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– will shape the phenotypic traits that jointly determine microbial fitness. However, the relative contribution of these different selective pressures on the adaptive trajectories of microbial symbionts are still poorly known. Here we characterized the dynamics of phenotypic adaptation and the underlying genetic bases during the experimental evolution of a plant pathogenic bacterium into a legume symbiont. We show that fast adaptation was supported by the sequential fixation of mutational cohorts within populations, and dominated by selection for host entry competitiveness over within-host proliferation. Computer simulations predict that this effect emerges from the presence of a strong selective bottleneck at host entry before bacterial multiplication. However, identifying and characterizing adaptive mutations revealed recurrent couplings between the two phenotypic traits, further increasing the efficiency of host-mediated selection of new symbionts.

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 bacteria can move[1]. Lifestyle changes may arise due to ecological (resource availability, host environment changes or host shifts) or genomic (mutations, acquisition of new genetic material) modifications. These new interactions are often initially sub-optimal for the bacterial partner[2-4], which

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

70

71

72 **Results**

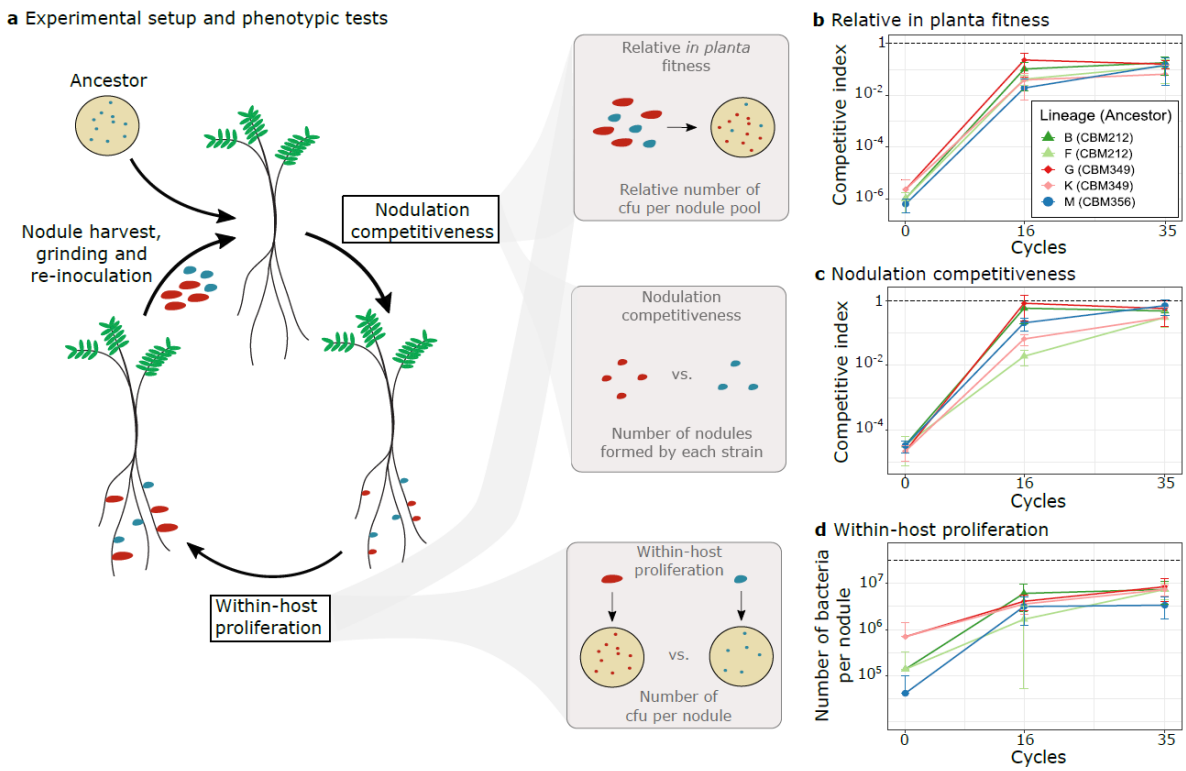
73

74 **Fast adaptation of new legume symbionts during the first cycles of evolution**

75 To investigate the genetic and evolutionary conditions that promote the evolution of new rhizobia, we
76 previously transferred a symbiotic plasmid (pRalta) from the rhizobium *Cupriavidus taiwanensis*
77 LMG19424 into the plant pathogen *Ralstonia solanacearum* GMI1000. We obtained three nodulating
78 ancestors, two infecting nodules intracellularly (CBM212, CBM349) and one infecting nodules
79 extracellularly (CBM356)[14], that we evolved in 18 parallel bacterial lineages for 16 serial cycles of
80 nodulation on *Mimosa pudica* plants[16]. Here we continued the evolution of five of these lineages
81 (referred to as lineages B, F, G, K, and M) until cycle 35 (Fig. 1a, Extended Data Fig. 1, Supplementary
82 Table 1). After 35 cycles, no nitrogen fixation sustaining plant growth was observed. Therefore, we
83 focus exclusively on the analysis of changes in bacterial fitness in this manuscript. To analyse the
84 dynamics of fitness changes over time, we compared the relative fitness of the nodulating ancestors
85 and evolved clones from cycles 16 and 35 by replaying one nodulation cycle in competition with the
86 natural symbiont *C. taiwanensis.* Fitness trajectories were similar in the five lineages. While the three
87 nodulating ancestors were on average 10^6 times less fit than *C. taiwanensis*, fitness improved very
88 quickly during the first 16 cycles, then slower until cycle 35 (Fig. 1b,c). On average, evolved clones
89 isolated at cycle 16 and cycle 35 were 47 and 19 times less fit than *C. taiwanensis*, respectively (Fig. 1c,
90 and Supplementary Table 2).

91 Because bacterial fitness in our system depends on both the capacity of strains to enter the host and
92 induce nodule formation (nodulation competitiveness) and to multiply within these nodules (within-
93 host proliferation), we analysed how each of these two fitness components changed during the
94 experiment (Fig. 1b,c,d). In the five lineages, evolutionary trajectories of nodulation competitiveness
95 resembled that of fitness with fast improvement during the first 16 cycles, and a slow-down during the
96 last cycles. On average, nodulating ancestors were $5x10^4$ times less competitive than *C. taiwanensis*,
97 while cycle 16 and cycle 35 evolved clones were only 17 and 4 times less competitive than *C.*
98 *taiwanensis*, respectively (Fig. 1c and Supplementary Table 2). Moreover, evolved clones B16, B35,
99 G16, G35 and M35 were not statistically different from *C. taiwanensis* in terms of nodulation
100 competitiveness. Within-host proliferation also improved mostly during the first 16 cycles in all
101 lineages. The rate of improvement then slowed down in two lineages and continued to increase
102 significantly in the three others. As previously published, the three nodulating ancestors display
103 different capacities to infect nodules[14,16], the extracellularly infective ancestor CBM356 being the less
104 infective (750 times less than *C. taiwanensis*) and the intracellularly infective CBM349 being the most

105 infective (45 times less than *C. taiwanensis*). On average, cycle 16 and cycle 35 evolved clones were 8
106 and 5 times less infective than *C. taiwanensis* (Fig. 1d and Supplementary Table 2).
107 Altogether, the symbiotic properties of evolved clones improved rapidly during the first cycles of the
108 evolution experiment and then adaptation slowed down. Both symbiotic traits, nodulation
109 competitiveness (*i.e.* host entry) and within-host proliferation, improved greatly. However, gains in
110 nodulation competitiveness (average factor of 11,500 between ancestors and cycle 35 clones) were
111 *ca.* 150 times higher than gains in proliferation (average factor of 74 between ancestors and cycle 35
112 clones). Moreover, although the difference in nodulation competitiveness between *C. taiwanensis* and
113 the nodulating ancestors was much more important (>10^4 fold) than the difference in proliferation
114 (<10^3 fold), nodulation competitiveness of evolved clones reached the level of *C. taiwanensis* in three
115 lineages (B, G and M) while within host proliferation remained lower than *C. taiwanensis* in all of them
116 after 35 cycles.
117



