A selective bottleneck during host entry drives the evolution of new legume symbionts

Ginaini Grazielli Doin de Moura[1], Saida Mouffok[1], Nil Gaudu[1], Anne-Claire Cazalé[1], Marine Milhes[2], Tabatha Bulach[2], Sophie Valière[2], David Roche[3], Jean-Baptiste Ferdy[4], Catherine Masson-Boivin[1], Delphine Capela[1] and Philippe Remigi[1]

[1]LIPME, Université de Toulouse, INRAE, CNRS, Castanet-Tolosan, France

[2]INRAE, US1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France

[3]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France

[4]CNRS, UMR5174 EDB, Université Toulouse 3 Paul Sabatier, Toulouse, France.

Email for correspondence: delphine.capela@inrae.fr; philippe.remigi@inrae.fr

Abstract

During the emergence of new host-microbe symbioses, multiple selective pressures –acting at the different steps of the microbial life cycle– shape the phenotypic traits that jointly determine microbial fitness. However, the relative contribution of these different selective pressures to the adaptive trajectories of microbial symbionts are still poorly known. Here we characterized the dynamics of phenotypic adaptation and its underlying genetic bases during the experimental evolution of a plant pathogenic bacterium into a legume symbiont. We observed that fast adaptation was predominantly driven by selection acting on competitiveness for host entry, which outweighed selection acting on within-host proliferation. Whole-population sequencing of evolved bacteria revealed that phenotypic adaptation was supported by the continuous accumulation of new mutations and the sequential sweeps of cohorts of mutations with similar temporal trajectories. The identification of adaptive mutations within the fixed mutational cohorts showed that all of them improved competitiveness for host entry, while only a subset of those also improved within host proliferation. Computer simulations predict that this effect emerges from the presence of a strong selective bottleneck at host entry. Together, these results show how selective bottlenecks can alter the relative influence of selective pressures acting during bacterial adaptation to multistep infection processes.

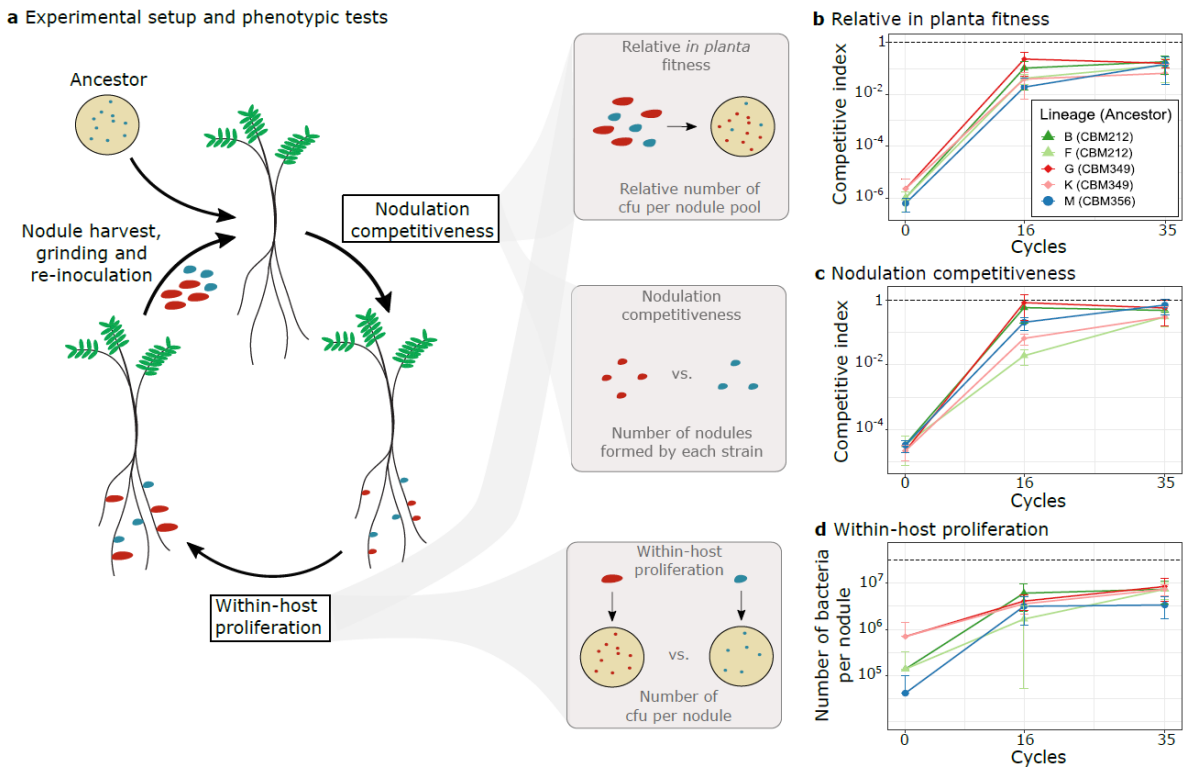
Main

Many bacterial lineages have evolved the capacity to establish symbiotic associations, either beneficial, neutral or parasitic, with eukaryotic hosts. These interactions form a dynamic continuum along which

35 bacteria can move[1]. Lifestyle changes may arise due to ecological (resource availability, host
36 environment changes or host shifts) or genomic (mutations, acquisition of new genetic material)
37 modifications. These new interactions are often initially sub-optimal for the bacterial partner[2-4], which
38 may then adapt to the selective pressures associated with its new life-cycle. For instance, horizontally-
39 transmitted microbes alternate between at least two different habitats, the host and the environment,
40 where they will face a variety of constraints and selective pressures[5,6]. Biphasic life cycles therefore
41 require bacteria not only to be able to enter and exit their hosts, but also to replicate and persist within
42 each habitat, which entails adapting to abiotic stressors, host immunity, competitors, as well as being
43 able to use specific nutrient sources. However, the relative influence of the different selective
44 pressures on the dynamics and the trajectory of bacterial adaptation to a new interaction is poorly
45 known.
46 Rhizobia are examples of facultative host-associated bacteria that can either live freely in soil or in
47 symbiotic mutualistic associations with legume plants[7]. In most legumes, rhizobia penetrate the root
48 tissue through the formation of so-called infection threads (ITs). Most of the time, only one bacterium
49 attaches to the root hair and initiates the formation of ITs[8], thus creating a strong population
50 bottleneck between the rhizosphere and internal root tissues. Bacteria then divide clonally within ITs,
51 while at the same time a nodule starts to develop at the basis of the infected root hair. Hundreds of
52 millions of bacteria are then released from ITs inside the cells of the developing nodule and
53 differentiate into nitrogen fixing bacteroids[9,10]. After several months, in nature, nodule senescence
54 leads to the release of a part of bacterial nodule population in the surrounding soil. To fulfil their life
55 cycle, rhizobia rely on the activity of numerous bacterial genes, allowing signal exchanges with the host
56 plant, proliferation and metabolic exchanges within nodule cells, while maintaining free-living
57 proficiency[7]. Somewhat surprisingly given their complex and specialized lifestyle, rhizobia evolved
58 several times independently. Horizontal transfer of key symbiotic genes is a necessary, though often
59 not sufficient, condition for a new rhizobium to emerge[11-14]. Additional steps of adaptation of the
60 recipient bacterium, occurring during evolution under plant selection, are believed to be needed to
61 actualize the symbiotic potential of emerging rhizobia[15].
62 In a previous study, we used experimental evolution to convert the plant pathogen *Ralstonia*
63 *solanacearum* into an intracellular legume symbiont[16]. Here, we analysed the dynamics of molecular
64 evolution in 5 lineages evolved for 35 cycles of nodulation through whole-population genome
65 sequencing and examined the selective and genetic bases of adaptation. Adaptation proceeded rapidly
66 during the first cycles of evolution, and was underpinned by the fixation of successive cohorts of
67 mutations within populations. We then identified adaptive mutations in two lineages and evaluated
68 their effect on each symbiotic stage of the rhizobial life cycle. Our experimental data indicated that
69 selection for nodulation competitiveness outweighs selection for multiplication within host. Computer

70 simulations further showed that the selective bottleneck at host entry and the chronology of symbiotic
71 events are critical drivers of this evolutionary pattern.
72
73 **Results**
74
75 **Fast adaptation of new legume symbionts during the first cycles of evolution**
76 To investigate the genetic and evolutionary conditions that promote the evolution of new rhizobia, we
77 previously transferred a symbiotic plasmid (pRalta) from the rhizobium *Cupriavidus taiwanensis*
78 LMG19424 into the plant pathogen *Ralstonia solanacearum* GMI1000. We obtained three nodulating
79 ancestors, two infecting nodules intracellularly (CBM212, CBM349) and one infecting nodules
80 extracellularly (CBM356)[14], that we evolved in 18 parallel bacterial lineages for 16 serial cycles of
81 nodulation on *Mimosa pudica* plants[16]. Here we continued the evolution of five of these lineages
82 (referred to as lineages B, F, G, K, and M) until cycle 35 (Figure 1a, Figure 1 – figure supplement 1 and
83 Supplementary file 1). After 35 cycles, no nitrogen fixation sustaining plant growth was observed
84 (Figure 1 – figure supplement 2). Therefore, in this manuscript, we focussed exclusively on the analysis
85 of changes in bacterial *in planta* fitness (estimated for each strain as the frequency of bacteria present
86 in nodules relative to a competing reference strain). To analyse the dynamics of fitness changes over
87 time, we compared the relative fitness of the nodulating ancestors and evolved clones from cycles 16
88 and 35 by replaying one nodulation cycle in competition with the natural symbiont *C. taiwanensis.*
89 Fitness trajectories were similar in the five lineages. While the three nodulating ancestors were on
90 average 10^6 times less fit than *C. taiwanensis*, fitness improved very quickly during the first 16 cycles,
91 then slower until cycle 35 (Figure 1b). On average, evolved clones isolated at cycle 16 and cycle 35
92 were 47 and 19 times less fit than *C. taiwanensis*, respectively (Figure 1b).
93 Because bacterial fitness in our system depends on both the capacity of strains to enter the host and
94 induce nodule formation (nodulation competitiveness) and to multiply within these nodules (within-
95 host proliferation), we analysed how each of these two fitness components changed during the
96 experiment (Figure 1c and d). In the five lineages, evolutionary trajectories of nodulation
97 competitiveness resembled that of fitness with fast improvement during the first 16 cycles, and a slow-
98 down during the last cycles. On average, nodulating ancestors were $5x10^4$ times less competitive than
99 *C. taiwanensis*, while cycle 16 and cycle 35 evolved clones were only 17 and 4 times less competitive
100 than *C. taiwanensis*, respectively (Figure 1c). Moreover, evolved clones B16, B35, G16, G35 and M35
101 were not statistically different from *C. taiwanensis* in terms of nodulation competitiveness. Within-
102 host proliferation also improved mostly during the first 16 cycles in all lineages. The rate of
103 improvement then slowed down in two lineages and continued to increase significantly in the three
104 others. As previously published, the three nodulating ancestors display different capacities to infect

105 nodules[14,16], the extracellularly infective ancestor CBM356 being the less infective (750 times less than
106 *C. taiwanensis*) and the intracellularly infective CBM349 being the most infective (45 times less than *C.*
107 *taiwanensis*). On average, cycle 16 and cycle 35 evolved clones were 8 and 5 times less infective than
108 *C. taiwanensis* (Figure 1d).
109 Altogether, the symbiotic properties of evolved clones improved rapidly during the first cycles of the
110 evolution experiment and then adaptation slowed down. Both symbiotic traits, nodulation
111 competitiveness (*i.e.* host entry) and within-host proliferation, improved greatly. However, gains in
112 nodulation competitiveness (average factor of 11,500 between ancestors and cycle 35 clones) were
113 *ca.* 150 times higher than gains in proliferation (average factor of 74 between ancestors and cycle 35
114 clones). Moreover, although the difference in nodulation competitiveness between *C. taiwanensis* and
115 the nodulating ancestors was much more important (>10^4 fold) than the difference in proliferation
116 (<10^3 fold), nodulation competitiveness of evolved clones reached the level of *C. taiwanensis* in three
117 lineages (B, G and M) while within host proliferation remained lower than *C. taiwanensis* in all of them
118 after 35 cycles.
119



120
121 **Figure 1: Evolution of the symbiotic properties of *Ralstonia* clones along evolution cycles.**
122 a, Left: overview of the evolution cycles showing the symbiotic steps determining bacterial fitness (nodulation
123 competitiveness and within-host proliferation) and human intervention (nodule harvest, grinding and re-inoculation).
124 Nodule colors reflect bacterial genotypes: blue represents the ancestor, and red represents a mutant with increased fitness.
125 Right: schematic representation of the main phenotypic measurements performed in this work, where the symbiotic
126 phenotypes of one evolved strain can be compared to those of a reference strain (e.g. the reference symbiont *C.*

127 *taiwanensis*). Competitive indexes for *in planta* fitness are calculated as the ratio of evolved vs. reference clones in nodule
128 bacterial populations normalized by the ratio of strains in the inoculum. Competitive indexes for nodulation
129 competitiveness are calculated as the ratio of nodules formed by each strain normalized by the inoculum ratio. Within-host
130 proliferation is measured in independent single inoculations of each strain. b-d, Relative *in planta* fitness (b), nodulation
131 competitiveness (c) and within-host proliferation (d) of nodulating ancestors (cycle 0) and evolved clones isolated from
132 cycles 16 and 35 were compared to *C. taiwanensis*. Values correspond to means ± standard deviations. Data were obtained
133 from at least three independent experiments. For each experiment, nodules were harvested from 10 plants (b), 20 plants
134 (c) and 6 plants (d). The sample size (n) is equal to n=3 (b,c) or comprised between n=15-18 (d). Raw data are available in
135 Figure 1 – source data 1. cfu: colony-forming units.
136 **Figure 1 – figure supplement 1:** Experimental evolution of *Ralstonia solanacearum* GMI1000 pRalta through serial cycles of
137 plant (*Mimosa pudica*) inoculation-isolation of nodule bacteria.
138 **Figure 1 – figure supplement 2:** Dry weights of *M. pudica* plants inoculated with cycle 35 evolved clones.
139 **Figure 1 – source data 1:** Raw data obtained for the phenotypic characterization of *Ralstonia* evolved clones compared to *C.*
140 *taiwanensis.*
141 **Figure 1 – figure supplement 2 – source data 1:** Raw data obtained for the measurements of the dry weights of *M. pudica*
142 plants inoculated with cycle 35 clones.
143

144 **The dynamics of molecular evolution is characterized by multiple selective sweeps of large**
145 **mutational cohorts**

146 The rapid adaptation in this experiment was unexpected given that strong population bottlenecks,
147 such as the one occurring at the nodulation step, limit effective population size at each cycle (see
148 Supplementary file 1) and are generally known to limit the rate of adaptive evolution[17]. To describe
149 the dynamics of evolution at the molecular level, we performed whole-population sequencing of
150 evolved lineages using the Illumina sequencing technology. Populations of the five lineages were
151 sequenced every other cycle until cycle 35 with a minimum sequencing coverage of 100x and a median
152 of 347x (Supplementary file 2). We detected a very large number of mutations in all sequenced
153 populations (a total of 4114 mutations, between 382 and 1204 mutations per lineage with 23.6 new
154 mutations per cycle on average) (Supplementary file 3 and Figure 2 - source data 1). This high number
155 of mutations is a result of transient hypermutagenesis occurring in the rhizosphere[18]. Mutations
156 accumulated in the populations throughout the experiment, showing no sign of slowing down until the
157 end of the experiment, which suggests that no anti-mutator mutations established in these
158 populations as we might have expected[19]. We evaluated genetic parallelism among mutations
159 detected above a frequency of 5% in the populations by computing G scores[20], a statistics used to point
160 out genes that could be mutated more often than expected by chance (Supplementary file 4). In spite
161 of high mutation rates that may obscure signals of genetic convergence, we detected signatures of
162 parallelism at the gene level among our list of mutations (observed sum of G scores of 7,540.5,
163 compared to a mean sum of 5,401.01 after 1,000 randomized simulations, Z = 31.23, $P < 10^{-200}$). The

randomized simulations identified 171 genes with a higher number of mutations than expected by chance (Bonferroni adjusted *P* value<0.01, Supplementary file 4).

