.E., de Fátima Loureiro, M., De Faria, S.M., Sprent, J.I., James, E.K., Young, J.P. (2010) *Burkholderia* species are ancient symbionts of legumes. *Mol. Ecol.* 19, 44–52.
8. Bournaud, C., De Faria, S.M., Dos Santos, J.M.F., Tisseyre, P., Silva, M., Chaintreuil, C., Gross, E., James, E.K., Prin, Y., Moulin, L. (2013) *Burkholderia* species are the most

- common and preferred nodulating symbionts of the *Piptadenia* Group (tribe Mimoseae). PLoS ONE 8, e63478.
9. Chen, W.M., Laevens, S., Lee, T.M., Coenye, T., De Vos, P., Mergeay, M., Vandamme, P. (2001) *Ralstonia taiwanensis* sp. nov., isolated from root nodules of *Mimosa* species and sputum of a cystic fibrosis patient. Int. J. Syst. Evol. Microbiol. 51, 1729–1735.
 10. Chen, W.M., Moulin, L., Bontemps, C., Vandamme, P., Béna, G., Boivin-Masson, C. (2003) Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature. J. Bacteriol. 185, 7266–7272.
 11. Chen, W.M., James, E.K., Chou, J.H., Sheu, S.Y., Yang, S.Z., Sprent, J.I. (2005) β -Rhizobia from *Mimosa pigra*, a newly discovered invasive plant in Taiwan. New Phytol. 168, 661–675.
 12. Chen, W.M., James, E.K., Coenye, T., Chou, J.H., Barrios, E., de Faria, S.M., Elliott, G.N., Sheu, S.Y., Sprent, J.I., Vandamme, P. (2006) *Burkholderia mimosarum* sp. nov., isolated from root nodules of *Mimosa* spp. from Taiwan and South America. Int. J. Syst. Evol. Microbiol. 56, 1847–51.
 13. Creager, R.A. (1992) Seed germination, physical and chemical control of catclaw *Mimosa* (*Mimosa pigra* var. *pigra*). Weed Technol. 6, 884–891.
 14. Didelot, X., Falush, D. (2007) Inference of bacterial microevolution using multilocus sequence data. Genetics 175, 1251–66.
 15. da Silva, K., Florentino, L.A., da Silva, K.B., de Brandt, E., Vandamme, P., de Souza Moreira, F.M. (2012) *Cupriavidus necator* isolates are able to fix nitrogen in symbiosis with different legume species. Syst. Appl. Microbiol. 35, 175–182.
 16. De Meyer, S., Cnockaert, M., Ardley, J.K., Maker, G., Yates, R., Howieson, J.G., Vandamme, P. (2013) *Burkholderia sprentiae* sp. nov. isolated from *Lebeckia ambigua* root nodules. Int. J. Syst. Evol. Microbiol. 63, 3950–3957.