118
119 **Fig. 1 | Evolution of the symbiotic properties of Ralstonia clones along evolution cycles.**
120 a, Left: overview of the evolution cycles showing the symbiotic steps determining bacterial fitness (nodulation
121 competitiveness and within-host proliferation) and human intervention (nodule harvest, grinding and re-inoculation).
122 Nodule colors reflect bacterial genotypes: blue represents the ancestor, and red represents a mutant with increased fitness.
123 Right: schematic representation of the main phenotypic measurements performed in this work, where the symbiotic
124 phenotypes of one evolved strain can be compared to those of a reference strain (e.g. the reference symbiont C.
125 taiwanensis). Competitive indexes for in planta fitness are calculated as the ratio of evolved vs. reference clones in nodule
126 bacterial populations normalized by the ratio of strains in the inoculum. Competitive indexes for nodulation

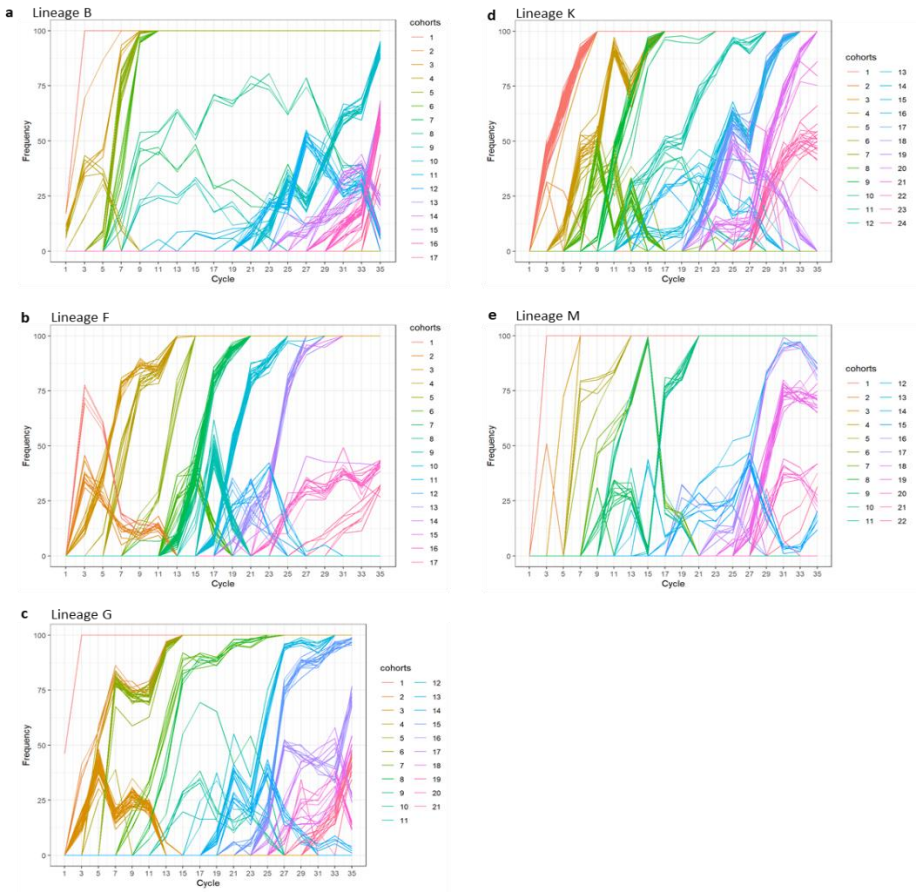
competitiveness are calculated as the ratio of nodules formed by each strain normalized by the inoculum ratio. Within-host proliferation is measured in independent single inoculations of each strain. b-d, Relative in planta fitness (b), nodulation competitiveness (c) and within-host proliferation (d) of nodulating ancestors (cycle 0) and evolved clones isolated from cycles 16 and 35 were compared to C. taiwanensis. Values correspond to means ± standard deviations. Data were obtained from at least three independent experiments. For each experiment, nodules were harvested from 10 plants (b), 20 plants (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 Supplementary Table 2. cfu: colony-forming units.

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

The rapid adaptation in this experiment was striking given that strong nodulation bottlenecks (limiting effective population size at each cycle (Supplementary Table 1)) are generally known to limit the rate of adaptive evolution[17]. To describe the dynamics of evolution at the molecular level, we performed whole-population sequencing of evolved lineages using the Illumina sequencing technology. Populations of the five lineages were sequenced every other cycle until cycle 35 with a minimum sequencing coverage of 100x and a median of 347x (Supplementary Table 3). We detected a very large number of mutations in all sequenced populations (a total of 4114 mutations, between 382 and 1204 mutations per lineage with 23.6 new mutations per cycle on average) (Extended Data Table 1 and Supplementary Tables 4-8). This high number of mutations is a result of transient hypermutagenesis occurring in the rhizosphere[18]. Mutations accumulated in the populations throughout the experiment, showing no sign of slowing down until the end of the experiment, which suggests that no anti-mutator mutations established in these populations as we might have expected[19]. We evaluated genetic parallelism among mutations detected above a frequency of 5% in the populations by computing G scores[20], a statistics used to point out genes that could be mutated more often than expected by chance (Extended Data Table 2). In spite of high mutation rates that may obscure signals of genetic convergence, we detected signatures of parallelism at the gene level among our list of mutations (observed sum of G scores of 7,540.5, 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 (Extended Data Table 2).

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%. We observed that groups of mutations showed correlated temporal trajectories and clustered them in cohorts[21,22] (Fig. 2 and Extended Data Table 3). 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) (Extended Data Table 3), which suggests hitchhiking of neutral or slightly deleterious

163 mutations with one or several beneficial mutations acting as driver (or co-drivers) of the cohort[21,23].
164 Among the mutations that rose above 30% frequency at some stage of the experiment, 35% later
165 declined until extinction. This is indicative of clonal interference, *i.e.* the co-occurrence of multiple
166 subpopulations competing against each other within populations[24], which can readily be observed on
167 Muller plots for lineages B and G (Extended Data Fig. 2).
168 Overall, in these 5 lineages, the pattern of molecular evolution was characterized by a steady
169 accumulation of mutations along the successive cycles and the formation of mutational cohorts, some
170 of which containing a large number of mutations. Despite strong selective population bottlenecks at
171 the root entry, mutation supply was not limiting as evidenced by the co-occurrence of competing
172 adaptive mutations. Yet, multiple (and sometimes rapid) selective sweeps occurred throughout the 35
173 cycles in all lineages, indicating that strong selection is acting on these populations.
174



175
176 **Fig. 2 | Dynamics of molecular evolution**
177 Allele frequency trajectories of mutations that attained a frequency of 30% in at least one population of the B, F, G, K and M
178 lineages (a,b,c,d,e). Mutations with similar trajectories were clustered in cohorts and represented by different colors. For
179 simplicity, mutations travelling alone were also called cohorts.
180