To simplify the analysis of mutational trajectories, we then focused on the 819 mutations (out of a total of 4414 detected mutations) that rose above a frequency of 30%. These mutations did not show independent trajectories. Instead, we observed that groups of mutations arose synchronously and showed correlated temporal trajectories, enabling us to cluster them in cohorts[21,22] (Figure 2 and Supplementary file 5). Cohorts increased in frequency with variable speed, some being fixed within 1 to 3 cycles (representing *ca*. 25-75 bacterial generations) while others reached fixation in up to 21 cycles (representing more than 500 bacterial generations). Fixed cohorts can be very large (up to 30 mutations) (Supplementary file 5), which suggests hitchhiking of neutral or slightly deleterious mutations with one or several beneficial mutations acting as driver (or co-drivers) of the cohort[21,23]. Among the mutations that rose above 30% frequency at some stage of the experiment, 35% later declined until extinction. This is indicative of clonal interference, *i.e.* the co-occurrence of multiple subpopulations competing against each other within populations[24], which can readily be observed on Muller plots for lineages B and G (Figure 2 – figure supplement 1).

Overall, in these 5 lineages, the pattern of molecular evolution was characterized by a steady accumulation of mutations along the successive cycles and the formation of mutational cohorts, some of which containing a large number of mutations. Despite strong population bottlenecks at the root entry, mutation supply was not limiting as evidenced by the co-occurrence of competing adaptive mutations. Yet, multiple (and sometimes rapid) selective sweeps occurred throughout the 35 cycles in all lineages, suggesting that strong selection is acting on these populations.

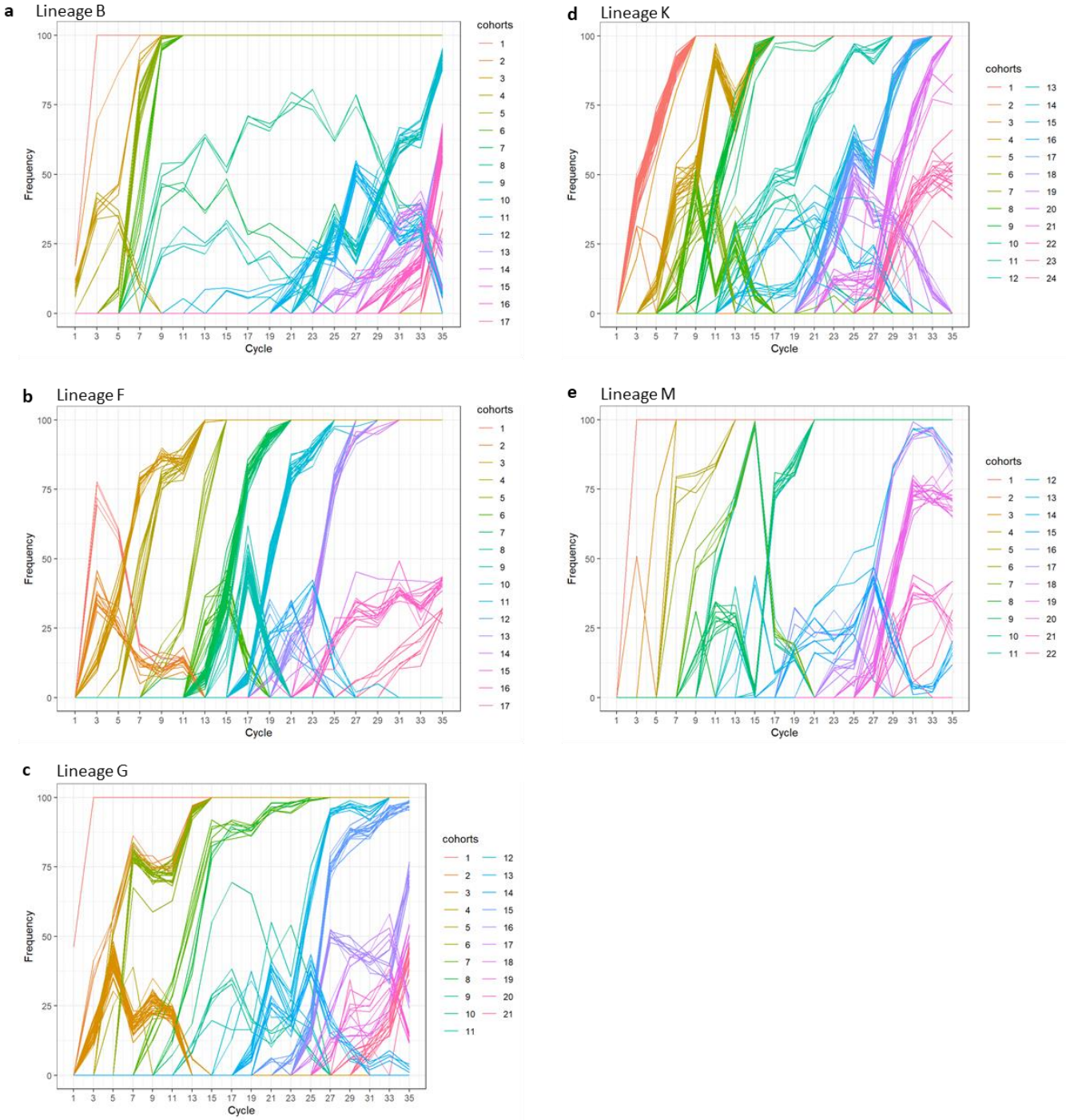


Figure 2: Dynamics of molecular evolution.

Allele frequency trajectories of mutations that attained a frequency of 30% in at least one population of the B, F, G, K and M lineages (**a**,**b**,**c**,**d**,**e**). Mutations with similar trajectories were clustered in cohorts and represented by different colors. For simplicity, mutations travelling alone were also called cohorts.

Figure 2 – figure supplement 1: Evolution of population composition over time in lines B and G.

Figure 2 – source data 1: Mutations identified in the evolved populations from all lineages

Figure 2 – source data 2: Mutations present in the evolved clones

196 **Large fixed cohorts contain multiple adaptive mutations**

197 In our system, strong nodulation bottlenecks lead to a small effective population size and to the
198 possibility that genetic drift might be responsible for the observed allelic sweeps. To determine
199 whether drift or adaptation are causing these sweeps, we searched for adaptive mutations in fixed (or
200 nearly fixed, >90% frequency) cohorts from the two lineages that have the highest symbiotic fitness
201 after 35 cycles: B and G. To do so, we introduced (by genetic engineering) individual mutations from
202 each cohort of interest into an evolved clone carrying all (or, when not available, almost all) previously
203 fixed cohorts. This allowed us to test the fitness effect associated with each mutation in a relevant
204 genetic background, taking into account possible epistatic effects arising from mutations that were
205 previously acquired in this clone. Competitive index (ratio of strains carrying the mutant vs. the wild-
206 type alleles in nodule populations normalized by the inoculum ratio) were measured for 44 mutations
207 belonging to 14 cohorts from the two lineages (Figure 3a and b). Twenty-eight mutations significantly
208 increased the fitness of evolved clones and were thus beneficial for symbiosis while 11 mutations were
209 neutral and 5 were slightly deleterious for symbiosis (Figure 3a and b, Table 1 and Supplementary file
210 6). Among the 44 genes targeted by the reconstructed mutations, 8 have a higher number of mutations
211 than expected by chance (Supplementary file 4) and adaptive mutations were identified in 6 of them.
212 Several interesting results emerge from this data set. First, some mutations show very strong adaptive
213 effects, improving more than 100 times the fitness of bacteria, in particular during the first cycles of
214 evolution. Second, many cohorts carry more than 1 adaptive mutation: six cohorts contain between 2
215 and 7 adaptive mutations. Third, 35% of adaptive mutations were found to be synonymous mutations,
216 confirming that these mutations can play a significant role in adaptive evolution[25]. All adaptive
217 synonymous mutations except one converted frequently used codons in *R. solanacearum* into unusual
218 ones (Supplementary file 7), which may explain their functional effects. Moreover, when inspecting
219 gene annotations, we found that the 28 beneficial mutations target various biological functions (Table
220 1). Strikingly, nearly 40% affect regulatory functions such as global transcription regulators (*efpR* and
221 Rsc0965)[26], putative quorum quenching (RSp0595), signal transduction (RSc1598), unknown
222 transcription regulator (RSc0243), protein dephosphorylation (Rsp1469), protein folding (*ppiB*),
223 degradosome (*rhlE1*), and DNA methylation (RSc1982). Another significant proportion of mutations
224 (32%) affect metabolism and transport genes (*gloA*, *phbB*, RSc0194, RSc2277, Rsp1116, RSp1417,
225 RSp1422, RSp1590 and pRALTA_0398) and 3 mutations are located in genes belonging to the type VI
226 secretion system (T6SS) operon (*tssJ*, RSp0759, Rsp0769). The observed genetic architecture (*i.e*. the
227 large number of genes and molecular functions targeted by adaptive mutations) indicates that
228 symbiotic traits are complex and constitute a large mutational target for adaptive evolution.

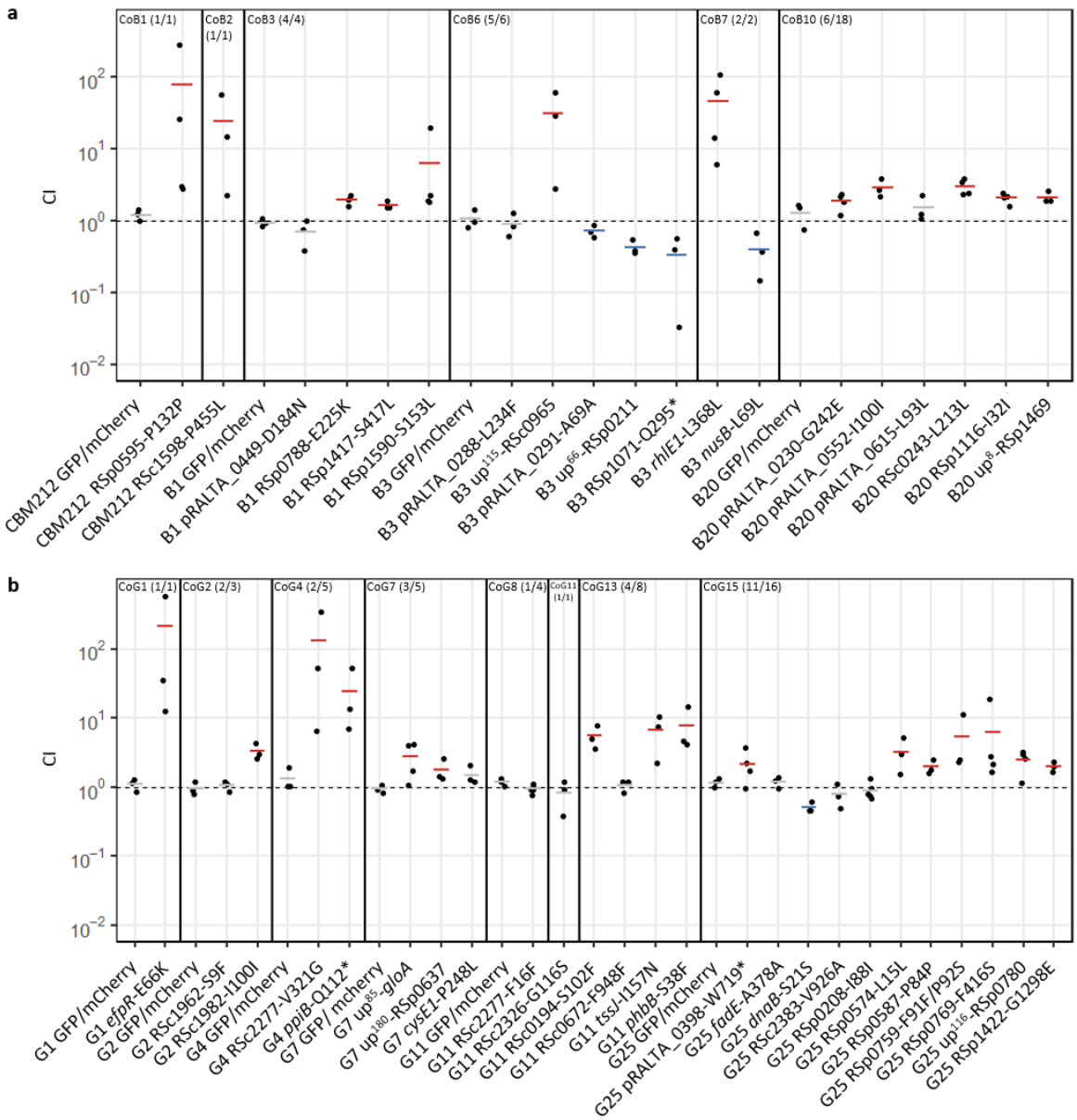


Figure 3: *In planta* fitness of reconstructed mutants from lineages B and G.