17. dos Reis, F.B., Simon, M.F., Gross, E., Boddey, R.M., Elliott, G.N., Neto, N.E., Loureiro Mde, F., de Queiroz, L.P., Scotti, M.R., Chen, W.M., Norén, A., Rubio, M.C., de Faria, S.M., Bontemps, C., Goi, S.R., Young, J.P., Sprent, J.I., James, E.K. (2010) Nodulation and nitrogen fixation by *Mimosa* spp. in the Cerrado and Caatinga biomes of Brazil. *New Phytol.* 186, 934–946.
18. Elliott, G.N., Chen, W.M., Bontemps, C., Chou, J.H., Young, J.P., Sprent, J.I., James, E.K. (2007) Nodulation of *Cyclopia* spp. (Leguminosae, Papilionoideae) by *Burkholderia tuberum*. *Ann. Bot.* 100, 1403–1411.
19. Elliott, G.N., Chen, W.M., Chou, J.H., Wang, H.C., Sheu, S.Y., Perin, L., Reis, V.M., Moulin, L., Simon, M.F., Bontemps, C., Sutherland, J.M., Bessi, R., de Faria, S.M., Trinick, M.J., Prescott, A.R., Sprent, J.I., James, E.K. (2007) *Burkholderia phymatum* is a highly effective nitrogen- fixing symbiont of *Mimosa* spp. and fixes nitrogen ex planta. *New Phytol.* 173, 168-180.
20. Elliott, G.N., Chou, J.H., Chen, W.M., Bloemberg, G.V., Bontemps, C., Martinez-Romero, E., Velazquez, E., Young, J.P.W., Sprent, J.I., James, E.K. (2009) *Burkholderia* spp. are the most competitive symbionts of *Mimosa*, particularly under N-limited conditions. *Environ. Microbiol.* 11, 762-778.
21. Fraser, C., Alm, E.J., Polz, M.F., Spratt, B.G., Hanage, W.P. (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 323, 741-746.
22. Garau, G., Yates, R.J., Deiana, P., Howieson, J.G. (2009) Novel strains of nodulating *Burkholderia* have a role in nitrogen fixation with papilionoid herbaceous legumes adapted to acid, infertile soils. *Soil Biol. Biochem.* 41, 125–134.
23. Gehlot, H.S., Tak, N., Kaushik, M., Mitra, S., Chen, W.M., Poweleit, N., Panwar, D., Poonar, N., Parihar, R., Tak, A., Sankhla, I.S., Ojha, A., Rao, S.R., Simon, M.F., Reis Junior, F.B., Perigolo, N., Tripathi, A.K., Sprent, J.I., Young, J.P., James, E.K., Gyaneshwar, P. (2013)

- An invasive *Mimosa* in India does not adopt the symbionts of its native relatives. *Ann. Bot.* 112, 179-196.
24. Gyaneshwar, P., Hirsch, A.M., Moulin, L., Chen, W.M., Elliot, G.N., Bontemps, C., Estrada-de Los Santos, P., Gross, E., Dos Reis, F.B., Sprent, J.I., Young, J.P., James, E.K. (2011) Legume-nodulating betaproteobacteria: diversity, host range, and future prospects. *Mol. Plant-Microbe Int.* 24, 1276-1288.
 25. Howieson, J.G., De Meyer, S.E., Vivas-Marfisi, A.V., Ratnayake, S., Ardley, J.K., Yates, R.J. (2013) Novel *Burkholderia* bacteria isolated from *Lebeckia ambigua* – a perennial suffrutescent legume of the fynbos. *Soil Biol. Biochem.* 60, 55–64.
 26. Huson, D.H., Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254-267.
 27. Kim, S., Picardal, F. (2001) Microbial growth on dichlorobiphenyls chlorinated on both rings as a sole carbon and energy source. *Appl. Environ. Microbiol.* 67, 1953-1955.
 28. Klonowska, A., Chaintreuil, C., Tisseyre, P., Miche, L., Melkonian, R., Ducousso, M., Laguerre, G., Brunel, B., Moulin, L. (2012) Biodiversity of *Mimosa pudica* rhizobial symbionts (*Cupriavidus taiwanensis*, *Rhizobium mesoamericanum*) in New Caledonia and their adaptation to heavy metal-rich soils. *FEMS Microbiol. Ecol.* 81, 618–635.
 29. Librado, P., Rozas, J. (2009) DnaSP v5: A software package for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451-1452.
 30. Liu, X., Wei, S., Wang F, James, E.K., Guo, X., Zagar, C., Xia, L.G., Dong, X., Wang, Y.P. (2012) *Burkholderia* and *Cupriavidus* spp. are the preferred symbionts of *Mimosa* spp. in southern China. *FEMS Microbiol. Ecol.* 80, 417-26.
 31. Melkonian, R., Moulin, L., Béna, G., Tisseyre, P., Chaintreuil, C., Heulin, K., Rezkallah, N., Klonowska, A., Gonzalez, S., Simon, M., Chen, W.M., James, E.K., Laguerre, G. (2014) The geographical patterns of symbiont diversity in the invasive legume *Mimosa pudica* can