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

182 Next, we searched for driver mutations in fixed (or nearly fixed, >90% frequency) cohorts from the two
183 lineages that have the highest symbiotic fitness after 35 cycles: B and G. To do so, we identified, for
184 each cohort of interest, a clone carrying all (or, when not available, almost all) previously fixed cohorts
185 of mutations and used it to reconstruct individual point mutations from this cohort. This allowed us to
186 test the fitness effect associated with each mutation in a relevant genetic background, taking into
187 account possible epistatic effects arising from mutations that were previously acquired in this clone.
188 Competitive index (ratio of strains carrying the mutant vs. the wild-type alleles in nodule populations
189 normalized by the inoculum ratio) were measured for 44 mutations belonging to 14 cohorts from the
190 two lineages (Fig. 3a,d and Supplementary Table 7). Twenty-eight mutations significantly increased the
191 fitness of evolved clones and were thus beneficial for symbiosis while 11 mutations were neutral and
192 5 were slightly deleterious for symbiosis (Fig. 3a,d, Table 1 and Extended Data Table 4). Among the 44
193 genes targeted by the reconstructed mutations, 8 have a higher number of mutations than expected
194 by chance (Extended Data Table 2) and adaptive mutations were identified in 6 of them.
195 Several interesting results emerge from this data set. First, some mutations show very strong adaptive
196 effects, improving more than 100 times the fitness of bacteria, in particular during the first cycles of
197 evolution. Second, many cohorts carry more than 1 adaptive mutation: six cohorts contain between 2
198 and 7 adaptive mutations. Third, 35% of adaptive mutations were found to be synonymous mutations,
199 confirming that these mutations can play a significant role in adaptive evolution[25]. All adaptive
200 synonymous mutations except one converted frequently used codons in *R. solanacearum* into unusual
201 ones (Extended Data Table 5), which may explain their functional effects. Moreover, when inspecting
202 gene annotations, we found that the 28 beneficial mutations target various biological functions (Table
203 1). Strikingly, nearly 40% affect regulatory functions such as global transcription regulators (*efpR* and
204 Rsc0965)[26], putative quorum quenching (RSp0595), signal transduction (RSc1598), unknown
205 transcription regulator (RSc0243), protein dephosphorylation (Rsp1469), protein folding (*ppiB*),
206 degradosome (*rhlE1*), and DNA methylation (RSc1982). Another significant proportion of mutations
207 (32%) affect metabolism and transport genes (*gloA*, *phbB*, RSc0194, RSc2277, Rsp1116, RSp1417,
208 RSp1422, RSp1590 and pRALTA_0398) and 3 mutations are located in genes belonging to the type VI
209 secretion system (T6SS) operon (*tssJ*, RSp0759, Rsp0769). The observed genetic architecture (*i.e*. the
210 large number of genes and molecular functions targeted by adaptive mutations) indicates that
211 symbiotic traits are complex and constitute a large mutational target for adaptive evolution.
212
213

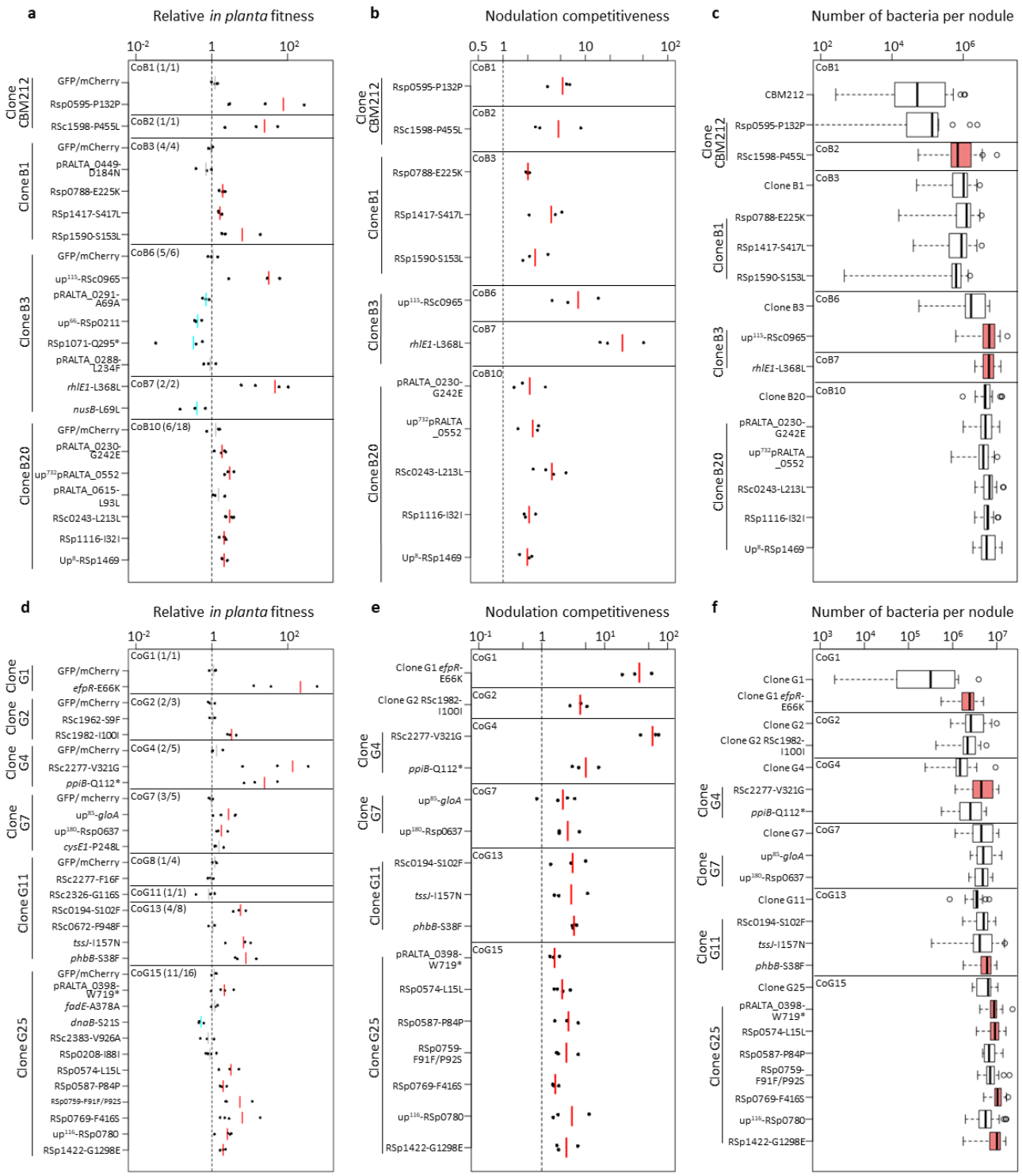


Fig. 3| Symbiotic phenotypes of reconstructed mutants and evolved clones from lineages B and G.

a,d, In planta relative fitness of evolved clones carrying reconstructed mutations from fixed mutational cohorts. Evolved clones in which mutations were reconstructed are indicated on the left of the graph. Horizontal 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. b,e, Nodulation competitiveness effect of adaptive mutants. CI were calculated as the ratio of the number