In planta relative fitness of evolved clones carrying reconstructed mutations from fixed mutational cohorts identified in lineages B (a) and G (b). Evolved clones in which mutations were reconstructed are indicated at the bottom of the graph. Vertical lines separate the different mutational cohorts. Cohort number and the number of tested mutations from the cohort on the total number of mutations present in the cohort are indicated in brackets. Competitive indexes (CI) were calculated as the ratio of the mutant strain on the isogenic parental strain in bacterial nodule populations normalized by the inoculum ratio. GFP/mCherry correspond to control co-inoculation experiments of strains derived from the same evolved clone labelled with different fluorophore either GFP or mCherry. Horizontal segments correspond to mean values of CI. Red segments indicate significantly beneficial mutations, gray segments indicate neutral mutations and blue segments indicate significantly deleterious mutations ($P<0.05$, *t*-test). Data were obtained from 3 to 5 independent experiments. The sample size (n) is comprised between n=3-5. Each CI value was obtained from pools of 40 to 227 nodules harvested from 10 plants.

Figure 3 – source data 1: Raw data obtained for the characterization of the *in planta* fitness of evolved clones carrying fixed mutations.

244 **Table 1: Symbiotic phenotypes of beneficial mutations improving *in planta* fitness**

Lineage	Gene ID[a]	Gene name	Mutation	Type of mutation[b]	Description	Cohort	Mean CI for *in planta* fitness[c]	Mean CI for plant culture medium colonization[d]	Mean CI for rhizosphere colonization[d]	Mean CI for nodulation competitiveness[c]	Mean gain factor in within-host proliferation[e]	Symbiotic phenotypes[f]	G score[g]	Adjusted *P* value[h]
B	RSp0595		P132P	Syn	Putative N-acyl-homoserine lactonase	CoB1	**77**	1	1	**5.2**	1.6	Nod+	2.80	1
	RSc1598		P455L	Nonsyn	Transmembrane sensor histidine kinase	CoB2	**24.2**	1	1.4	**4.7**	**6.4**	Nod+ Pro+	1.04	1
	RSp0788		E225K	Nonsyn	Carbamoyl-transferase	CoB3	**1.9**	1	1.1	**2**	1.2	Nod+	-0.09	1
	RSp1417		S417L	Nonsyn	Multidrug efflux transmembrane protein	CoB3	**1.6**	0.8	1.4	**3.9**	0.9	Nod+	12.47	**2.58E-04**
	RSp1590		S153L	Nonsyn	FAD-dependent oxireductase	CoB3	**6.3**	1	0.9	**2.4**	0.6	Nod+	4.10	1
	up[115]-RSc0965		G/A	Intergenic	Unknown function, negatively regulates *efpR* expression	CoB6	**31.1**	0.9	**2.9**	**8.2**	**2.6**	Rhizo+ Nod+ Pro+	nd	nd
	RSc0539	*rhlE1*	L368L	Syn	ATP-dependent RNA helicase protein	CoB7	**46.2**	1.1	**4**	**28.2**	**2.4**	Rhizo+ Nod+ Pro+	2.88	1
	RSc0243		L213L	Syn	LysR-type transcription regulator protein	CoB10	**3**	1.3	1.4	**3.9**	1.1	Rhizo+ Nod+	1.16	1
	RSp1116		I32I	Syn	Putative low specificity L-threonine aldolase	CoB10	**2.1**	1.2	1.2	**2.1**	0.9	Nod+	0.84	1
			L33F	Nonsyn										
	up[8]-RSp1469		C/T	Intergenic	Putative tyrosine specific protein phosphatase	CoB10	**2.1**	0.9	1	**2**	1.0	Nod+	nd	nd
	pRALTA_0230		G242E	Nonsyn	Putative ATP-dependent DNA helicase	CoB10	**1.5**	1.5	1.5	**2.1**	0.9	Nod+	6.34	1
	pRALTA_0552		I100I	Syn	Pseudogene, putative membrane bound hydrogenase	CoB10	**2.9**	1.2	1.4	**2.3**	0.8	Nod+	nd	nd
G	RSc1097	*efpR*	E66K	Nonsyn	Transcriptional regulator	CoG1	**212.1**	1.3	**4**	**35**	**3.7**	Rhizo+ Nod+ Pro+	3.17	1
	RSc1982		I100I	Syn	DNA methyltransferase	CoG2	**3.3**	3.1	1.3	**4.0**	0.7	Nod+	1.44	1
	RSc1164	*ppiB*	Q112*	Nonsense	Peptyl-prolyl-cis/trans isomerase	CoG4	**24.5**	1	1.0	**5.0**	1.4	Nod+	2.26	1
	RSc2277		V321G	Nonsyn	Putative transporter protein	CoG4	**133.9**	9.3	10.1	**58**	**2.5**	Nod+ Pro+	36.95	**4.68E-50**
	up[85]-RSc0520	*gloA*	T/C	Intergenic	Lactoylglutathione lyase	CoG7	**2.8**	1.2	**1.5**	**2.1**	1.1	Nod+	nd	nd
	up[180]-RSp0637		G/C	Intergenic	Hypothetical protein	CoG7	**1.8**	1	1.2	**2.6**	0.9	Nod+	nd	nd
	RSc0194		S102F	Nonsyn	Zinc-dependent alcohol dehydrogenase	CoG13	**5.5**	1.1	1.6	**3.1**	1.4	Nod+	4.67	1
	RSc1633	*phbB*	S38F	Nonsyn	Acetoacetyl-CoA reductase	CoG13	**7.9**	2	2.1	**3.3**	1.7	Rhizo+ Nod+ Pro+	1.62	1
	RSp0741	*tssJ*	I157N	Nonsyn	Type VI secretion system, lipoprotein TssJ	CoG13	**6.7**	1	1.1	**2.9**	1.7	Nod+	6.99	1
	RSp0574		L15L	Syn	Hypothetical protein	CoG15	**3.2**	1.1	1.3	**2.1**	**1.6**	Nod+ Pro+	4.36	1
	RSp0587		P84P	Syn	Putative signal peptide hypothetical protein	CoG15	**2**	1.1	0.8	**2.6**	1.3	Nod+	28.74	**1.04E-46**
	RSp0759		F91F	Syn	Putative type VI secretion-associated protein	CoG15	**5.4**	0.7	0.9	**2.5**	1.4	Nod+	1.00	1
			P92S	Nonsyn										
	RSp0769		F416S	Nonsyn	Putative type VI secretion-associated protein	CoG15	**6.3**	1.1	1.3	**1.6**	**1.8**	Nod+ Pro+	17.20	**3.98E-10**
	up[116]-RSp0780		C/T	Intergenic	Hypothetical protein	CoG15	**2.5**	1.1	1.1	**3**	1.1	Nod+	nd	nd
	RSp1422		G1298E	Nonsyn	Non-ribosomal peptide synthase	CoG15	**2**	1.2	1.2	**2.4**	1.6	Nod+	18.28	**2.79E-03**
	pRALTA_0398		W719*	Nonsense	Phosphoenol pyruvate synthase	CoG15	**2.2**	0.8	0.8	**1.6**	**1.6**	Nod+ Pro+	151.77	**0**

[a]up[x], intergenic mutations located x nucleotides upstream the gene indicated. [b]Syn, synonymous mutations. Nonsyn, non synonymous mutations. [c]Mean values of competitive indexes (CI) obtained from 3 to 4 independent experiments. Figures in bold are statistically different from 1 (Student *t*-test $P<0.05$). dMean values obtained from at least 15 measurements from 3 to 4 independent experiments. Figures in bold are statistically significant (Wilcoxon test, $P<0.05$). [e]Mean values obtained from at least 15 measurements from 3 to 4 independent experiments. Figures in bold are statistically different from 1 (Wilcoxon test, $P<0.05$ with the Benjamini-Hochberg correction). [f]Rhizo+, improvement in rhizosphere colonization. Nod+, improvement in nodulation competitiveness. Pro+, improvement in within-host proliferation. [g]G scores were calculated as proposed by Tenaillon et al. (2016). nd, G scores were not determined for mutations in pseudogenes or intergenic regions. [h]P-values associated with G scores were adjusted using a Bonferroni correction to evaluate genetic parallelism.

245

246 **Adaptive mutations predominantly improve nodulation competitiveness**

247 Next, we measured the effect of each adaptive mutation on nodulation competitiveness (reflecting
248 host entry) and within-host proliferation in order to uncover which of these two symbiotic traits can
249 explain the observed fitness gain. All mutations increasing fitness improved nodulation
250 competitiveness of evolved clones (Figure 4a and b). Gains in nodulation competitiveness ranged from
251 1.6 to 58-fold, with three mutations providing the highest gains: RSc2277-V321G (58-fold), *efpR*-E66K
252 (35-fold) and *rhlE1*-L368L (28-fold). By contrast, within-host proliferation was improved by only a
253 subset of adaptive mutations (Figure 5a and b). Only 3 mutations from lineage B (RSc1598-P455L, up^{115}-
254 RSc0965 and *rhlE1*-L368L) and 7 mutations from lineage G (*efpR*-E66K, RSc2277-V321G, *phbB*-S38F,
255 pRALTA_0398-W719*, RSp0574–L15L, RSp0769-F416S and RSp1422-G1298E) significantly improved
256 this fitness component. Moreover, gains in *in planta* proliferation were generally lower than gains in
257 nodulation competitiveness and ranged from 1.6 to 6.4-fold. The highest proliferation gains were
258 produced by the five mutations RSc1598-P455L (6.4-fold), *efpR*-E66K (3.7-fold), up^{115}-RSc0965 (2.6-
259 fold), RSc2277-V321G (2.5 fold) and *rhlE1*-L368L (2.4-fold), three of which also produced the highest
260 nodulation gains.

261 To investigate whether improvements in nodulation competitiveness could be due to a better
262 colonization of the plant culture medium or rhizosphere, we measured the survival of the 28 mutants
263 in competition with their isogenic parental strain in these two compartments. Results from these
264 assays showed that none of the adaptive mutations improved bacterial ability to colonize the culture
265 medium and only 4 of them, including up^{115}-RSc0965, *efpR*-E66K and *rhlE1*-L368L, improved the
266 colonization of the rhizosphere (Figure 4 – figure supplement 1) although by smaller factors than
267 improvements in nodulation competitiveness. These results showed that improvements in nodulation
268 competitiveness were generally not associated with a better colonization of the culture medium or of
269 the rhizosphere, indicating that host entry is mainly controlled directly by the plant and is the dominant
270 selective force driving the adaptation of legume symbionts in this evolution experiment.