- be explained by the competitiveness of its symbionts and by the host genotype. *Environ. Microbiol.* 16, 2099-2111.
32. Mishra, R.P., Tisseyre, P., Melkonian, R., Chaintreuil, C., Miche, L., Klonowska, A., Gonzalez, S., Bena, G., Laguerre, G., Moulin, L. (2012) Genetic diversity of *Mimosa pudica* rhizobial symbionts in soils of French Guiana: investigating the origin and diversity of *Burkholderia phymatum* and other beta-rhizobia. *FEMS Microbiol. Ecol.* 79, 487–503.
 33. Ohtsubo, Y., Fujita, N., Nagata, Y., Tsuda, M., Iwasaki, T., Hatta, T. (2013) Complete genome sequence of *Ralstonia pickettii* DTP0602, a 2,4,6-trichlorophenol degrader. *Genome Announc.* 1, e00903-13.
 34. Pohlmann, A., Fricke, W.F., Reinecke, F., Kusian, B., Liesegang, H., Cramm, R., Eitinger, T., Ewering, C., Pötter, M., Schwartz, E., Strittmatter, A., Voss, I., Gottschalk, G., Steinbüchel, A., Friedrich, B., Bowien, B. (2006) Genome sequence of the bioplastic-producing “Knallgas” bacterium *Ralstonia eutropha* H16. *Nature Biotechnol.* 24, 1257 - 1262.
 35. Ronquist, F., Huelsenbeck, J.P. (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572-1574.
 36. Slatkin, M. (1985) Rare alleles as indicators of gene flow. *Evolution* 39, 53-65.
 37. Taulé, C., Zabaleta, M., Mareque, C., Platero, R., Sanjurjo, L., Sicardi, M., Frioni, L., Battistoni, F., Fabiano, E. (2012) New betaproteobacterial *Rhizobium* strains able to efficiently nodulate *Parapiptadenia rigida* (Benth.) Brenan. *Appl. Environ. Microbiol.* 78, 1692–1700.
 38. Vandamme, P., Coenye, T. (2004) Taxonomy of the genus *Cupriavidus*: a tale of lost and found. *Int. J. Syst. Evol. Microbiol.* 54, 2285-2289.
 39. Vos, M., Didelot, X. (2009) A comparison of homologous recombination rates in bacteria and archaea. *ISME Journal* 3, 199–208.

40. Wilkinson, H.H., Spoerke, J.M., Parker, M.A. (1996) Divergence in symbiotic compatibility in legume-*Bradyrhizobium* mutualism. *Evolution* 50, 1470-1477.

Table 1. Nucleotide and haplotype diversity in 80 *Cupriavidus* isolates.

	haplotypes	$\pi^a \pm 1SD$	haplotype diversity $\pm 1SD$
<i>sym</i> plasmid loci			
<i>nodA</i> (488 bp)	7	0.014 ± 0.002	0.538 ± 0.061
<i>nodC</i> (797 bp)	7	0.016 ± 0.002	0.539 ± 0.062
<i>nifH</i> (473 bp)	6	0.011 ± 0.002	0.453 ± 0.063
<i>fixA</i> (676 bp)	5	0.016 ± 0.002	0.452 ± 0.063
<i>phaZ</i> (524 bp)	5 ^b	0.012 ± 0.002	0.414 ± 0.065
chromosomal housekeeping loci			
<i>glnD</i> (742 bp)	21	0.046 ± 0.004	0.939 ± 0.010
<i>gyrB</i> (730 bp)	25	0.024 ± 0.002	0.946 ± 0.009
<i>recA</i> (701 bp)	19	0.027 ± 0.001	0.910 ± 0.015
<i>rpoB</i> (661 bp)	20	0.023 ± 0.002	0.940 ± 0.009
23S rRNA (390 bp)	17	0.020 ± 0.001	0.874 ± 0.021

^a mean proportion of nucleotide differences among isolate pairs

^b three isolates apparently lacked a *phaZ* gene; this could be considered a potential sixth haplotype

Table 2. Prevalence of private single-nucleotide variants (SNVs restricted to only one of the five sampled populations).

	total	private	private SNV mean frequency
<i>sym</i> plasmid loci	SNVs	SNVs (%)	within their population
<i>nodA</i>	36	18 (50%)	0.333
<i>nodC</i>	69	46 (67%)	0.319
<i>nifH</i>	27	19 (70%)	0.339
<i>fixA</i>	46	24 (52%)	0.338
<i>phaZ</i>	30	18 (60%)	0.354
combined ^a	208	125 (60%) ^b	0.333 ^c
chromosomal			
housekeeping loci			
<i>glnD</i>	126	10 (8%)	0.175
<i>gyrB</i>	88	24 (27%)	0.169
<i>recA</i>	73	8 (11%)	0.102
<i>rpoB</i>	70	11 (16%)	0.318
23S rRNA	24	2 (8%)	0.250
combined	381	55 (14%)	0.193

^a combined length for the five *sym* plasmid loci was 2958 bp; combined length for the five housekeeping loci was 3224 bp.