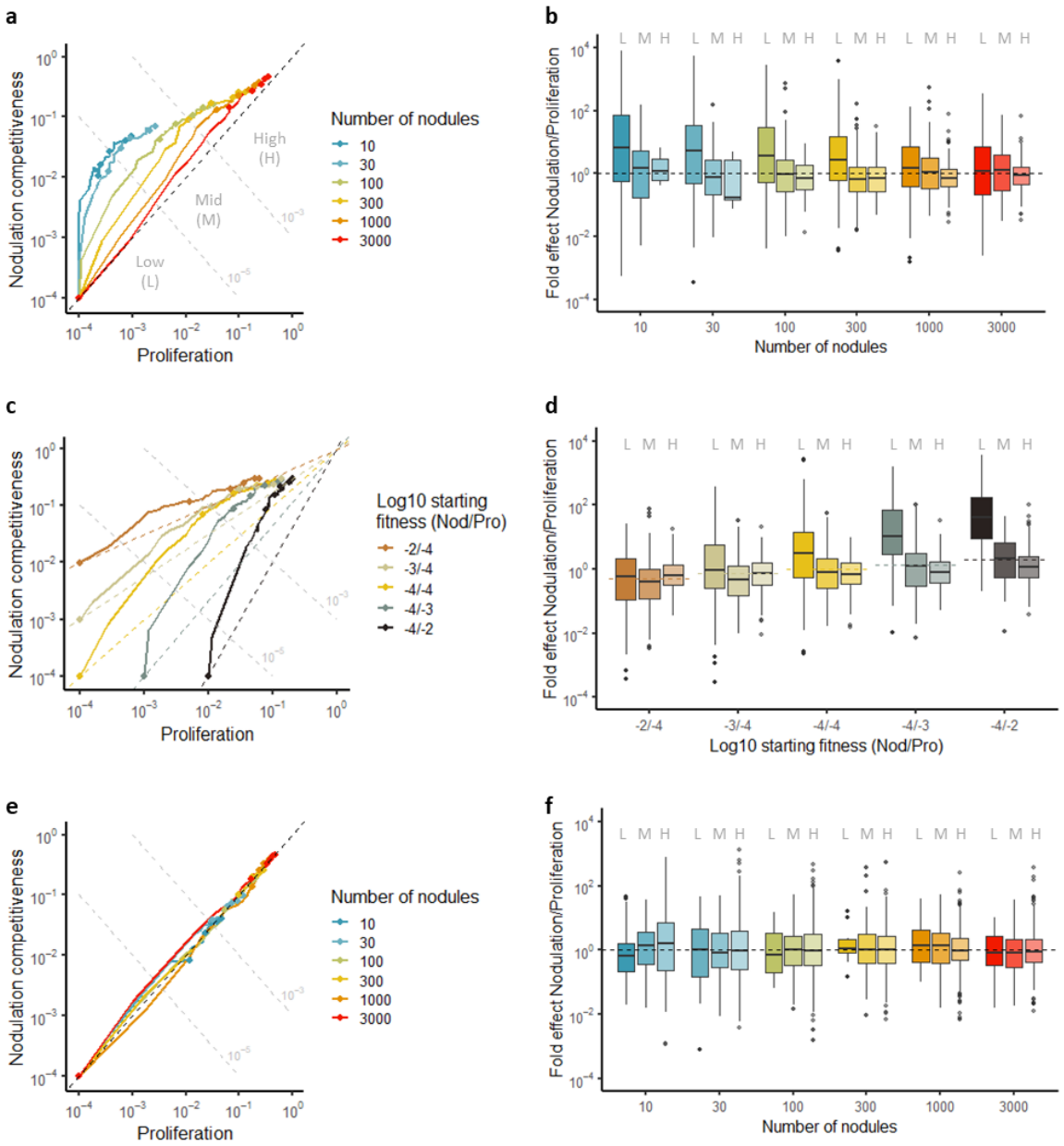
223 of nodules formed by the mutant strain on the number of nodules formed by the isogenic parental strain normalized by the
224 inoculum ratio. a,b,d,e, vertical segments correspond to mean values of CI. Red segments indicate significantly beneficial
225 mutations, gray segments indicate neutral mutations and blue segments indicate significantly deleterious mutations ($P<0.05$,
226 t-test). c,f, Distribution of within-host proliferation values (number of bacteria per nodule) of adaptive mutants and the
227 corresponding isogenic parental clones. Rectangles span the first quartiles to the third quartiles, bold segments inside the
228 rectangle show the median, unfiled circles represent outliers, whiskers above and below each box show the minimum and
229 the maximum in the absence of suspected outlying data. Red stars indicate the mutations that significantly improved
230 infectivity compared to the parental evolved clone ($P<0.05$, Wilcoxon test). a,b,c,d,e,f, Data were obtained from 3 to 5
231 independent experiments. For each experiment, nodules were harvested from 10 plants (a,d), 20 plants (b,e) and 6 plants
232 (c,f). The sample size (n) is comprised between n=3-5 (a,d), n=3-4 (b,e) and n=15-24 (c,f). Raw data are available in
233 Supplementary Table 9.
234

235 **Adaptive mutations predominantly improve nodulation competitiveness**

236 Next, we measured the effect of each adaptive mutation on nodulation competitiveness (reflecting
237 host entry) and within-host proliferation in order to uncover which of these two symbiotic traits can
238 explain the observed fitness gain. All mutations increasing fitness improved nodulation
239 competitiveness of evolved clones (Fig. 3b,e and Supplementary Table 9). Gains in nodulation
240 competitiveness ranged from 1.6 to 58-fold, with three mutations providing the highest gains:
241 RSc2277-V321G (58-fold), *efpR*-E66K (35-fold) and *rhlE1*-L368L (28-fold). By contrast, within-host
242 proliferation was improved by only a subset of adaptive mutations (Fig. 3c,f and Supplementary Table
243 9). Only 3 mutations from lineage B (RSc1598-P455L, up^{115}-RSc0965 and *rhlE1*-L368L) and 7 mutations
244 from lineage G (*efpR*-E66K, RSc2277-V321G, *phbB*-S38F, pRALTA_0398-W719*, RSp0574–L15L,
245 RSp0769-F416S and RSp1422-G1298E) significantly improved this fitness component. Moreover, gains
246 in *in planta* proliferation were generally lower than gains in nodulation competitiveness and ranged
247 from 1.6 to 6.4-fold. The highest proliferation gains were produced by the five mutations RSc1598-
248 P455L (6.4-fold), *efpR*-E66K (3.7-fold), up^{115}-RSc0965 (2.6-fold), RSc2277-V321G (2.5 fold) and *rhlE1*-
249 L368L (2.4-fold), three of which also produced the highest nodulation gains.
250 To investigate whether improvements in nodulation competitiveness could be due to a better
251 colonization of the plant culture medium or rhizosphere, we measured the survival of the 28 mutants
252 in competition with their isogenic parental strain in these two compartments. Results from these
253 assays showed that none of the adaptive mutations improved bacterial ability to colonize the culture
254 medium and only 4 of them, including up^{115}-RSc0965, *efpR*-E66K and *rhlE1*-L368L, improved the
255 colonization of the rhizosphere (Extended Data Fig. 3 and Supplementary Table 10) although by smaller
256 factors than improvements in nodulation competitiveness. These results showed that improvements
257 in nodulation competitiveness are generally not associated with a better colonization of the culture
258 medium or of the rhizosphere, indicating that host entry is mainly controlled directly by the plant and

259 is the dominant selective force driving the adaptation of legume symbionts in this evolution
260 experiment.
261
262 **Evolutionary modelling predicts that selection for host entry drives adaptation**
263 The preferential selection of nodulation competitiveness in our experiment prompted us to investigate
264 the origin of this phenomenon. In particular, we wondered if stronger selection for host entry over
265 within-host proliferation could be a general feature of symbiotic life cycles, or, instead, if it is more
266 likely to be a specific property of our experimental system that may arise due to genetic constraints
267 on symbiotic traits and/or specific features of the selective regime. We used computer simulations to
268 model the evolution of bacterial populations that cycle between two compartments: the rhizosphere,
269 where bacteria accumulate mutations, and root nodules, where bacteria reproduce clonally (see
270 Supplementary Information for full details on the model and Supplementary Fig. 1). In this setting (and
271 similar to our experiment), bacterial fitness is mediated by two phenotypic components:
272 competitiveness to enter the host (resulting in nodulation) and within-host proliferation. The
273 nodulation step imposes a very stringent selective bottleneck during host entry, since (i) in most cases,
274 each nodule is founded by one single bacterium[9,27], (ii) there were usually between 100 and 300
275 nodules per lineage collected at each cycle (Supplementary Table 1) and (iii) bacterial genotypes differ
276 in their competitive ability to form nodules (Fig. 3b,e). Once inside the host, bacteria multiply to reach
277 a carrying capacity that is directly proportional to their proliferation fitness (no competition assumed),
278 before returning to the rhizosphere.
279 When following the evolution of bacterial phenotypes in populations founded by an ancestor with low
280 initial fitness (10^{-4} fold compared to the theoretical optimum, for each phenotypic component), we
281 observed, in most cases, a faster increase in nodulation competitiveness relative to within-host
282 proliferation (Fig. 4a). Accordingly, the distribution of mutations that were selected at the early steps
283 of the adaptive process was biased towards stronger improvement of nodulation over proliferation
284 (Fig. 4b, Low fitness domain). This dominance decreased, and was sometimes reverted, as populations
285 progressed towards high fitness values. Decreasing the strength of the nodulation bottleneck also
286 reduced the dominance of early selection for nodulation competitiveness, until abolishing it for the
287 largest bottleneck size tested (3000 nodules).
288 In our experiment, the ancestral nodulating strains had a much higher potential for improvement in
289 nodulation competitiveness relative to proliferation ($5x10^4$ vs. 10^2-10^3 fold, respectively, compared to
290 the natural symbiont *C. taiwanensis*) which could explain the faster evolution of the former trait. We
291 performed additional simulations to test how evolutionary trajectories are affected by the genotype
292 of the ancestor. As expected, starting from a higher proliferation fitness in the ancestor further
293 increased the overall fold effect of early selected mutations (Fig. 4c,d). In contrast, starting with a

294 higher initial nodulation competitiveness value of 10^{-3} or 10^{-2} in the ancestor was not sufficient to lead
295 to a stronger selection for proliferation, further confirming the asymmetry between selective
296 pressures acting on the two phenotypes. These findings were qualitatively robust to our assumptions
297 regarding the genetic architecture of the two symbiotic phenotypes (Supplementary Information,
298 Supplementary Fig. 2-11).