271

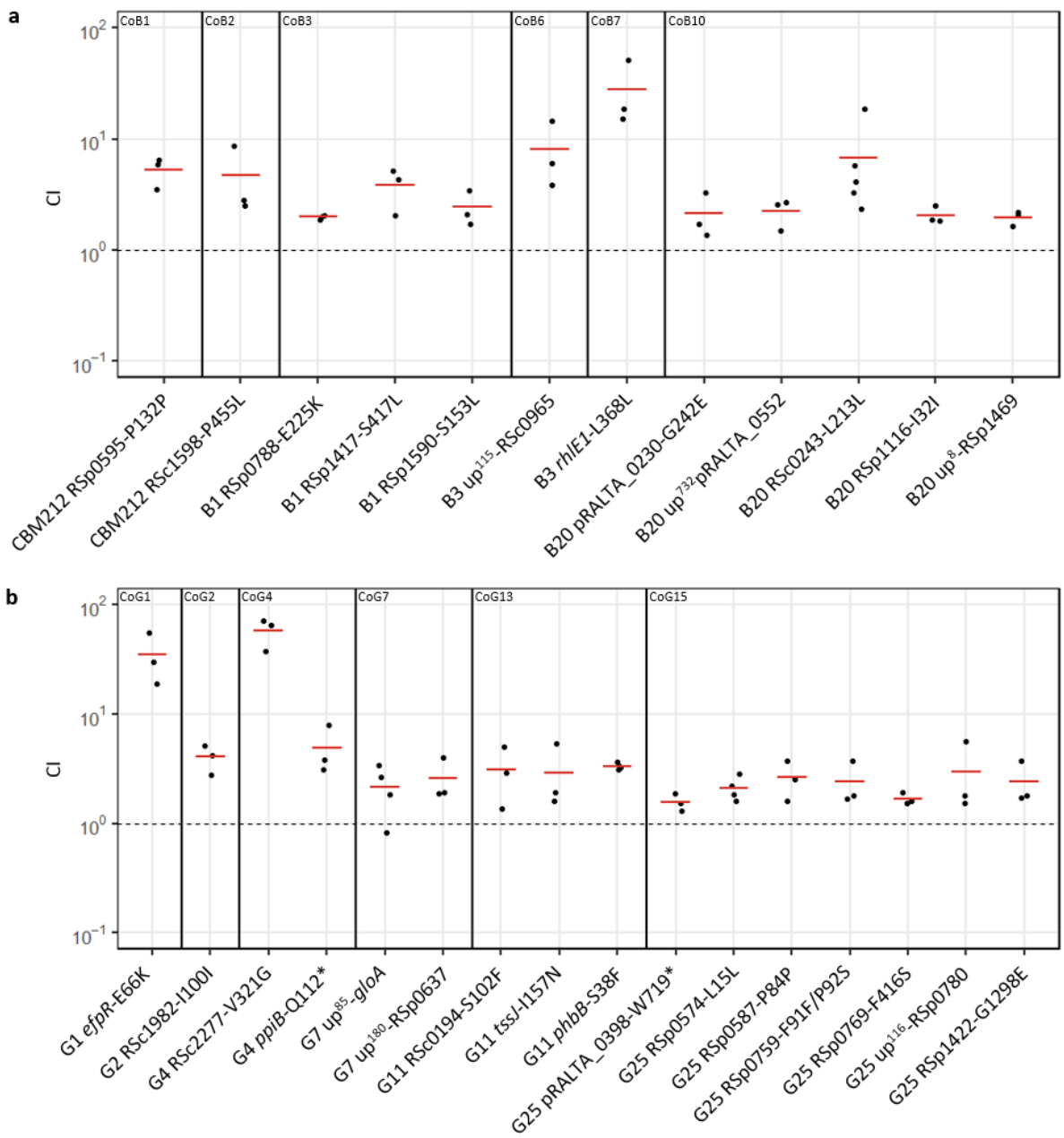
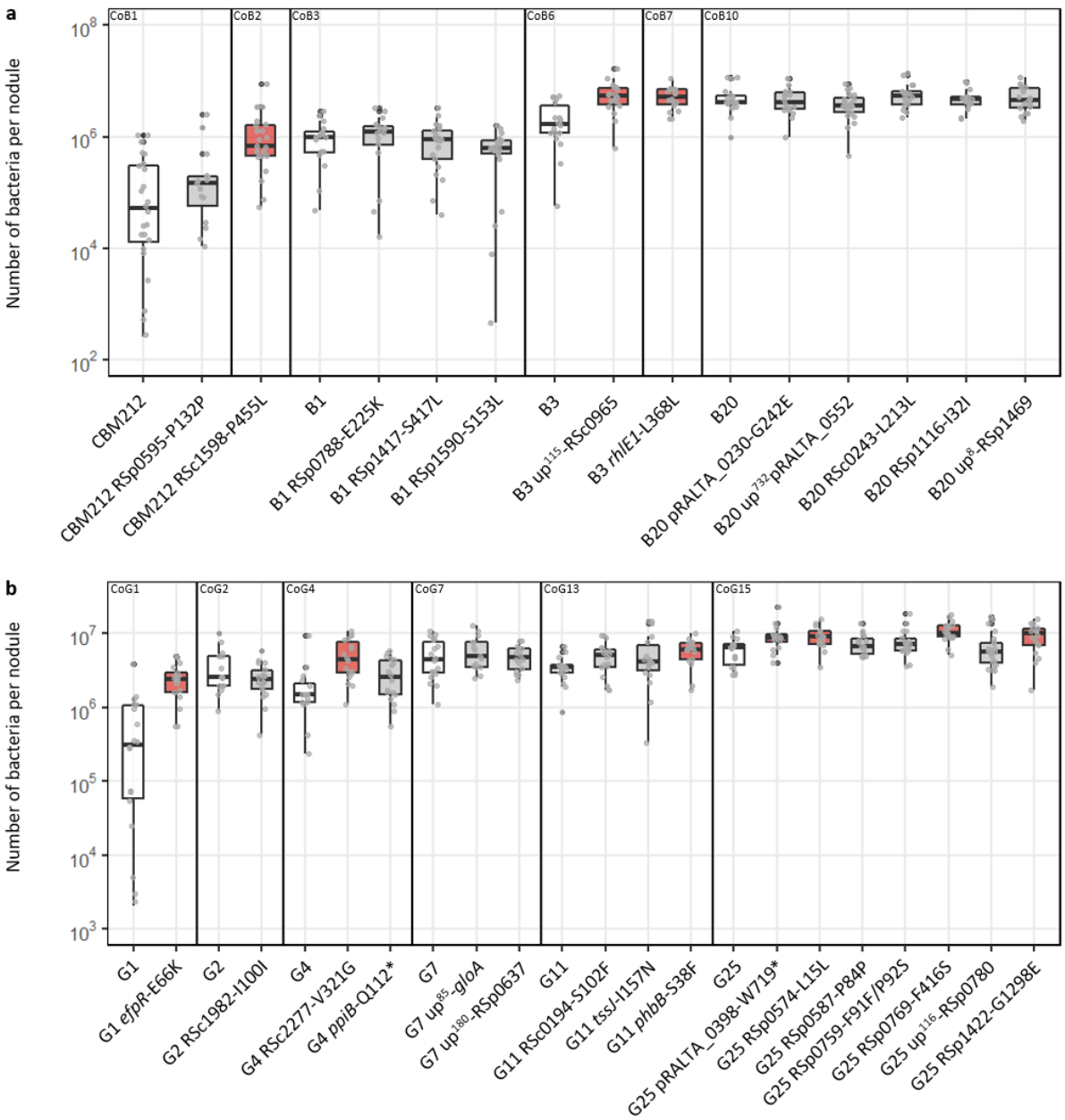


Figure 4: Nodulation competitiveness of adaptive mutants

Nodulation competitiveness effect of adaptive mutants from lineages B (**a**) and G (**b**). Evolved clones in which mutations were reconstructed are indicated at the bottom of the graph. Vertical lines separate the different mutational cohorts. Cohort number and the number of tested mutations from the cohort on the total number of mutations present in the cohort are indicated in brackets. CI were calculated as the ratio of the number of nodules formed by the mutant strain on the number of nodules formed by the isogenic parental strain normalized by the inoculum ratio. Horizontal segments correspond to mean values of CI. Red segments indicate significantly beneficial mutations (P<0.05, t-test). Data were obtained from 3 to 4 independent experiments. The sample size (n) is comprised between n=3-4. Each CI value was obtained from 95-98 nodules harvested from 20 plants.

Figure 4 – figure supplement 1: Survival of reconstructed mutants from lineages B and G in the Jensen medium and rhizosphere.

284 **Figure 4 – source data 1:** Raw data obtained for the characterization of nodulation competitiveness of evolved clones carrying
285 fixed mutations.
286 **Figure 4 – figure supplement 1– source data 1**: Raw data obtained for the Jensen culture medium colonization and
287 rhizosphere colonization assays.
288



289
290 **Figure 5: Within-host proliferation of adaptive mutants and corresponding isogenic parental clones**
291 Distribution of the number of bacteria per nodule recovered for adaptive mutants from lineage B (a) and G (b) and their
292 corresponding isogenic parental clones at 21 dpi. Rectangles span the first quartiles to the third quartiles, bold segments
293 inside the rectangle show the median, unfiled circles represent outliers, whiskers above and below each box show the
294 minimum and the maximum in the absence of suspected outlying data. Red boxes indicate the mutations that significantly
295 improved infectivity compared to the parental evolved clone (P<0.05, Wilcoxon test). Data were obtained from 3 to 5

independent experiments. For each experiment, nodules were harvested from 6 plants. The sample size (n) is comprised between n=15-24.

Figure 5 – source data 1: Raw data obtained for the characterization of the capacity of evolved clones carrying fixed mutations to proliferate within the host.

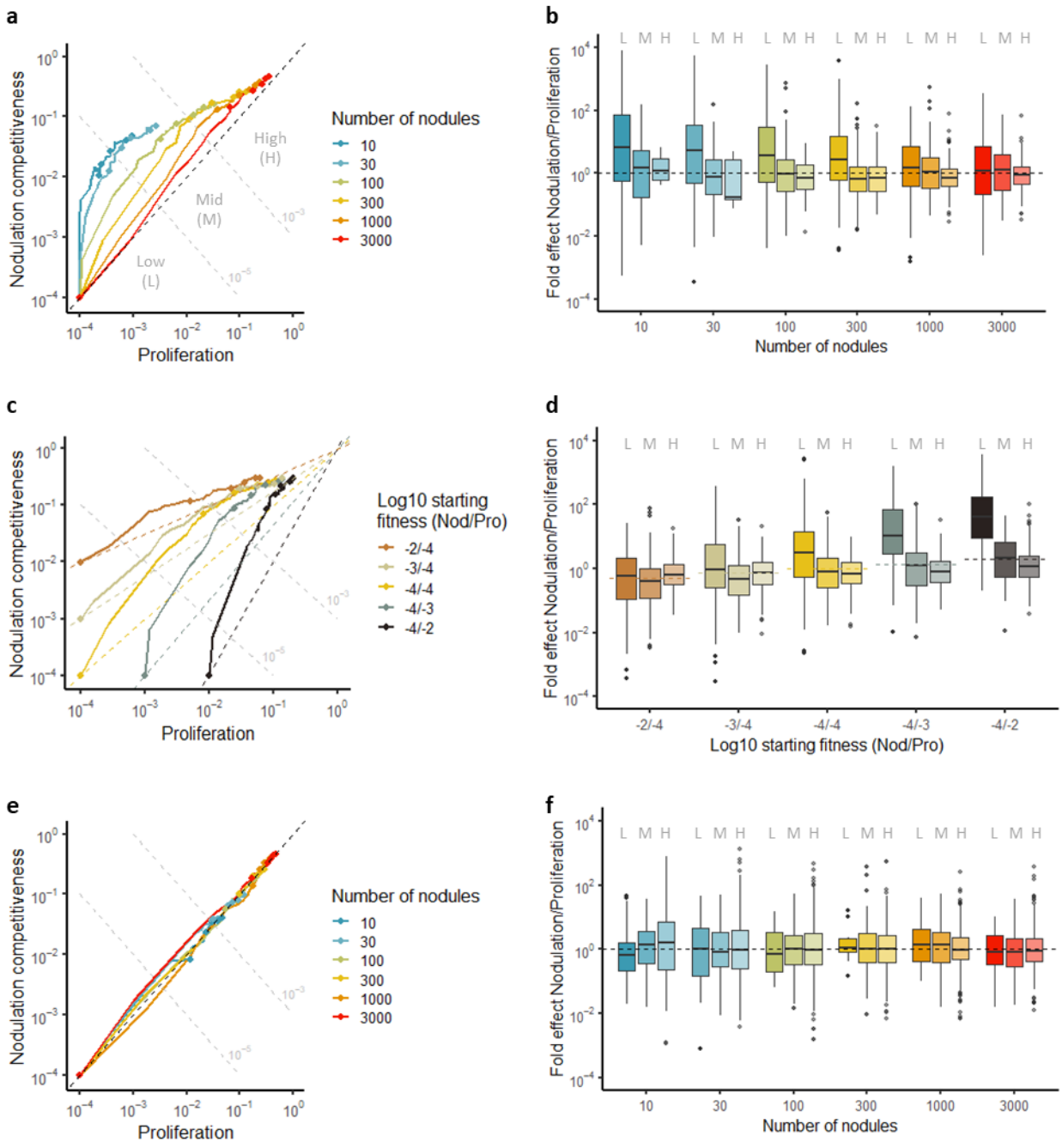
Evolutionary modelling predicts that selection for host entry drives adaptation

The preferential selection of nodulation competitiveness in our experiment prompted us to investigate the origin of this phenomenon. In particular, we wondered if stronger selection for host entry over within-host proliferation could be a general feature of symbiotic life cycles, or, instead, if it is more likely to be a specific property of our experimental system that may arise due to genetic constraints on symbiotic traits and/or specific features of the selective regime. We used computer simulations to model the evolution of bacterial populations that cycle between two compartments: the external environment (rhizosphere) and the host tissues (root nodules). In our model, bacteria accumulate mutations in the rhizosphere due to transient hypermutagenesis[18] while we consider that the bacterial population remains clonal within nodules[18] (see Supplementary file 8 for full details on the model and Figure 6 – figure supplement 1). These assumptions are relevant not only to rhizobium-legume interactions but also to other horizontally transmitted symbioses where the hosts are exposed to highly diverse environmental bacterial populations and accommodate more homogenous populations within their tissues[27-29]. In our experimental system, bacterial fitness is mediated by two phenotypic components: competitiveness to enter the host (resulting in nodulation) and within-host proliferation. Host entry imposes a very stringent selective bottleneck since (i) in most cases, each nodule is founded by one single bacterium[9,30], (ii) there were usually between 100 and 300 nodules per lineage collected at each cycle (Supplementary file 1) while *ca*. 10^6-10^7 bacteria were inoculated at each cycle and (iii) bacterial genotypes differ in their competitive ability to form nodules (Figure 4a and b). Once inside the host, bacteria multiply to reach a carrying capacity that is directly proportional to their proliferation fitness, before returning to the external environment.

When following the evolution of bacterial phenotypes in populations founded by an ancestor with low initial fitness (10^{-4} fold compared to the theoretical optimum, for each phenotypic component), we observed, in most cases, a faster increase in nodulation competitiveness relative to within-host proliferation (Figure 6a). Accordingly, the distribution of mutations that were selected at the early steps of the adaptive process was biased towards stronger improvement of nodulation competitiveness over proliferation (Figure 6b, Low fitness domain). This dominance decreased, and was sometimes reverted, as populations progressed towards high fitness values. Decreasing the strength of the host entry bottleneck also reduced the dominance of early selection for nodulation competitiveness, until abolishing it for the largest bottleneck size tested (3000 nodules).