^b the five *sym* plasmid loci had a significantly higher prevalence of private SNVs than the five housekeeping loci (Mann-Whitney test, $P = 0.012$)

^c Mann-Whitney test comparing frequencies for 125 *sym* and 55 housekeeping private SNVs indicated that *sym* SNVs were significantly more common ($P < 0.0001$)

Figure legends.

Fig. 1. Bayesian tree for four concatenated *sym* plasmid loci (2434 bp) in 80 *Cupriavidus* isolates and 11 reference strains. The outgroup was deleted before plotting. Source population abbreviations: MA, Costa Rica Manuel Antonio site; TP, Costa Rica Tres Piedras site; TX, Texas; PR, Puerto Rico; Guad, Guadeloupe.

Fig. 2. Majority-rule consensus network from ClonalFrame analysis of four housekeeping loci. Clades or strains that are legume symbionts are marked with a box around the name. For the seven clades that encompassed the 80 *Cupriavidus* isolates, their *sym* plasmid haplotypes and geographic distribution (see Fig. 1 legend) are given. Branches marked with a dot were present in >99.6% of sampled trees.

Fig. 3. Growth of two *Mimosa* species when inoculated with ten *Cupriavidus* strains representing each of the five main *sym* plasmid types. Bars show the mean and 1 SE for mass gain relative to uninoculated controls of each species. ANOVA indicated that all main effects and interaction terms (*sym* plasmid type x plant species, strain within *sym* plasmid type x plant species) were highly significant. The 10 strains used were: mpj6(A1), tpud28.1(A1), mpp1.1(B), mpp1.3(B), cmp17(C1), cmp47(C1), cmp52(D), cmp60(D), mcg1(E), mcg5(E).

Fig. 4. Growth of two *Mimosa* species when inoculated with eleven *Cupriavidus* strains that all had the A1 *sym* plasmid haplotype (Fig. 1). ANOVA indicated highly significant effects of bacterial strain, host plant species, and strain x host plant interaction.

Fig. 1. Bayesian tree for four concatenated *sym* plasmid loci (2434 bp) in 80 *Cupriavidus* isolates and 11 reference strains. The outgroup was deleted before plotting. Source population abbreviations: MA, Costa Rica Manuel Antonio site; TP, Costa Rica Tres Piedras site; TX, Texas; PR, Puerto Rico; Guad, Guadeloupe.

Fig. 2. Majority-rule consensus network from ClonalFrame analysis of four housekeeping loci. Clades or strains that are legume symbionts are marked with a box around the name. For the seven clades that encompassed the 80 *Cupriavidus* isolates, their *sym* plasmid haplotypes and geographic distribution (see Fig. 1 legend) are given. Branches marked with a dot were present in >99.6% of sampled trees.

Fig. 3. Growth of two *Mimosa* species when inoculated with ten *Cupriavidus* strains representing each of the five main *sym* plasmid types. Bars show the mean and 1 SE for mass gain relative to uninoculated controls of each species. ANOVA indicated that all main effects and interaction terms (*sym* plasmid type x plant species, strain within *sym* plasmid type x plant species) were highly significant. The 10 strains used were: mpj6(A1), tpud28.1(A1), mpp1.1(B), mpp1.3(B), cmp17(C1), cmp47(C1), cmp52(D), cmp60(D), mcg1(E), mcg5(E).

Fig. 4. Growth of two *Mimosa* species when inoculated with eleven *Cupriavidus* strains that all had the A1 *sym* plasmid haplotype (Fig. 1). ANOVA indicated highly significant effects of bacterial strain, host plant species, and strain x host plant interaction.