299

300 **Fig. 4 | Relative strength of selection for nodulation competitiveness and proliferation**

301 a,c,e, Median fitness trajectories of 100 simulated populations evolving under different sizes of nodulation bottleneck (10 to
302 3000 nodules). Points indicate fitness values at cycles 0, 10, 20, 30, 40 and 50. Dotted grey lines represent iso-fitness lines
303 (defined as the product of nodulation competitiveness and proliferation) used to delineate 3 fitness domains: Low (L; fitness
304 < 10-5), Mid (M; 10-5 < fitness < 10-3), and High (H; fitness > 10-3). The black dotted line represents the diagonal, along which
305 population would improve both phenotypic traits equally well. b,d,f, Fold effect (nodulation competitiveness effect divided
306 by proliferation effect) of mutations that reached a frequency of at least 30% in simulated populations. Color shading indicate

307 the fitness domain (Low, Mid or High) in which each group of mutations arose. Note that fold effects of mutations are
308 indicated even when median fitness trajectories do not reach the High fitness domain (e.g. 10 and 30 nodule lines in a)
309 because some individual replicate simulations (shown in Supplementary Figures) do reach this fitness domain. The effect of
310 different parameters on adaptive trajectories was tested: the size of nodulation bottleneck (a, b), the initial fitness values of
311 the ancestor (c, d), and the size of nodulation bottleneck under a scenario where the nodulation bottleneck occurs after
312 bacterial clonal proliferation (e, f).
313
314 Finally, we wondered if the chronology of symbiotic steps impacts on selective pressures and
315 evolutionary trajectories. We modified the simulations so that clonal multiplication of bacteria
316 (according to their proliferation fitness) occurs before the nodulation bottleneck. Although this model
317 is not relevant in the context of the legume-rhizobium symbiosis, it may apply to other symbiotic or
318 pathogenic life cycles, for example when a selective bottleneck occurs during dissemination in the
319 environment or within the host[28,29] or in the case of vertically transmitted symbioses[30]. In this context,
320 the dominance of nodulation competitiveness was strongly reduced, bringing the two selective forces
321 close to equilibrium (Fig. 4e,f and Supplementary Fig. 12), in agreement with the expectation that each
322 fitness component should have a similar influence on adaptation when they have a multiplicative effect
323 on global fitness. Therefore, the dominance of nodulation competitiveness in the previous simulation
324 runs is likely a consequence of the selective bottleneck occurring just after the transient
325 hypermutagenesis phase in the rhizosphere.
326 Altogether, these simulations show that the selective pressures that are experienced by symbiotic
327 bacteria are asymmetric, with a dominance of selection for nodulation competitiveness over within-
328 host proliferation. This asymmetry occurs when a highly diverse bacterial population from the
329 rhizosphere is exposed to the strong, selective nodulation bottleneck, allowing the efficient selection
330 of the most competitive clones by the host plant. In our evolution experiment, the fact that the
331 ancestor has a lower nodulation competitiveness value compared to proliferation (relative to that of
332 the wild-type rhizobium *C. taiwanensis*) is an additional factor expected to have contributed to the
333 stronger selection on nodulation competitiveness.
334
335 **Discussion**
336 Horizontally transmitted symbiotic bacteria have complex lifecycles during which they have to face
337 multiple environmental constraints, both outside and inside the host. Assessing the relative influence
338 of each of these selective pressures on the adaptive trajectories of emerging symbiotic bacteria is
339 critically needed to better understand the eco-evolutionary dynamics of microbial populations[5].
340 Here we show that selection on competitiveness for host entry is the dominant selective pressure
341 shaping early adaptation of new rhizobia. This trait, mainly mediated by the plant, improved very fast
342 during our evolution experiment, to a level comparable to that of the *Mimosa* rhizobium

343 *C. taiwanensis*, and all adaptive mutations identified in lineages B and G improved nodulation
344 competitiveness. Evolutionary simulations led us to propose that the dominance of the selection on
345 host entry over the selection on within-host proliferation that we observed in our system is strongly
346 dependent on the strength of the selective nodulation bottleneck. These results shed light on the
347 importance of selective bottlenecks in shaping the evolutionary trajectories of emerging microbial
348 symbionts. Most theoretical[31,32] and experimental[33-38] works on the influence of bottlenecks on
349 microbial adaptation have focused on non-selective bottlenecks that randomly purge genetic diversity
350 and reduce the efficiency of natural selection. Yet, another aspect of bottlenecks emerges when
351 considering that transmission and host colonization can be, at least partially, dependent on microbial
352 genotype, prompting us to consider infection bottlenecks as selective events[39]. Selective bottlenecks
353 have already been described during virus infections[28,40-43] but received less attention in bacteria, except
354 for their role in the evolution of phenotypic heterogeneity[17,44-46]. However selective bottlenecks are
355 likely common both in mutualistic and pathogenic interactions. Indeed, potentially all symbiotic
356 microorganisms (whether commensal, pathogenic or mutualists) experience bottlenecks of various
357 intensities (but sometimes as strong as only a few cells[47]) during their life cycles[48,49]. Given that
358 microbes rely on complex genetic networks to disseminate and infect their hosts[7,50-53], one can expect
359 these traits to display extensive genetic and phenotypic variation within populations and evolve
360 through competition with other members of the microbiome. Interestingly, strong selection for host
361 entry in experimentally evolved symbionts was recently observed in several studies[54-56], reinforcing the
362 idea that selective bottlenecks for host entry may play a crucial role in the evolution of diverse
363 symbiotic interactions.
364 The molecular dissection of bacterial adaptation confirmed that the genetic architecture of rhizobial
365 symbiotic traits is complex, as it involves many different genes and physiological functions [57,58]. Our
366 study also unveiled a new characteristic of the symbiotic genotype-to-phenotype map: the recurrent
367 genetic coupling between nodulation competitiveness and within-host proliferation. Indeed, all the 10
368 identified mutations that improved within-host proliferation also improved nodulation competiveness,
369 while the reverse was not true. This result corroborates our previous data showing that mutations
370 promoting nodule cell infection always improved nodule formation[14,59]. It is conceivable that the
371 intracellular release and proliferation of rhizobia in nodule cells is dependent on the efficiency of the
372 earliest symbiotic events, *i. e*. the entry and progression of bacteria in infection threads and the
373 concomitant divisions of nodule cells preparing for the accommodation of bacteria. Consistently,
374 delayed progression of infection threads across the root cell layers was shown to impair the release of
375 bacteria in nodule cells[60]. Moreover plant receptors (LYK3, NFR1, Sym37, SYMRK) and transcription
376 factors (NIN, ERN1, ERN2) involved in both nodule organogenesis and rhizobial intracellular infection
377 were identified[61-64], suggesting that common mechanisms controlling bacterial invasion are exerted by