331 In our experiment, the ancestral nodulating strains had a much higher potential for improvement in
332 nodulation competitiveness relative to proliferation ($5x10^4$ vs. 10^2-10^3 fold, respectively, compared to
333 the natural symbiont *C. taiwanensis*) which could explain the faster evolution of the former trait. We
334 performed additional simulations to test how evolutionary trajectories are affected by the genotype
335 of the ancestor. As expected, starting from a higher proliferation fitness in the ancestor further
336 increased the overall fold effect of early selected mutations on nodulation competitiveness (Figure 6c
337 and d). In contrast, starting with a higher initial nodulation competitiveness value of 10^{-3} or 10^{-2} in the
338 ancestor was not sufficient to lead to a stronger selection for proliferation, further confirming the
339 asymmetry between selective pressures acting on the two phenotypes. These findings were
340 qualitatively robust to our assumptions regarding the genetic architecture of the two symbiotic
341 phenotypes (Supplementary file 8, Figure 6 – figure supplement 2-11).
342 Finally, we wondered if the chronology of symbiotic steps impacts on selective pressures and
343 evolutionary trajectories. We modified the simulations so that clonal multiplication of bacteria
344 (according to their proliferation fitness) occurs before the selective bottleneck. Although this model is
345 not relevant in the context of the legume-rhizobium symbiosis, it may apply to other symbiotic or
346 pathogenic life cycles, for example when a selective bottleneck occurs during dissemination in the
347 environment or within the host[31,32], as well as in the case of vector-mediated transmitted
348 pathogens[33,34] or vertically transmitted symbionts[35]. In this context, the dominance of nodulation
349 competitiveness was strongly reduced, bringing the two selective forces close to equilibrium (Figure
350 6e and f and Figure 6 – figure supplement 12), in agreement with the expectation that each fitness
351 component should have a similar influence on adaptation when they have a multiplicative effect on
352 global fitness. Therefore, the dominance of nodulation competitiveness in the previous simulation runs
353 is likely a consequence of the selective bottleneck occurring before within-host bacterial multiplication
354 and just after the transient hypermutagenesis phase in the rhizosphere.
355 Altogether, our simulations show that the selective pressures that are experienced by symbiotic
356 bacteria are asymmetric, with a dominance of selection for nodulation competitiveness over within-
357 host proliferation. This asymmetry occurs when a highly diverse bacterial population from the
358 rhizosphere is exposed to the strong selective bottleneck at host entry, allowing the efficient selection
359 of the most competitive clones by the host plant. The fact that the ancestor from our evolution
360 experiment has a lower nodulation competitiveness value compared to proliferation (relative to that
361 of the wild-type rhizobium *C. taiwanensis*) is an additional factor expected to have contributed to the
362 stronger selection on nodulation competitiveness. Importantly, the key factors responsible for the
363 dominance of selection for host entry competitiveness in our model (strong bottleneck at host entry
364 and phenotypic diversity of environmental microbial populations) are found in many symbiotic

365 systems, suggesting that competition for host-entry is probably a widespread driver of bacterial
366 adaptation in diverse symbiotic associations.
367



368
369 **Figure 6: Relative strength of selection for nodulation competitiveness and within-host proliferation**
370 **a,c,e**, Median fitness trajectories of 100 simulated populations evolving under different sizes of host entry bottleneck (10 to
371 3000 nodules). Points indicate fitness values at cycles 0, 10, 20, 30, 40 and 50. Dotted grey lines represent iso-fitness lines
372 (defined as the product of nodulation competitiveness and within-host proliferation values) used to delineate 3 fitness
373 domains: Low (L; fitness < 10^{-5}), Mid (M; 10^{-5} < fitness < 10^{-3}), and High (H; fitness > 10^{-3}). The black dotted line represents the
374 diagonal, along which population would improve both phenotypic traits equally well. **b,d,f**, Fold effect (nodulation
375 competitiveness effect divided by proliferation effect) of mutations that reached a frequency of at least 30% in simulated
376 populations. Color shading indicates the fitness domain (Low, Mid or High) in which each group of mutations arose. Note that

377 fold effects of mutations are indicated even when median fitness trajectories do not reach the High fitness domain (*e.g.* 10
378 and 30 nodule lines in **a**) because some individual replicate simulations (shown in Supplementary Figures) do reach this fitness
379 domain. The effect of different parameters on adaptive trajectories was tested: the size of host entry bottleneck (**a, b**), the
380 initial fitness values of the ancestor (**c, d**), and the size of population bottleneck under a scenario where the bottleneck occurs
381 after bacterial clonal proliferation (**e, f**).
382 **Figure 6 – figure supplement 1:** Schematic representation of the modelling framework.
383 **Figure 6 – figure supplement 2**: Distribution of fitness effects of new mutations for each of the two fitness components
384 (nodulation competitiveness for host entry and proliferation), computed for three phenotypic dimensions.
385 **Figure 6 – figure supplement 3:** Representative distributions of fitness effects of new mutations for three levels of pleiotropy
386 between nodulation competitiveness and within-host proliferation.
387 **Figure 6 – figure supplement 4:** Effect of the nodulation bottleneck on the relative strength of selection for nodulation
388 competitiveness and proliferation (evolutionary parameters: k = l = 10 and m = 0).
389 **Figure 6 – figure supplement 5:** Effect of the nodulation bottleneck on the relative strength of selection for nodulation
390 competitiveness and proliferation, with a higher probability of beneficial mutations (evolutionary parameters: k = l = 3 and m
391 = 0).
392 **Figure 6 – figure supplement 6:** Effect of the nodulation bottleneck on the relative strength of selection for nodulation
393 competitiveness and proliferation, with a lower probability of beneficial mutations (evolutionary parameters: k = l = 20 and
394 m = 0).
395 **Figure 6 – figure supplement 7:** Effect of the fitness of the ancestor on the relative strength of selection for nodulation
396 competitiveness and proliferation (evolutionary parameters: k = l = 10 and m = 0).
397 **Figure 6 – figure supplement 8:** Effect of the fitness of the ancestor on the relative strength of selection for nodulation
398 competitiveness and proliferation, with a higher probability of beneficial mutations (evolutionary parameters: k = l = 3 and m
399 = 0).
400 **Figure 6 – figure supplement 9:** Effect of the fitness of the ancestor on the relative strength of selection for nodulation
401 competitiveness and proliferation, with a lower probability of beneficial mutations (evolutionary parameters: k = l = 20 and
402 m = 0).
403 **Figure 6 – figure supplement 10:** Effect of the nodulation bottleneck on the relative strength of selection for nodulation
404 competitiveness and proliferation under weak partial pleiotropy (evolutionary parameters: k = l = 6 and m = 4).
405 **Figure 6 – figure supplement 11:** Effect of the nodulation bottleneck on the relative strength of selection for nodulation
406 competitiveness and proliferation under strong partial pleiotropy (evolutionary parameters: k = l = 2 and m = 8).
407 **Figure 6 – figure supplement 12:** Effect of the chronology of symbiotic events and the size of nodulation bottleneck on the
408 relative strength of selection for nodulation competitiveness and proliferation (evolutionary parameters: k = l = 10 and m =
409 0).
410 **Figure 6 – source data 1:** Raw data from evolutionary simulations are available at: https://doi.org/10.15454/QYB2S9
411
412

413 **Discussion**

414 Horizontally transmitted symbiotic bacteria have complex lifecycles during which they have to face
415 multiple environmental constraints, both outside and inside the host. Assessing the relative influence
416 of each of these selective pressures on the adaptive trajectories of emerging symbiotic bacteria is
417 critically needed to better understand the eco-evolutionary dynamics of microbial populations[5].

418 Here we show that selection on competitiveness for host entry is the dominant selective pressure
419 shaping early adaptation of new rhizobia. This trait, mainly mediated by the plant, improved very fast
420 during our evolution experiment, to a level comparable to that of the natural *Mimosa* symbiont
421 *C. taiwanensis*, and all adaptive mutations identified in lineages B and G improved nodulation
422 competitiveness. Evolutionary simulations led us to propose that the dominance of the selection on
423 host entry over the selection on within-host proliferation that we observed in our system is strongly
424 dependent on the strength of the selective bottleneck that occurs at host entry and its position the life
425 cycle of symbionts. These results shed light on the importance of selective bottlenecks in shaping the
426 evolutionary trajectories of emerging microbial symbionts. Most theoretical[36,37] and experimental[38-43]
427 works on the influence of bottlenecks on microbial adaptation have focused on non-selective
428 bottlenecks that randomly purge genetic diversity and reduce the efficiency of natural selection. Yet,
429 another aspect of bottlenecks emerges when considering that transmission and host colonization can
430 be, at least partially, dependent on microbial genotype, prompting us to consider infection bottlenecks
431 as selective events[44]. Selective bottlenecks have already been described during virus infections[31,45-48]
432 but received less attention in bacteria, except for their role in the evolution of phenotypic
433 heterogeneity[17,49-51]. Yet, potentially all symbiotic bacterial populations (whether commensal,
434 pathogenic or mutualistic) experience bottlenecks of various intensities (sometimes going down to
435 only a few cells[52]) during their life cycles[27,53,54]. Since competitiveness for host entry is a complex,
436 polygenic phenotypic trait[7,55-58] that often displays extensive variation within natural populations[59,60],
437 bottlenecks are expected to represent major selective events (instead of purely stochastic events) in
438 many symbiotic life cycles. Interestingly, strong selection for host entry in experimentally evolved
439 mutualistic and pathogenic symbionts was recently observed in several studies[61,62] reinforcing the idea
440 that selective bottlenecks at host entry may play a crucial role in the evolution of diverse host-microbe
441 interactions.
442 In multistep infection processes, the different infection steps can be functionally linked either by
443 couplings or trade-offs[63,64]. Our study unveiled a new characteristic of the rhizobium-legume
444 interactions: the recurrent genetic coupling between nodulation competitiveness and within-host
445 proliferation. Indeed, all the 10 identified mutations that improved within-host proliferation also
446 improved nodulation competiveness, while the reverse was not true. This result corroborates our
447 previous data showing that mutations promoting nodule cell infection always improved nodule
448 formation[14,65]. It is conceivable that the intracellular release and proliferation of rhizobia in nodule cells
449 is dependent on the efficiency of the earliest symbiotic events, *i. e*. the entry and progression of
450 bacteria in infection threads and the concomitant divisions of nodule cells preparing for the
451 accommodation of bacteria. Consistently, delayed progression of infection threads across the root cell
452 layers was shown to impair the release of bacteria in nodule cells[66]. Moreover the identification of

453 plant receptors (LYK3, NFR1, Sym37, SYMRK) and transcription factors (NIN, ERN1, ERN2) involved in
454 both nodule organogenesis and rhizobial intracellular infection[67-70] supports the existence of common
455 mechanisms controlling the two processes on the plant side. Future functional analyses of the adaptive
456 mutations identified in this study will expand our understanding of the molecular bases of nodulation
457 competitiveness in rhizobia[60,71,72], and its relationship to within-host proliferation. Including
458 mutualistic traits (nitrogen fixation and host growth promotion) in the analysis of the genetic couplings
459 (or trade-offs) between the different symbiotic traits will be important to fully characterize the genetic
460 constraints shaping the evolution of rhizobium-legume interactions[72-74]. The genetic links between
461 different host colonization phases are generally poorly documented in other symbiotic systems, but
462 they were analyzed in two recent experimental evolution studies with opposite outcomes. Robinson
463 et al. (2018) showed that early bacterial adaptation to zebrafish gut favored the improvement of host
464 entry and inter-host transmission without affecting within-host proliferation[61]. In another study,
465 experimental evolution of *Vibrio fisherii* in symbiosis with squids led to the fixation of mutations in a
466 global regulatory gene (*binK*) that improved both the initiation and maintenance of the interaction,
467 through its action on several bacterial phenotypic traits[75]. Conversely, an example of trade-off
468 between early and late infection stages was recently evidenced in *Xenorhabdus nematophila*
469 interacting with insects[76,77]. In the same line, a trade-off between within-vector and within-host fitness
470 was evidenced in the mosquito-borne parasite *Plasmodium falciparum* causing malaria[78]. These
471 examples highlight the diversity and context-dependence of genetic correlations that exist between
472 bacterial phenotypic traits involved in symbiotic interactions.
473 In all lineages of this experiment, the rate of adaptation was very high during the first cycles of
474 evolution and then tended to decrease over time. This classical pattern of evolution, due to diminishing
475 return epistasis among beneficial mutations[24,79,80], was likely favored by the low initial fitness values of
476 nodulating ancestors compared to the natural symbiont *C. taiwanensis* allowing the rapid acquisition
477 of highly beneficial mutations during the first cycles (Table 1). Elevated mutation rate in the
478 rhizosphere, prior to the entry of bacteria in roots[18], has also played an important role in the dynamic
479 of adaptation by fueling the extensive genetic diversification of bacteria and exposing these
480 populations to plant-mediated selection. A probable consequence of high mutation rate combined to
481 strong selection is the presence of large cohorts of mutations that carry multiple beneficial mutations
482 along with some neutral or slightly deleterious ones. Mutational cohorts were already described in
483 previous evolution experiments[21,81], with cases of cohorts carrying 2 co-driver mutations[22]. Here we
484 identified up to 7 adaptive mutations in one of our cohorts. The occurrence of multi-driver mutations
485 per cohort might be explained by the nested, sequential fixation of multiple adaptive mutations in one
486 lineage before it reaches detectable frequency (5% in our case)[22]. This phenomenon, creating a
487 'travelling wave' of adaptation[82], was recently observed with high-resolution sequencing[23] and is

488 possibly amplified by the strong selection/high mutation regime of our experiment. Alternatively, it
489 was proposed that, in the case of hypermutagenesis, the likelihood of multiple adaptive mutations
490 arising simultaneously in a given genome becomes non-negligible and might promote saltational
491 evolution[83].
492 In conclusion, our work shows that the relative strengths of the selective forces imposed by hosts
493 strongly influence the evolutionary trajectory of microbial populations. Selection for host entry is likely
494 the predominant life-history trait selected during the evolution of new horizontally-transmitted
495 symbiotic interactions, such as the rhizobium-legume symbiosis. This effect emerges from the
496 presence of a selective bottleneck at host entry, although its predominance can be modulated by other
497 genetic and ecological factors. Investigating the role of selective bottlenecks in other symbiotic
498 interactions will improve our ability to manage and predict the eco-evolutionary dynamics of host-
499 associated microbial populations.
500
501 **Methods**
502
503 **Bacterial strains and growth conditions**
504 Strains and plasmids used in this study are listed in the Key ressources table. *Cupriavidus taiwanensis*
505 strains were grown at 28°C on tryptone-yeast (TY) medium[84]. *Ralstonia solanacearum* strains were
506 grown at 28°C either on BG medium[85] or on minimal MP medium[86] supplemented with 2% glycerol.
507 *Escherichia coli* strains were grown at 37°C on Luria-Bertani medium[87]. Antibiotics were used at the
508 following concentrations: trimethoprim at 100 µg ml^{-1}, spectinomycin at 40 µg ml^{-1}, streptomycin at
509 200 µg ml^{-1} and kanamycin at 25 µg ml^{-1} (for *E. coli*) or 50 µg ml^{-1} (for *R. solanacearum*).
510
511 **Experimental evolution**
512 Symbiotically evolved clones and populations were generated as previously described[16]. Five lineages,
513 two (B and F) derived from the CBM212 ancestor, two (G and K) derived from the CBM349 ancestor
514 and one (M) derived from the CBM356 ancestor, previously evolved for 16 cycles were further evolved
515 until cycle 35. For each lineage, at each cycle, 30 plants were inoculated with nodule bacterial
516 populations from the previous cycle. Twenty-one days after inoculation, all nodules from the 30 plants
517 were pooled, surface-sterilized with 2.6% sodium hypochlorite for 15 min, rinsed with sterile water
518 and crushed. Then, 10% of the nodule crush was used to inoculate a new set of 30 plants the same
519 day. Serial dilutions of each nodule crush were plated and one clone was randomly selected from the
520 highest dilution and purified. The selected clones, the rest of the nodule crushes, and a 24-h culture of
521 bacteria from an aliquot of nodule crushes were stored at –80°C at each cycle.
522

523 **Sequencing bacterial populations and clones**

524 Aliquots of frozen nodule bacterial populations or single colonies from purified clones were grown
525 overnight in BG medium supplemented with trimethoprim. Bacterial DNA was extracted from 1 ml of
526 culture using the Wizard genomic DNA purification kit (Promega). Evolved population DNAs were
527 sequenced at the GeT-PlaGe core facility (https://get.genotoul.fr/), INRAE Toulouse. DNA-seq libraries
528 have been prepared according to Illumina's protocols using the Illumina TruSeq Nano DNA HT Library
529 Prep Kit. Briefly, DNA was fragmented by sonication, size selection was performed using SPB beads (kit
530 beads) and adaptators were ligated to be sequenced. Library quality was assessed using an Advanced
531 Analytical Fragment Analyzer and libraries were quantified by qPCR using the Kapa Library
532 Quantification Kit (Roche). Sequencing has been performed on a NovaSeq6000 S4 lane (Illumina,
533 California, USA) using a paired-end read length of 2x150 pb with the Illumina NovaSeq Reagent Kits.
534 Evolved clones were re-sequenced either by C.E.A/IG/Genoscope using the Illumina GA2X technology
535 (clones B8, B16, F16, G8, G16, K16 and M16), or by the GeT_PlaGe core facility using the Illumina
536 technology HiSeq2000 (clones B1, B2, B3, B4, B5, B6, B7, B9, B10, B11, B12, B13, B14, B15, G1, G2, G3,
537 G4, G5, G6, G7, G9, G10, G11, G12, G13, G14, and G15), or HiSeq3000 (clones B20, B25, B30, G25 and
538 G30), or NovaSeq6000 (clones B35, F35, G35, K35 and M35).