378 the plant at the two entry points, the root hairs (host entry) and the nodule cells (within-host
379 proliferation). Yet, this coupling is not strict, since some mutations increase nodulation
380 competitiveness without increasing proliferation. These mutations may target bacterial functions
381 specifically required for nodulation. Alternatively, the fact that most mutations with an effect on
382 proliferation are the ones with the highest contribution to nodulation competitiveness could be
383 explained by a threshold effect, *i.e.* the existence of a minimal level above which mutations that
384 increase nodulation competitiveness also increase proliferation. Surprisingly, mutations only
385 improving proliferation were not found in our study. At this stage, we cannot distinguish between two
386 hypotheses, namely if these mutations do not exist (proliferation is always linked to nodulation and
387 host entry) or if they were not yet selected in this experiment due to the dominance of nodulation
388 competitiveness in the first steps of adaptation. Future functional analyses of the adaptive mutations
389 identified in this study will expand our understanding of the molecular bases of nodulation
390 competitiveness in rhizobia[65-67], and its relationship to within-host proliferation. Interestingly, these
391 two traits primarily selected by the plant, although necessary for the symbiotic interaction, are not, in
392 themselves, beneficial for the host. Including nitrogen fixation and host growth promotion in the
393 analysis of genetic couplings between the different symbiotic traits, will be fundamental to better
394 understand the evolution of mutualism[67].
395 In all lineages of this experiment, the rate of adaptation was very high during the first cycles of
396 evolution and then tended to decrease over time. This classical pattern of evolution, due to diminishing
397 return epistasis among beneficial mutations[24,68,69], was likely favored by the low initial fitness values of
398 nodulating ancestors compared to the natural symbiont *C. taiwanensis* allowing the rapid acquisition
399 of highly beneficial mutations during the first cycles (Table 1). Elevated mutation rate in the
400 rhizosphere, prior to the entry of bacteria in roots[18], has also played an important role in the dynamic
401 of adaptation by fueling the extensive genetic diversification of bacteria and exposing these
402 populations to plant-mediated selection. A probable consequence of high mutation rate combined to
403 strong selection is the presence of large cohorts of mutations that carry multiple beneficial mutations
404 along with some neutral or slightly deleterious ones. Mutational cohorts were already described in
405 previous evolution experiments[21,70], with cases of cohorts carrying 2 co-driver mutations[22]. Here we
406 identified up to 7 adaptive mutations in one of our cohorts. The occurrence of multi-driver mutations
407 per cohort might be explained by the nested, sequential fixation of multiple adaptive mutations in one
408 lineage before it reaches detectable frequency (5% in our case)[22]. This phenomenon, creating a
409 'travelling wave' of adaptation[71], was recently observed with high-resolution sequencing[23] and is
410 possibly amplified by the strong selection/high mutation regime of our experiment. Alternatively, it
411 was proposed that, in the case of hypermutagenesis, the likelihood of multiple adaptive mutations

412 arising simultaneously in a given genome becomes non-negligible and might promote saltational
413 evolution[72].
414 In conclusion, our work shows that the relative strengths of the selective forces imposed by hosts
415 strongly influence the evolutionary trajectory of microbial populations. Selection for host entry is likely
416 the predominant life-history trait selected during the evolution of new horizontally-transmitted
417 symbiotic interactions, such as the rhizobium-legume symbiosis. This effect emerges from the
418 presence of a selective bottleneck at host entry, although its predominance can be modulated by other
419 genetic and ecological factors. Investigating the role of selective bottlenecks in other symbiotic
420 interactions will improve our ability to manage and predict the eco-evolutionary dynamics of host-
421 associated microbial populations.
422
423 **Methods**
424
425 **Bacterial strains and growth conditions**
426 Strains and plasmids used in this study are listed in Supplementary Table 11. *Cupriavidus taiwanensis*
427 strains were grown at 28°C on tryptone-yeast (TY) medium[73]. *Ralstonia solanacearum* strains were
428 grown at 28°C either on BG medium[74] or on minimal MP medium[75] supplemented with 2% glycerol.
429 *Escherichia coli* strains were grown at 37°C on Luria-Bertani medium[76]. Antibiotics were used at the
430 following concentrations: trimethoprim at 100 µg ml^{-1}, spectinomycin at 40 µg ml^{-1}, streptomycin at
431 200 µg ml^{-1} and kanamycin at 25 µg ml^{-1} (for *E. coli*) or 50 µg ml^{-1} (for *R. solanacearum*).
432
433 **Experimental evolution**
434 Symbiotically evolved clones and populations were generated as previously described[16]. Five lineages,
435 two (B and F) derived from the CBM212 ancestor, two (G and K) derived from the CBM349 ancestor
436 and one (M) derived from the CBM356 ancestor, previously evolved for 16 cycles were further evolved
437 until cycle 35. For each lineage, at each cycle, 30 plants were inoculated with nodule bacterial
438 populations from the previous cycle. Twenty-one days after inoculation, all nodules from the 30 plants
439 were pooled, surface-sterilized with 2.6% sodium hypochlorite for 15 min, rinsed with sterile water
440 and crushed. Then, 10% of the nodule crush was used to inoculate a new set of 30 plants the same
441 day. Serial dilutions of each nodule crush were plated and one clone was randomly selected from the
442 highest dilution and purified. The selected clones, the rest of the nodule crushes, and a 24-h culture of
443 bacteria from an aliquot of nodule crushes were stored at –80°C at each cycle.
444
445 **Sequencing bacterial populations and clones**

446 Aliquots of frozen nodule bacterial populations or single colonies from purified clones were grown
447 overnight in BG medium supplemented with trimethoprim. Bacterial DNA was extracted from 1 ml of
448 culture using the Wizard genomic DNA purification kit (Promega). Evolved population DNAs were
449 sequenced at the GeT-PlaGe core facility (https://get.genotoul.fr/), INRAE Toulouse. DNA-seq libraries
450 have been prepared according to Illumina's protocols using the Illumina TruSeq Nano DNA HT Library
451 Prep Kit. Briefly, DNA was fragmented by sonication, size selection was performed using SPB beads (kit
452 beads) and adaptators were ligated to be sequenced. Library quality was assessed using an Advanced
453 Analytical Fragment Analyzer and libraries were quantified by qPCR using the Kapa Library
454 Quantification Kit (Roche). Sequencing has been performed on a NovaSeq6000 S4 lane (Illumina,
455 California, USA) using a paired-end read length of 2x150 pb with the Illumina NovaSeq Reagent Kits.
456 Evolved clones were re-sequenced either by C.E.A/IG/Genoscope using the Illumina GA2X technology
457 (clones B8, B16, F16, G8, G16, K16 and M16), or by the GeT_PlaGe core facility using the Illumina
458 technology HiSeq2000 (clones B1, B2, B3, B4, B5, B6, B7, B9, B10, B11, B12, B13, B14, B15, G1, G2, G3,
459 G4, G5, G6, G7, G9, G10, G11, G12, G13, G14, and G15), or HiSeq3000 (clones B20, B25, B30, G25 and
460 G30), or NovaSeq6000 (clones B35, F35, G35, K35 and M35).
461
462 **Detection of mutations and molecular analyses**
463 Sequencing reads from NovaSeq6000 runs (whole-populations and clones from cycle 35) were mapped
464 on the chimeric reference genome of the ancestral strain, comprising *R. solanacearum* GMI1000
465 chromosome (GenBank accession number: NC_003295.1) and megaplasmid (NC_003296.1) together
466 with *C. taiwanensis* symbiotic plasmid pRalta (CU633751). Mutations were detected using breseq
467 v0.33.1(Ref. [77]) with default parameters, either using the polymorphism mode (for whole-population
468 sequences) or the consensus mode (for individual clones). Mutation lists were curated manually in
469 order to remove mutations present in the ancestral strains as well as false positive hits arising from
470 reads misalignments in low complexity regions. Moreover, alleles showing aberrant trajectories in the
471 time-course whole-population sequencing data were checked manually, by inspecting either breseq
472 output files and/or read alignments with IGV[78], and corrected as needed. In this work, we focused our
473 attention on SNPs and indels detected above 5% frequency in the populations. Recombinations,
474 rearrangements and IS movements were not analyzed exhaustively. Mutations detected in populations
475 are listed in Supplementary Tables 4-8.
476 Sequencing data of the remaining clones were analyzed as described previously[26] with the PALOMA
477 bioinformatics pipeline implemented in the Microscope platform[79]. The complete list of mutational
478 events generated for these clones are available on the Microscope platform
479 (https://mage.genoscope.cns.fr/microscope/expdata/NGSProjectEvo.php, SYMPA tag). Mutations
480 detected in clones are listed in Supplementary Tables 12-16.

481 In order to identify mutations forming temporal clusters (cohorts) through populations, we selected
482 mutations having a frequency of 30% in at least one cycle and clustered their frequency using the *hclust*
483 function in R v3.6.1[80]. Then, we separated cohorts using a cutoff distance of 0.3.
484 Sub-population genealogies in lineages B and G, shown as Muller plots in Extended Data Figure 2, were
485 reconstructed by comparing mutations found in cohorts and in individually sequenced clones. Cohorts
486 for which ancestry cannot be ascertained (mutations not found in any clone) were not included in
487 these plots. Relative frequencies of genotypes were calculated manually and plotted with ggMuller[81].
488 G scores were calculated as described[20]. Since we found that synonymous mutations can be adaptive,
489 we used all types of mutations in coding regions to calculate this statistics. Moreover, we included all
490 mutations beyond 5% frequency in this analysis since we assumed that strong clonal interference may
491 prevent adaptive mutations to rise to high frequency. To evaluate statistical significance of G scores,
492 we ran 1,000 simulations where the total number of mutations used for the calculation of the observed
493 G scores (3330) were randomly attributed in the coding genes according to their respective length to
494 compute simulated G scores for each bacterial gene. The sum of simulated G scores was compared to
495 the observed sum and we computed a Z score and *P*-value from these simulated G statistics. These
496 simulations were also used to compute mean G scores for each gene, and to calculate the associated
497 Z scores and *P*-values (adjusted using a Bonferroni correction).
498
499 **Reconstruction of mutations in evolved clones**
500 Mutant alleles and constitutively expressed reporter genes (GFPuv, mCherry) were introduced into
501 *Ralstonia* evolved clones by co-transformation using the MuGent technique[26,82]. Briefly, two DNA
502 fragments, the first one carrying an antibiotic (kanamycin) resistance gene prepared from the pRCK-
503 P*ps*-GFP or pRCK-P*ps*-mCherry linearized plasmids allowing the integration of the resistance gene in
504 the intergenic region downstream the *glmS* gene, and the second one carrying the mutation to be
505 introduced prepared by PCR amplification of a 6 kb region using genomic DNA of evolved clones as
506 template and high fidelity Phusion polymerase (ThermoFisher Scientific), were co-transformed into
507 naturally competents cells of *Ralstonia* evolved clones. Co-transformants resistant to kanamycin were
508 screened by PCR using primers specifically amplifying mutant or wild-type alleles and verified by Sanger
509 sequencing. Primers used in mutation reconstructions are listed in Supplementary Table 17.
510
511 **Analyses of symbiotic phenotypes on *M. pudica***
512 *Mimosa pudica* seeds from LIPME production (2019 batch) were sterilized as described[83] by immersion
513 in 95% H_2SO_4 during 15 minutes and 2.4% sodium hypochlorite solution during 10 minutes and rinsed
514 in sterile distilled water 5 times. Seeds were soaked in sterile water at 28°C under agitation for 24
515 hours and then deposited on soft agar (9.375 g/L) and incubated at 28°C during 24 more hours in

516 darkness. Then, seedlings were cultivated in glass tubes (2 seedlings/tube) under N-free conditions,
517 each tube containing 20 mL of solid Fahraeus medium[84] and 40 mL of liquid Jensen medium[85] diluted
518 1:4 with sterile water. Plants were incubated in a culture chamber at 28°C under a photoperiod
519 day/night of 16 h/8 h.
520 For *in planta* relative fitness and nodulation competitiveness assays, two strains of *Ralstonia*
521 *solanacearum* expressing differential constitutive fluorophores (GFPuv or mCherry) or one strain of *R.*
522 *solanacearum* and one strain of *C. taiwanensis*, both expressing differential antibiotic resistance
523 (streptomycin and trimethoprim), were co-inoculated onto *M. pudica* plantlets grown for 3 to 4 days
524 in the culture chamber. Both strains were inoculated in equivalent proportion (*ca*. $5x10^5$ bacteria of
525 each strain per plant) except for the comparisons of the *Ralstonia* nodulating ancestors with *C.*
526 *taiwanensis* where strains were mixed in a 1000:1 ratio (*ca.* $5x10^3$ bacteria of *C. taiwanensis* and *ca.*
527 $5x10^5$ bacteria of nodulating ancestors per plant). Nodules were harvested 21 days after inoculation,
528 surface sterilized by immersion in 2.4% sodium hypochlorite solution for 15 minutes and rinsed with
529 sterile water. For *in planta* relative fitness measurements, sterilized nodules from 20 plants (see the
530 exact number of harvested nodules in Supplementary Table 1) were pooled and crushed in 1 mL of
531 sterile water, diluted and spread on selective solid medium using an easySpiral automatic plater
532 (Interscience). After 2-day incubation at 28°C, colonies were screened either by plating on selective
533 medium in case of co-inoculations of *Ralstonia* evolved clones with *C. taiwanensis* or based on
534 fluorescence using a stereo zoom microscope (AxioZoom V16, Zeiss, Oberkochen, Germany) for co-
535 inoculations of *Ralstonia* evolved clones with *Ralstonia* reconstructed mutants. For nodulation
536 competitiveness assays, *ca*. 96 individual nodules per experiment were crushed separately in 96-well
537 microtiter plates and droplets were deposited on selective medium. Nodule occupancy was
538 determined by screening bacteria grown in the droplets either on selective medium or based on their
539 fluorescence as described for fitness measurements. Both nodulation competitiveness and relative
540 fitness assays were measured in at least three independent experiments. Competitive indexes (CI)
541 were calculated by dividing the ratio of the number of test strain (evolved strains or reconstructed
542 mutants) vs. reference strain (*C. taiwanensis* or evolved parental strains, respectively) in nodules
543 normalized by the inoculum ratio. CI values above 1 were transformed by their inverse and compared
544 to the value 1 using a one-sided Student *t*-test ($P<0.05$).
545 For within-host proliferation assays, plantlets were inoculated with a single strain ($5x10^5$ bacteria per
546 plant). In each experiment, nodules from 6 individual plants were collected separately 21 days after
547 inoculation, surface sterilized for 15 minutes in a 2.4% sodium hypochlorite solution, rinsed with sterile
548 water and crushed in 1 ml of sterile water. Dilutions of nodule crushes were plated on selective solid
549 medium using an easySpiral automatic plater (Interscience). Two days after incubation at 28°C, the
550 number of colonies was counted. Within-host proliferation was estimated as the number of bacteria

551 per nodule. For each strain, 15 to 24 measurements from three independent experiments were
552 performed. Pairwise comparisons of proliferation values were compared using a two-sided Wilcoxon
553 rank sum test.
554 For Jensen culture medium and rhizosphere colonization assays, plantlets were co-inoculated with pair
555 of strains expressing differential constitutive fluorophores (GFPuv or mCherry) in equivalent
556 proportion ($5x10^6$ bacteria of each strain per plant). Seven days after inoculation, bacteria present in
557 the culture medium were diluted and plated on selective solid medium using an easySpiral automatic
558 plater (Interscience). Bacteria attached to the roots were resuspended in 4 ml of sterile water by
559 strongly vortexing for 1 minute, diluting and plating on selective medium using an easySpiral automatic
560 plater (Interscience). Two days after incubation at 28°C, colonies were screened for fluorescence using
561 a stereo zoom microscope (AxioZoom V16, Zeiss, Oberkochen, Germany). Competitive indexes (CI) for
562 Jensen medium or rhizosphere colonization were calculated as described above. CI values were
563 compared to the value 1 using a two-sided Wilcoxon rank sum test with Benjamini-Hochberg
564 correction.
565
566 **Modelling/simulations**
567 Full details on evolutionary simulations are available as Supplementary Information.
568
569 **Data availability**
570 Sequencing data are available under NCBI SRA BioProject ID PRJNA788708 and SRP353965. Raw
571 experimental data are available as Supplementary Information. Raw data generated from computer
572 simulations are deposited on the Data INRAE dataverse (https://doi.org/10.15454/QYB2S9).
573
574 **Code availability**
575 Computer code used to analyze genomic data and to perform computer simulations are deposited on
576 the Data INRAE dataverse (https://doi.org/10.15454/QYB2S9).
577
578 **References**
579 1. Drew, G. C., Stevens, E. J. & King, K. C. Microbial evolution and transitions along the parasite-
580 mutualist continuum. *Nat. Rev. Microbiol*. **19**, 623-638, doi:10.1038/s41579-021-00550-7 (2021).
581 2. Nandasena, K. G., O'Hara, G. W., Tiwari, R. P., Sezmis, E. & Howieson, J. G. *In situ* lateral transfer
582 of symbiosis islands results in rapid evolution of diverse competitive strains of mesorhizobia
583 suboptimal in symbiotic nitrogen fixation on the pasture legume *Biserrula pelecinus* L. *Environ.*
584 *Microbiol.* **9**, 2496-2511, doi:10.1111/j.1462-2920.2007.01368.x (2007).