539

540 **Detection of mutations and molecular analyses**

541 Sequencing reads from NovaSeq6000 runs (whole-populations and clones from cycle 35) were mapped
542 on the chimeric reference genome of the ancestral strain, comprising *R. solanacearum* GMI1000
543 chromosome (GenBank accession number: NC_003295.1) and megaplasmid (NC_003296.1) together
544 with *C. taiwanensis* symbiotic plasmid pRalta (CU633751). Mutations were detected using breseq
545 v0.33.1(Ref. [88]) with default parameters, either using the polymorphism mode (for whole-population
546 sequences) or the consensus mode (for individual clones). Mutation lists were curated manually in
547 order to remove mutations present in the ancestral strains as well as false positive hits arising from
548 reads misalignments in low complexity regions. Moreover, alleles showing aberrant trajectories in the
549 time-course whole-population sequencing data were checked manually, by inspecting either breseq
550 output files and/or read alignments with IGV[89], and corrected as needed. In this work, we focused our
551 attention on SNPs and indels detected above 5% frequency in the populations. Recombinations,
552 rearrangements and IS movements were not analyzed exhaustively. Mutations detected in populations
553 are listed in Figure 2 – source data 1.

554 Sequencing data of the remaining clones were analyzed as described previously[26] with the PALOMA
555 bioinformatics pipeline implemented in the Microscope platform[90]. The complete list of mutational
556 events generated for these clones are available on the Microscope platform

557 (https://mage.genoscope.cns.fr/microscope/expdata/NGSProjectEvo.php, SYMPA tag). Mutations
558 detected in clones are listed in Figure 2 – source data 2.
559 In order to identify mutations forming temporal clusters (cohorts) through populations, we selected
560 mutations having a frequency of 30% in at least one cycle and clustered their frequency using the *hclust*
561 function in R v3.6.1[91]. Then, we separated cohorts using a cutoff distance of 0.3.
562 Sub-population genealogies in lineages B and G, shown as Muller plots in Figure 2 – figure supplement
563 1, were reconstructed by comparing mutations found in cohorts and in individually sequenced clones.
564 Cohorts for which ancestry cannot be ascertained (mutations not found in any clone) were not
565 included in these plots. Relative frequencies of genotypes were calculated manually and plotted with
566 the R package MullerPlot[92].
567 G scores were calculated as described[20]. Since we found that synonymous mutations can be adaptive,
568 we used all types of mutations in coding regions to calculate this statistics. Moreover, we included all
569 mutations beyond 5% frequency in this analysis since we assumed that strong clonal interference may
570 prevent adaptive mutations to rise to high frequency. To evaluate statistical significance of G scores,
571 we ran 1,000 simulations where the total number of mutations used for the calculation of the observed
572 G scores (3330) were randomly attributed in the coding genes according to their respective length to
573 compute simulated G scores for each bacterial gene. The sum of simulated G scores was compared to
574 the observed sum and we computed a Z score and *P*-value from these simulated G statistics. These
575 simulations were also used to compute mean G scores for each gene, and to calculate the associated
576 Z scores and *P*-values (adjusted using a Bonferroni correction).
577
578 **Reconstruction of mutations in evolved clones**
579 Mutant alleles and constitutively expressed reporter genes (GFPuv, mCherry) were introduced into
580 *Ralstonia* evolved clones by co-transformation using the MuGent technique[26,93]. Briefly, two DNA
581 fragments, the first one carrying an antibiotic (kanamycin) resistance gene prepared from the pRCK-
582 P*ps*-GFP or pRCK-P*ps*-mCherry linearized plasmids allowing the integration of the resistance gene in
583 the intergenic region downstream the *glmS* gene, and the second one carrying the mutation to be
584 introduced prepared by PCR amplification of a 6 kb region using genomic DNA of evolved clones as
585 template and high fidelity Phusion polymerase (ThermoFisher Scientific), were co-transformed into
586 naturally competents cells of *Ralstonia* evolved clones. Co-transformants resistant to kanamycin were
587 screened by PCR using primers specifically amplifying mutant or wild-type alleles and verified by Sanger
588 sequencing. Primers used in mutation reconstructions are listed in the Key ressources table.
589
590 **Analyses of symbiotic phenotypes on *M. pudica***

591 *Mimosa pudica* seeds (LIPME 2019 production obtained from one commercial seed (B&T World Seed,
592 Paguignan, France) of Australian origin) were sterilized as described[94] by immersion in 95% H_2SO_4
593 during 15 minutes and 2.4% sodium hypochlorite solution during 10 minutes and rinsed in sterile
594 distilled water 5 times. Seeds were soaked in sterile water at 28°C under agitation for 24 hours and
595 then deposited on soft agar (9.375 g/L) and incubated at 28°C during 24 more hours in darkness. Then,
596 seedlings were cultivated in glass tubes (2 seedlings/tube) under N-free conditions, each tube
597 containing 20 mL of solid Fahraeus medium[95] and 40 mL of liquid Jensen medium[96] diluted 1:4 with
598 sterile water. Plants were incubated in a culture chamber at 28°C under a photoperiod day/night of 16
599 h/8 h and positioned randomly in the different experiments.
600 For *in planta* relative fitness and nodulation competitiveness assays, two strains of *Ralstonia*
601 *solanacearum* expressing differential constitutive fluorophores (GFPuv or mCherry) or one strain of *R.*
602 *solanacearum* and one strain of *C. taiwanensis*, both expressing differential antibiotic resistance
603 (streptomycin and trimethoprim), were co-inoculated onto *M. pudica* plantlets grown for 3 to 4 days
604 in the culture chamber. Both strains were inoculated in equivalent proportion (*ca*. $5x10^5$ bacteria of
605 each strain per plant) except for the comparisons of the *Ralstonia* nodulating ancestors with *C.*
606 *taiwanensis* where strains were mixed in a 1000:1 ratio (*ca.* $5x10^3$ bacteria of *C. taiwanensis* and *ca.*
607 $5x10^5$ bacteria of nodulating ancestors per plant). Nodules were harvested 21 days after inoculation,
608 surface sterilized by immersion in 2.4% sodium hypochlorite solution for 15 minutes and rinsed with
609 sterile water. For *in planta* relative fitness measurements, sterilized nodules from 20 plants (see the
610 exact number of harvested nodules in Supplementary file 1) were pooled and crushed in 1 mL of sterile
611 water, diluted and spread on selective solid medium using an easySpiral automatic plater
612 (Interscience). After 2-day incubation at 28°C, colonies were screened either by plating on selective
613 medium in case of co-inoculations of *Ralstonia* evolved clones with *C. taiwanensis* or based on
614 fluorescence using a stereo zoom microscope (AxioZoom V16, Zeiss, Oberkochen, Germany) for co-
615 inoculations of *Ralstonia* evolved clones with *Ralstonia* reconstructed mutants. For nodulation
616 competitiveness assays, *ca*. 96 individual nodules per experiment were crushed separately in 96-well
617 microtiter plates and droplets were deposited on selective medium. Nodule occupancy was
618 determined by screening bacteria grown in the droplets either on selective medium or based on their
619 fluorescence as described for fitness measurements. Both nodulation competitiveness and relative
620 fitness assays were measured in at least three independent biological replicates. Competitive indexes
621 (CI) were calculated by dividing the ratio of the number of test strain (evolved strains or reconstructed
622 mutants) vs. reference strain (*C. taiwanensis* or evolved parental strains, respectively) in nodules
623 normalized by the inoculum ratio. When CI values were all above 1, CI values were transformed by
624 their inverse and compared to the value 1 using a one-sided Student *t*-test ($P<0.05$).

625 For within-host proliferation assays, plantlets were inoculated with a single strain ($5x10^5$ bacteria per
626 plant). In each experiment, nodules from 6 individual plants were collected separately 21 days after
627 inoculation, surface sterilized for 15 minutes in a 2.4% sodium hypochlorite solution, rinsed with sterile
628 water and crushed in 1 ml of sterile water. Dilutions of nodule crushes were plated on selective solid
629 medium using an easySpiral automatic plater (Interscience). Two days after incubation at 28°C, the
630 number of colonies was counted. Within-host proliferation was estimated as the number of bacteria
631 per nodule. For each strain, 15 to 24 measurements from three independent biological replicates were
632 performed. Pairwise comparisons of proliferation values were compared using a two-sided Wilcoxon
633 rank sum test.
634 For Jensen culture medium and rhizosphere colonization assays, plantlets were co-inoculated with pair
635 of strains expressing differential constitutive fluorophores (GFPuv or mCherry) in equivalent
636 proportion ($5x10^6$ bacteria of each strain per plant). Seven days after inoculation, bacteria present in
637 the culture medium were diluted and plated on selective solid medium using an easySpiral automatic
638 plater (Interscience). Bacteria attached to the roots were resuspended in 4 ml of sterile water by
639 strongly vortexing for 1 minute, diluting and plating on selective medium using an easySpiral automatic
640 plater (Interscience). Two days after incubation at 28°C, colonies were screened for fluorescence using
641 a stereo zoom microscope (AxioZoom V16, Zeiss, Oberkochen, Germany). Competitive indexes (CI) for
642 Jensen medium or rhizosphere colonization were calculated as described above. CI values were
643 compared to the value 1 using a two-sided Wilcoxon rank sum test with Benjamini-Hochberg
644 correction.
645 Sample sizes were determined based on our previous studies[26,65,97]. In cases where we performed only
646 three independent replicates, each replicate was based on the analysis of a large number of plants or
647 nodules. Independent biological replicates were performed on different days, with different plants and
648 different bacterial cultures.
649
650 **Modelling/simulations**
651 Full details on evolutionary simulations and the list and values of parameters used in the simulations
652 are available in Supplementary files 8 and 9, respectively.
653
654 **Data availability statement**
655 Sequencing data are available under NCBI SRA BioProject ID PRJNA788708 and SRP353965. Raw
656 experimental data are available as Supplementary Information. Raw data generated from computer
657 simulations are deposited on the Data INRAE dataverse (https://doi.org/10.15454/QYB2S9).
658
659 **Code availability statement**

Computer code used to analyze genomic data and to perform computer simulations are deposited on the Data INRAE dataverse (https://doi.org/10.15454/QYB2S9).

Material availability statement

Requests for biological resources should be directed to the corresponding authors P.R. or D.C. *Ralstonia solanacearum* strains will be shared upon fulfilment of appropriate biosecurity procedures for quarantine bacteria.

Acknowledgements

We are grateful to A. Carlier and F. Roux for careful reading of the manuscript, to O. Tenaillon for helpful discussions on the project, to O. Tenaillon and H. Le Nagard for sharing their genetic algorithm script, and to Ludovic Legrand for archiving sequencing data and their associated metadata. We acknowledge the Genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12) for providing computing and storage resources. G.G.D.d.M was supported by a fellowship from the French Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI). P.R. received funding from the EU in the framework of the Marie-Curie FP7 COFUND People Programme, through the award of an AgreenSkills+ fellowship (under grant agreement n°609398) and from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 845838.
This study was supported by the Fédération de Recherche Agrobiosciences, Interactions et Biodiversité, the French National Research Agency (ANR-16-CE20-0011-01 and ANR-21-CE02-0019-01) and the "Laboratoires d'Excellence (LABEX)" TULIP (ANR-10-LABX-41)" and the "École Universitaire de Recherche (EUR)" TULIP-GS (ANR-18-EURE-0019).

Authors contributions

C.M.-B, D.C. and P.R. designed the project. G.G.D.M. performed most experiments and analyze the data. S.M., N.G., A.-C.C., D.C. and P.R. performed experiments. J.-B.F. and P.R. performed computer simulations. M.M., T.B. and S.V. performed NGS sequencing at GeT-PlaGe. D.R. analyzed mutations in clones and deposit data in the microScope platform. D.C. and P.R. supervised experimental work and data analysis. C.B.-M., J.-B.F and P.R. acquired funding for the project. G.G.D.M., J.B-F., D.C. and P.R. wrote the manuscript, which was approved by all authors.

Competing interests

The authors declare no competing interests.

695 **References**

696 1. Drew, G. C., Stevens, E. J. & King, K. C. Microbial evolution and transitions along the parasite-
697 mutualist continuum. *Nat. Rev. Microbiol*. **19**, 623-638, doi:10.1038/s41579-021-00550-7 (2021).
698 2. Nandasena, K. G., O'Hara, G. W., Tiwari, R. P., Sezmis, E. & Howieson, J. G. *In situ* lateral transfer
699 of symbiosis islands results in rapid evolution of diverse competitive strains of mesorhizobia
700 suboptimal in symbiotic nitrogen fixation on the pasture legume *Biserrula pelecinus* L. *Environ.*
701 *Microbiol.* **9**, 2496-2511, doi:10.1111/j.1462-2920.2007.01368.x (2007).
702 3. Longdon, B. et al. The causes and consequences of changes in virulence following pathogen host
703 shifts. *PLoS Pathog*. **11**, e1004728, doi:10.1371/journal.ppat.1004728 (2015).
704 4. Bonneaud, C. & Longdon, B. Emerging pathogen evolution: using evolutionary theory to
705 understand the fate of novel infectious pathogens. *EMBO Rep.* **21**, e51374,
706 doi:10.15252/embr.202051374 (2020).
707 5. Obeng, N., Bansept, F., Sieber, M., Traulsen, A. & Schulenburg, H. Evolution of microbiota-host
708 associations: the microbe's perspective. *Trends Microbiol*. **29**, 779-787,
709 doi:10.1016/j.tim.2021.02.005 (2021).
710 6. Robinson, C. D., Bohannan, B. J. & Britton, R. A. Scales of persistence: transmission and the
711 microbiome. *Curr. Opin. Microbiol*. **50**, 42-49, doi:10.1016/j.mib.2019.09.009 (2019).
712 7. Wheatley, R. M. et al. Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. *Proc.*
713 *Natl. Acad. Sci. U. S. A*. **117**, 23823-23834, doi:10.1073/pnas.2009094117 (2020).
714 8. Gage, D. J. Analysis of infection thread development using Gfp- and DsRed-expressing
715 *Sinorhizobium meliloti*. *J. Bacteriol.* **184**, 7042-7046, doi:10.1128/jb.184.24.7042-7046.2002
716 (2002).
717 9. Gage, D. J. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation
718 of temperate legumes. *Microbiol. Mol. Biol. Rev*. **68**, 280-300, doi:10.1128/mmbr.68.2.280-
719 300.2004 (2004).
720 10. Coba de la Peña, T., Fedorova, E., Pueyo, J. J. & Lucas, M. M. The symbiosome: legume and rhizobia
721 co-evolution toward a nitrogen-fixing organelle? *Front. Plant Sci.* **8**, 2229,
722 doi:10.3389/fpls.2017.02229 (2017).
723 11. Hirsch, A. M. et al. *Rhizobium meliloti* nodulation genes allow *Agrobacterium tumefaciens* and
724 *Escherichia coli* to form pseudonodules on alfalfa. *J. Bacteriol*. **158**, 1133-1143 (1984).
725 12. Abe, M. et al. Transfer of the symbiotic plasmid from *Rhizobium leguminosarum* biovar *trifolii* to
726 *Agrobacterium tumefaciens*. *J. Gen. Appl. Microbiol*. **44**, 65-74 (1998).
727 13. Nakatsukasa, H. et al. Transposon mediation allows a symbiotic plasmid of *Rhizobium*
728 *leguminosarum* bv. *trifolii* to become a symbiosis island in *Agrobacterium* and *Rhizobium*. *J. Gen.*
729 *Appl. Microbiol.* **54**, 107-118, doi:10.2323/jgam.54.107 (2008).

730 14. Marchetti, M. et al. Experimental evolution of a plant pathogen into a legume symbiont. *PLoS Biol*.
731 **8**, doi:e1000280 10.1371/journal.pbio.1000280 (2010).
732 15. Masson-Boivin, C., Giraud, E., Perret, X. & Batut, J. Establishing nitrogen-fixing symbiosis with
733 legumes: how many rhizobium recipes? *Trends Microbiol.* **17**, 458-466,
734 doi:10.1016/j.tim.2009.07.004 (2009).
735 16. Marchetti, M. et al. Experimental evolution of rhizobia may lead to either extra- or intracellular
736 symbiotic adaptation depending on the selection regime. *Mol. Ecol.* **26**, 1818-1831,
737 doi:10.1111/mec.13895 (2017).
738 17. Moxon, R. & Kussell, E. The impact of bottlenecks on microbial survival, adaptation, and
739 phenotypic switching in host-pathogen interactions. *Evolution* **71**, 2803-2816,
740 doi:10.1111/evo.13370 (2017).
741 18. Remigi, P. et al. Transient hypermutagenesis accelerates the evolution of legume endosymbionts
742 following horizontal gene transfer. *PLoS Biol.* **12**, e1001942, doi:10.1371/journal.pbio.1001942
743 (2014).
744 19. Lynch, M. et al. Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet*.
745 **17**, 704-714, doi:10.1038/nrg.2016.104 (2016).
746 20. Tenaillon, O. et al. Tempo and mode of genome evolution in a 50,000-generation experiment.
747 *Nature* **536**, 165-170, doi:10.1038/nature18959 (2016).
748 21. Lang, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast
749 populations. *Nature* **500**, 571-574, doi:10.1038/nature12344 (2013).
750 22. Buskirk, S. W., Peace, R. E. & Lang, G. I. Hitchhiking and epistasis give rise to cohort dynamics in
751 adapting populations. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 8330-8335,
752 doi:10.1073/pnas.1702314114 (2017).
753 23. Nguyen Ba, A. N. et al. High-resolution lineage tracking reveals travelling wave of adaptation in
754 laboratory yeast. *Nature* **575**, 494-499, doi:10.1038/s41586-019-1749-3 (2019).
755 24. McDonald, M. J. Microbial experimental evolution - a proving ground for evolutionary theory and
756 a tool for discovery. *EMBO Rep*. **20**, e46992, doi:10.15252/embr.201846992 (2019).
757 25. Bailey, S. F., Alonso Morales, L. A. & Kassen, R. Effects of synonymous mutations beyond codon
758 bias: the evidence for adaptive synonymous substitutions from microbial evolution experiments.
759 Genome *Biol. Evol*. **13**, doi:10.1093/gbe/evab141 (2021).
760 26. Capela, D. et al. Recruitment of a lineage-specific virulence regulatory pathway promotes
761 intracellular infection by a plant pathogen experimentally evolved into a legume symbiont. *Mol.*
762 *Biol. Evol.* **34**, 2503-2521, doi:10.1093/molbev/msx165 (2017).
763 27. Fronk, D. C. & Sachs, J. L. Symbiotic organs: the nexus of host-microbe evolution. *Trends Ecol. Evol*.
764 doi:10.1016/j.tree.2022.02.014 (2022).

765 28 Nyholm, S. V. & McFall-Ngai, M. J. The winnowing: establishing the squid-*vibrio* symbiosis. *Nat.*
766 *Rev. Microbiol.* **2**, 632-642, doi:10.1038/nrmicro957 (2004).
767 29 Kumar, N. et al. Bacterial genospecies that are not ecologically coherent: population genomics of
768 *Rhizobium leguminosarum*. Open Biol. **5**, 140133, doi: 10.1098/rsob.140133 (2015).
769 30. Daubech, B. et al. Spatio-temporal control of mutualism in legumes helps spread symbiotic
770 nitrogen fixation. *eLife* **6**, doi:10.7554/eLife.28683 (2017).
771 31. Joseph, S. B., Swanstrom, R., Kashuba, A. D. & Cohen, M. S. Bottlenecks in HIV-1 transmission:
772 insights from the study of founder viruses. *Nat. Rev. Microbiol*. **13**, 414-425,
773 doi:10.1038/nrmicro3471 (2015).
774 32. Hullahalli, K. & Waldor, M. K. Pathogen clonal expansion underlies multiorgan dissemination and
775 organ-specific outcomes during murine systemic infection. *eLife* **10**, doi:10.7554/eLife.70910
776 (2021).
777 33 Goodrich-Blair, H. & Clarke, D. J. Mutualism and pathogenesis in *Xenorhabdus* and *Photorhabdus*:
778 two roads to the same destination. *Mol. Microbiol.* **64**, 260-268, doi:10.1111/j.1365-
779 2958.2007.05671.x (2007).
780 34 Koinuma, H. et al. Spatiotemporal dynamics and quantitative analysis of phytoplasmas in insect
781 vectors. Sci. Rep. **10**, 4291, doi:10.1038/s41598-020-61042-x (2020).
782 35. Perreau, J., Zhang, B., Maeda, G. P., Kirkpatrick, M. & Moran, N. A. Strong within-host selection in
783 a maternally inherited obligate symbiont: *Buchnera* and aphids. *Proc. Natl. Acad. Sci. U. S. A*. **118**,
784 doi:10.1073/pnas.2102467118 (2021).
785 36. Bergstrom, C. T., McElhany, P. & Real, L. A. Transmission bottlenecks as determinants of virulence
786 in rapidly evolving pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 5095-5100,
787 doi:10.1073/pnas.96.9.5095 (1999).
788 37. LeClair, J. & Wahl, L. The impact of population bottlenecks on microbial adaptation. *J. Stat. Physics*
789 **172**, 114–125, doi:10.1007/s10955-017-1924-6 (2018).
790 38. Izutsu, M., Lake, D., Matson, Z., Dodson, J. & Lenski, R. Effects of periodic bottlenecks on the
791 dynamics of adaptive evolution in microbial populations. *bioRxiv* doi:10.1101/2021.12.29.474457
792 (2022).
793 39. Wein, T. & Dagan, T. The Effect of population bottleneck size and selective regime on genetic
794 diversity and evolvability in bacteria. *Genome Biol. Evol.* **11**, 3283-3290, doi:10.1093/gbe/evz243
795 (2019).
796 40. Garoff, L. et al. Population bottlenecks strongly influence the evolutionary trajectory to
797 fluoroquinolone resistance in *Escherichia coli*. *Mol. Biol. Evol.* **37**, 1637-1646,
798 doi:10.1093/molbev/msaa032 (2020).

41. Mahrt, N. et al. Bottleneck size and selection level reproducibly impact evolution of antibiotic resistance. *Nat. Ecol. Evol*. **5**, 1233-1242, doi:10.1038/s41559-021-01511-2 (2021).
42. Windels, E. M. et al. Population bottlenecks strongly affect the evolutionary dynamics of antibiotic persistence. *Mol. Biol. Evol*. **38**, 3345-3357, doi:10.1093/molbev/msab107 (2021).
43. Sousa, A. et al. Recurrent Reverse Evolution Maintains Polymorphism after Strong Bottlenecks in Commensal Gut Bacteria. *Mol. Biol. Evol*. **34**, 2879-2892, doi:10.1093/molbev/msx221 (2017).
44. Handel, A. & Bennett, M. R. Surviving the bottleneck: transmission mutants and the evolution of microbial populations. *Genetics* **180**, 2193-2200, doi:10.1534/genetics.108.093013 (2008).
45. Wilker, P. R. et al. Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 *Influenza* viruses. *Nat. Commun*. **4**, 2636, doi:10.1038/ncomms3636 (2013).
46. Moncla, L. H. et al. Selective bottlenecks shape evolutionary pathways taken during mammalian adaptation of a 1918-like avian *Influenza* virus. *Cell Host Microbe* **19**, 169-180, doi:10.1016/j.chom.2016.01.011 (2016).
47. Sobel Leonard, A. et al. Deep sequencing of *Influenza* a virus from a human challenge study reveals a selective bottleneck and only limited intrahost genetic diversification. *J. Virol.* **90**, 11247-11258, doi:10.1128/JVI.01657-16 (2016).
48. Huang, S. W. et al. A selective bottleneck shapes the evolutionary mutant spectra of Enterovirus A71 during viral dissemination in humans. *J. Virol.* **91**, doi:10.1128/JVI.01062-17 (2017).
49. Libby, E. & Rainey, P. B. Exclusion rules, bottlenecks and the evolution of stochastic phenotype switching. *Proc. Biol. Sci.* **278**, 3574-3583, doi:10.1098/rspb.2011.0146 (2011).
50. De Ste Croix, M. et al. Selective and non-selective bottlenecks as drivers of the evolution of hypermutable bacterial loci. *Mol. Microbiol*. 113, 672-681, doi:10.1111/mmi.14453 (2020).
51. Schmutzer, M. & Wagner, A. Gene expression noise can promote the fixation of beneficial mutations in fluctuating environments. *PLoS Comput. Biol.* **16**, e1007727, doi:10.1371/journal.pcbi.1007727 (2020).
52. Gama, J. A., Abby, S. S., Vieira-Silva, S., Dionisio, F. & Rocha, E. P. Immune subversion and quorum-sensing shape the variation in infectious dose among bacterial pathogens. *PLoS Pathog*. **8**, e1002503, doi:10.1371/journal.ppat.1002503 (2012).
53. Abel, S., zur Wiesch, PA, Davis, B. & Waldor, M. Analysis of bottlenecks in experimental models of infection. *PLoS Pathogen* **11**(6):e1004823doi:10.1371/journal.ppat.1004823 (2015).
54. Papkou & A Gokhale, C. T., A Schulenburg, H. Host-parasite coevolution: why changing population size matters. *Zoology* **117**, 330-338, doi:10.1016/j.zool.2016.02.001 (2016).
55. Cossart, P. & Sansonetti, P. J. Bacterial invasion: the paradigms of enteroinvasive pathogens. *Science* **304**, 242-248, doi:10.1126/science.1090124 (2004).

834 56. Visick, K. L., Stabb, E. V. & Ruby, E. G. A lasting symbiosis: how *Vibrio fischeri* finds a squid partner
835 and persists within its natural host. *Nat. Rev. Microbiol.* **19**, 654-665, doi:10.1038/s41579-021-
836 00557-0 (2021).
837 57. Rowe, H. M. et al. Bacterial factors required for transmission of *Streptococcus pneumoniae* in
838 mammalian hosts. Cell Host Microbe **25**, 884-891.e886, doi:10.1016/j.chom.2019.04.012 (2019).
839 58. Bright, M. & Bulgheresi, S. A complex journey: transmission of microbial symbionts. *Nat. Rev.*
840 *Microbiol.* **8**, 218-230, doi:10.1038/nrmicro2262 (2010).
841 59 Bongrand, C. et al. Using colonization assays and comparative genomics to discover symbiosis
842 behaviors and factors in *Vibrio fischeri*. *mBio* **11**, doi:10.1128/mBio.03407-19 (2020).
843 60 Mendoza-Suárez, M., Andersen, S. U., Poole, P. S. & Sánchez-Cañizares, C. Competition, nodule
844 occupancy, and persistence of inoculant strains: key factors in the rhizobium-legume symbioses.
845 *Front. Plant Sci*. **12**, 690567, doi:10.3389/fpls.2021.690567 (2021).
846 61. Robinson, C. D. et al. Experimental bacterial adaptation to the zebrafish gut reveals a primary role
847 for immigration. *PLoS Biol*. **16**, e2006893, doi:10.1371/journal.pbio.2006893 (2018).
848 62. Bacigalupe, R., Tormo-Mas, M., Penadés, J. R. & Fitzgerald, J. R. A multihost bacterial pathogen
849 overcomes continuous population bottlenecks to adapt to new host species. *Sci. Adv.* **5**, eaax0063,
850 doi:10.1126/sciadv.aax0063 (2019).
851 63. Ebert, D. et al. A population biology perspective on the stepwise infection process of the bacterial
852 pathogen *Pasteuria ramosa* in Daphnia. *Adv. Parasitol.* **91**, 265-310,
853 doi:10.1016/bs.apar.2015.10.001 (2016).
854 64. Hall, M. D., Routtu, J. & Ebert, D. Dissecting the genetic architecture of a stepwise infection
855 process. *Mol. Ecol*. **28**, 3942-3957, doi:10.1111/mec.15166 (2019).
856 65. Guan, S. H. et al. Experimental evolution of nodule intracellular infection in legume symbionts.
857 *ISME J*. **7**, 1367-1377, doi:10.1038/ismej.2013.24 (2013).
858 66. Xiao, T. T. et al. Fate map of *Medicago truncatula* root nodules. *Development* **141**, 3517-3528,
859 doi:10.1242/dev.110775 (2014).
860 67. Moling, S. et al. Nod factor receptors form heteromeric complexes and are essential for
861 intracellular infection in medicago nodules. *Plant Cell* **26**, 4188-4199, doi:10.1105/tpc.114.129502
862 (2014).
863 68. Cerri, M. R. et al. The symbiosis-related ERN transcription factors act in concert to coordinate
864 rhizobial host root infection. *Plant Physiol*. **171**, 1037-1054, doi:10.1104/pp.16.00230 (2016).
865 69. Capoen, W., Goormachtig, S., De Rycke, R., Schroeyers, K. & Holsters, M. SrSymRK, a plant
866 receptor essential for symbiosome formation. *Proc Natl Acad Sci U S A* **102**, 10369-10374,
867 doi:10.1073/pnas.0504250102 (2005).

868 70. Liu, J. *et al.* NIN is essential for development of symbiosomes, suppression of defence and
869 premature senescence in *Medicago truncatula* nodules. *New Phytol.* **230**, 290-303,
870 doi:10.1111/nph.17215 (2021).
871 71. Boivin, S. & Lepetit, M. Partner preference in the legume-rhizobia symbiosis and impact on legume
872 inoculation strategies. *Advances in Botanical Research*, **94**, 323-348,
873 doi:10.1016/bs.abr.2019.09.016 (2020).
874 72. Younginger, B. S. & Friesen, M. L. Connecting signals and benefits through partner choice in plant-
875 microbe interactions. *FEMS Microbiol. Lett*. **366**, doi:10.1093/femsle/fnz217 (2019).
876 73. Batstone, R. T., O'Brien, A. M., Harrison, T. L. & Frederickson, M. E. Experimental evolution makes
877 microbes more cooperative with their local host genotype. *Science* **370**, 476-478,
878 doi:10.1126/science.abb7222 (2020).
879 74 Quides, K. W. et al. Experimental evolution can enhance benefits of rhizobia to novel legume
880 hosts. *Proc. Biol. Sci.* **288**, 20210812, doi:10.1098/rspb.2021.0812 (2021).
881 75 Pankey, S. M. et al. Host-selected mutations converging on a global regulator drive an adaptive
882 leap towards symbiosis in bacteria. *Elife* **6**, doi:10.7554/eLife.24414 (2017).
883 76 Faucher, C. et al. Step-specific adaptation and trade-off over the course of an infection by GASP
884 mutation small colony variants. *mBio* **12**, doi:10.1128/mBio.01399-20 (2021).
885 77 Cambon, M. C. et al. Selection of bacterial mutants in late infections: when vector transmission
886 trades off against growth advantage in stationary phase. *mBio* **10**, doi:10.1128/mBio.01437-19
887 (2019).
888 78 Costa, G. et al. Non-competitive resource exploitation within mosquito shapes within-host malaria
889 infectivity and virulence. *Nat. Commun*. **9**, 3474, doi:10.1038/s41467-018-05893-z (2018).
890 79. Van den Bergh, B., Swings, T., Fauvart, M. & Michiels, J. Experimental design, population
891 dynamics, and diversity in microbial experimental evolution. *Microbiol. Mol. Biol. Rev*. **82**,
892 doi:10.1128/MMBR.00008-18 (2018).
893 80. Couce, A. & Tenaillon, O. A. The rule of declining adaptability in microbial evolution experiments.
894 *Front. Genet*. **6**, 99, doi:10.3389/fgene.2015.00099 (2015).
895 81. Maddamsetti, R., Lenski, R. E. & Barrick, J. E. Adaptation, clonal interference, and frequency-
896 dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics* **200**,
897 619-631, doi:10.1534/genetics.115.176677 (2015).
898 82. Neher, R. Genetic draft, selective interference, and population genetics of rapid adaptation. *Ann*.
899 *Rev. Ecol. Evol. Syst*. **44**, 195-215, doi:10.1146/annurev-ecolsys-110512-135920 (2013).
900 83. Katsnelson, M. I., Wolf, Y. I. & Koonin, E. V. On the feasibility of saltational evolution. *Proc. Natl.*
901 *Acad. Sci. U. S. A.* **116**, 21068-21075, doi:10.1073/pnas.1909031116 (2019).

902 84. Beringer. J. E. R factor transfer in *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **84**, 188-198, doi:
903 10.1099/00221287-84-1-188 (1974)
904 85. Boucher, C. A., Barberis, P. A., Trigalet, A. P. & Demery, D. A. Transposon mutagenesis of
905 *Pseudomonas solanacearum* isolation of Tn5-induced avirulent mutants. *J. Gen. Microbiol.* **131**,
906 2449-2457 (1985).
907 86. Plener, L., Manfredi, P., Valls, M. & Genin, S. PrhG, a transcriptional regulator responding to
908 growth conditions, is involved in the control of the type III secretion system regulon in *Ralstonia*
909 *solanacearum*. *J. Bacteriol*. **192**, 1011-1019, doi:10.1128/jb.01189-09 (2010).
910 87. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J.*
911 *Bacteriol.* **62**, 293-300, doi:10.1128/jb.62.3.293-300.1951 (1951).
912 88. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from
913 next-generation sequencing data using breseq. *Methods Mol. Biol.* **1151**, 165-188,
914 doi:10.1007/978-1-4939-0554-6_12 (2014).
915 89. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol*. **29**, 24-26,
916 doi:10.1038/nbt.1754 (2011).
917 90. Vallenet, D. et al. MicroScope: an integrated platform for the annotation and exploration of
918 microbial gene functions through genomic, pangenomic and metabolic comparative analysis.
919 *Nucleic Acids Res* **48**, D579-D589, doi:10.1093/nar/gkz926 (2020).
920 91. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical
921 (2019).
922 92. Farahpour, F., Saeedghalati, M. & Hoffmann, D. MullerPlot: Generates Muller Plot from
923 Population/Abundance/Frequency Dynamics Data. R package version 0.1.2. https://CRAN.R-
924 project.org/package=MullerPlot (2016)
925 93. Dalia, A. B., McDonough, E. & Camilli, A. Multiplex genome editing by natural transformation.
926 *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8937-8942, doi:10.1073/pnas.1406478111 (2014).
927 94. Chen, W. M. et al. Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in
928 nature. *J. Bacteriol.* **185**, 7266-7272, doi:10.1128/jb.185.24.7266-7272.2003 (2003).
929 95. Fahraeus, G. The infection of Clover root hairs by nodule bacteria studied by a simple glass slide
930 technique. *J. Gen. Microbiol*. **16**, 374-& (1957).
931 96. Jensen, H. L. Nitrogen fixation in leguminous plants. I. General characters of root nodule bacteria
932 isolated from species of *Medicago* and *Trifolium* in Australia. *Proc. Int. Soc. N.S.W*. **66**, 68–108
933 (1942).
934 97. Tang, M., Bouchez, O., Cruveiller, S., Masson-Boivin, C. & Capela, D. Modulation of quorum
935 sensing as an adaptation to nodule cell infection during experimental evolution of legume
936 symbionts. *mBio* **11**, doi:10.1128/mBio.03129-19 (2020).

937

938 **Supplementary information**

939 **Supplementary files**

940 **Supplementary file 1:** Number of nodules harvested at each evolution cycle.

941 **Supplementary file 2:** Coverage depth for whole-population sequencing

942 **Supplementary file 3:** Summary of detected mutations in the five lineages

943 **Supplementary file 4:** Analysis of genetic parallelism in the five lineages.

944 **Supplementary file 5:** Global analysis of mutational cohorts

945 **Supplementary file 6:** Neutral and deleterious reconstructed mutations

946 **Supplementary file 7:** Frequencies of usage of codons modified by adaptive synonymous mutations

947 **Supplementary file 8:** Supplementary information on the evolutionary simulations

948 **Supplementary file 9:** Parameter values used in computer simulations

949

950 **Source Data**

951 **Figure 1 – source data 1**. Raw data obtained for the phenotypic characterization of *Ralstonia* evolved

952 clones compared to *C. taiwanensis.*

953 **Figure 1 – figure supplement 2 – source data 1:** Raw data obtained for the measurements of the dry

954 weights of M. pudica plants inoculated with cycle 35 clones.

955 **Figure 2 – source data 1.** Mutations identified in the evolved populations from all lineages

956 **Figure 2 – source data 2:** Mutations present in the evolved clones

957 **Figure 3 – source data 1:** Raw data obtained for the characterization of the *in planta* fitness of evolved

958 clones carrying fixed mutations.

959 **Figure 4 – source data 1:** Raw data obtained for the characterization of nodulation competitiveness of

960 evolved clones carrying fixed mutations.

961 **Figure 4 – figure supplement 1– source data 1:** Raw data obtained for the Jensen culture medium

962 colonization and rhizosphere colonization assays.

963 **Figure 5 – source data 1**: Raw data obtained for the characterization of the capacity of evolved clones

964 carrying fixed mutations to proliferate within the host.

965 **Figure 6 – source data 1:** Raw data from evolutionary simulations are available at:

966 https://doi.org/10.15454/QYB2S9

967

968 **Key ressources table:** Strains, plasmids, oligonucleotides and software used and sequence data

969 generated in this study.

970

971 **Figure supplements**

972 **Figure 1 – figure supplement 1:** Experimental evolution of *Ralstonia solanacearum* GMI1000 pRalta
973 through serial cycles of plant (*Mimosa pudica*) inoculation-isolation of nodule bacteria.
974 **Figure 1 – figure supplement 2:** Dry weights of *M. pudica* plants inoculated with cycle 35 evolved
975 clones.
976 **Figure 2 - figure supplement 1:** Evolution of population composition over time in lines B and G
977 **Figure 4 - figure supplement 1:** Survival of reconstructed mutants from lineages B and G in the Jensen
978 medium and rhizosphere.
979 **Figure 6 – figure supplement 1:** Schematic representation of the modelling framework.
980 **Figure 6 – figure supplement 2**: Distribution of fitness effects of new mutations for each of the two
981 fitness components (nodulation competitivity and proliferation), computed for three phenotypic
982 dimensions.
983 **Figure 6 – figure supplement 3:** Representative distributions of fitness effects of new mutations for
984 three levels of pleiotropy between nodulation competitiveness and within-host proliferation.
985 **Figure 6 – figure supplement 4:** Effect of the nodulation bottleneck on the relative strength of selection
986 for nodulation competitiveness and proliferation (evolutionary parameters: k = l = 10 and m = 0).
987 **Figure 6 – figure supplement 5:** Effect of the nodulation bottleneck on the relative strength of selection
988 for nodulation competitiveness and proliferation, with a higher probability of beneficial mutations
989 (evolutionary parameters: k = l = 3 and m = 0).
990 **Figure 6 – figure supplement 6:** Effect of the nodulation bottleneck on the relative strength of selection
991 for nodulation competitiveness and proliferation, with a lower probability of beneficial mutations
992 (evolutionary parameters: k = l = 20 and m = 0).
993 **Figure 6 – figure supplement 7:** Effect of the fitness of the ancestor on the relative strength of selection
994 for nodulation competitiveness and proliferation (evolutionary parameters: k = l = 10 and m = 0).
995 **Figure 6 – figure supplement 8:** Effect of the fitness of the ancestor on the relative strength of selection
996 for nodulation competitiveness and proliferation, with a higher probability of beneficial mutations
997 (evolutionary parameters: k = l = 3 and m = 0).
998 **Figure 6 – figure supplement 9:** Effect of the fitness of the ancestor on the relative strength of selection
999 for nodulation competitiveness and proliferation, with a lower probability of beneficial mutations
1000 (evolutionary parameters: k = l = 20 and m = 0).
1001 **Figure 6 – figure supplement 10:** Effect of the nodulation bottleneck on the relative strength of
1002 selection for nodulation competitiveness and proliferation under weak partial pleiotropy (evolutionary
1003 parameters: k = l = 6 and m = 4).
1004 **Figure 6 – figure supplement 11:** Effect of the nodulation bottleneck on the relative strength of
1005 selection for nodulation competitiveness and proliferation under strong partial pleiotropy
1006 (evolutionary parameters: k = l = 2 and m = 8).

1007 **Figure 6 – figure supplement 12:** Effect of the chronology of symbiotic events and the size of
1008 nodulation bottleneck on the relative strength of selection for nodulation competitiveness and
1009 proliferation (evolutionary parameters: k = l = 10 and m = 0).