585 3. Longdon, B. et al. The causes and consequences of changes in virulence following pathogen host
586 shifts. *PLoS Pathog*. **11**, e1004728, doi:10.1371/journal.ppat.1004728 (2015).
587 4. Bonneaud, C. & Longdon, B. Emerging pathogen evolution: using evolutionary theory to
588 understand the fate of novel infectious pathogens. *EMBO Rep.* **21**, e51374,
589 doi:10.15252/embr.202051374 (2020).
590 5. Obeng, N., Bansept, F., Sieber, M., Traulsen, A. & Schulenburg, H. Evolution of microbiota-host
591 associations: the microbe's perspective. *Trends Microbiol*. **29**, 779-787,
592 doi:10.1016/j.tim.2021.02.005 (2021).
593 6. Robinson, C. D., Bohannan, B. J. & Britton, R. A. Scales of persistence: transmission and the
594 microbiome. *Curr. Opin. Microbiol*. **50**, 42-49, doi:10.1016/j.mib.2019.09.009 (2019).
595 7. Wheatley, R. M. et al. Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. *Proc.*
596 *Natl. Acad. Sci. U. S. A*. **117**, 23823-23834, doi:10.1073/pnas.2009094117 (2020).
597 8. Gage, D. J. Analysis of infection thread development using Gfp- and DsRed-expressing
598 *Sinorhizobium meliloti*. *J. Bacteriol.* **184**, 7042-7046, doi:10.1128/jb.184.24.7042-7046.2002
599 (2002).
600 9. Gage, D. J. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation
601 of temperate legumes. *Microbiol. Mol. Biol. Rev*. **68**, 280-300, doi:10.1128/mmbr.68.2.280-
602 300.2004 (2004).
603 10. Coba de la Peña, T., Fedorova, E., Pueyo, J. J. & Lucas, M. M. The symbiosome: legume and rhizobia
604 co-evolution toward a nitrogen-fixing organelle? *Front. Plant Sci.* **8**, 2229,
605 doi:10.3389/fpls.2017.02229 (2017).
606 11. Hirsch, A. M. et al. *Rhizobium meliloti* nodulation genes allow *Agrobacterium tumefaciens* and
607 *Escherichia coli* to form pseudonodules on alfalfa. *J. Bacteriol*. **158**, 1133-1143 (1984).
608 12. Abe, M. et al. Transfer of the symbiotic plasmid from *Rhizobium leguminosarum* biovar *trifolii* to
609 *Agrobacterium tumefaciens*. *J. Gen. Appl. Microbiol*. **44**, 65-74 (1998).
610 13. Nakatsukasa, H. et al. Transposon mediation allows a symbiotic plasmid of *Rhizobium*
611 *leguminosarum* bv. *trifolii* to become a symbiosis island in *Agrobacterium* and *Rhizobium*. *J. Gen.*
612 *Appl. Microbiol.* **54**, 107-118, doi:10.2323/jgam.54.107 (2008).
613 14. Marchetti, M. et al. Experimental evolution of a plant pathogen into a legume symbiont. *PLoS Biol*.
614 **8**, doi:e1000280 10.1371/journal.pbio.1000280 (2010).
615 15. Masson-Boivin, C., Giraud, E., Perret, X. & Batut, J. Establishing nitrogen-fixing symbiosis with
616 legumes: how many rhizobium recipes? *Trends Microbiol.* **17**, 458-466,
617 doi:10.1016/j.tim.2009.07.004 (2009).

618 16. Marchetti, M. et al. Experimental evolution of rhizobia may lead to either extra- or intracellular
619 symbiotic adaptation depending on the selection regime. *Mol. Ecol.* **26**, 1818-1831,
620 doi:10.1111/mec.13895 (2017).
621 17. Moxon, R. & Kussell, E. The impact of bottlenecks on microbial survival, adaptation, and
622 phenotypic switching in host-pathogen interactions. *Evolution* **71**, 2803-2816,
623 doi:10.1111/evo.13370 (2017).
624 18. Remigi, P. et al. Transient hypermutagenesis accelerates the evolution of legume endosymbionts
625 following horizontal gene transfer. *PLoS Biol.* **12**, e1001942, doi:10.1371/journal.pbio.1001942
626 (2014).
627 19. Lynch, M. et al. Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet*.
628 **17**, 704-714, doi:10.1038/nrg.2016.104 (2016).
629 20. Tenaillon, O. et al. Tempo and mode of genome evolution in a 50,000-generation experiment.
630 *Nature* **536**, 165-170, doi:10.1038/nature18959 (2016).
631 21. Lang, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast
632 populations. *Nature* **500**, 571-574, doi:10.1038/nature12344 (2013).
633 22. Buskirk, S. W., Peace, R. E. & Lang, G. I. Hitchhiking and epistasis give rise to cohort dynamics in
634 adapting populations. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 8330-8335,
635 doi:10.1073/pnas.1702314114 (2017).
636 23. Nguyen Ba, A. N. et al. High-resolution lineage tracking reveals travelling wave of adaptation in
637 laboratory yeast. *Nature* **575**, 494-499, doi:10.1038/s41586-019-1749-3 (2019).
638 24. McDonald, M. J. Microbial experimental evolution - a proving ground for evolutionary theory and
639 a tool for discovery. *EMBO Rep*. **20**, e46992, doi:10.15252/embr.201846992 (2019).
640 25. Bailey, S. F., Alonso Morales, L. A. & Kassen, R. Effects of synonymous mutations beyond codon
641 bias: the evidence for adaptive synonymous substitutions from microbial evolution experiments.
642 Genome *Biol. Evol*. **13**, doi:10.1093/gbe/evab141 (2021).
643 26. Capela, D. et al. Recruitment of a lineage-specific virulence regulatory pathway promotes
644 intracellular infection by a plant pathogen experimentally evolved into a legume symbiont. *Mol.*
645 *Biol. Evol.* **34**, 2503-2521, doi:10.1093/molbev/msx165 (2017).
646 27. Daubech, B. et al. Spatio-temporal control of mutualism in legumes helps spread symbiotic
647 nitrogen fixation. *eLife* **6**, doi:10.7554/eLife.28683 (2017).
648 28. Joseph, S. B., Swanstrom, R., Kashuba, A. D. & Cohen, M. S. Bottlenecks in HIV-1 transmission:
649 insights from the study of founder viruses. *Nat. Rev. Microbiol*. **13**, 414-425,
650 doi:10.1038/nrmicro3471 (2015).

651 29. Hullahalli, K. & Waldor, M. K. Pathogen clonal expansion underlies multiorgan dissemination and
652 organ-specific outcomes during murine systemic infection. *eLife* **10**, doi:10.7554/eLife.70910
653 (2021).
654 30. Perreau, J., Zhang, B., Maeda, G. P., Kirkpatrick, M. & Moran, N. A. Strong within-host selection in
655 a maternally inherited obligate symbiont: *Buchnera* and aphids. *Proc. Natl. Acad. Sci. U. S. A*. **118**,
656 doi:10.1073/pnas.2102467118 (2021).
657 31. Bergstrom, C. T., McElhany, P. & Real, L. A. Transmission bottlenecks as determinants of virulence
658 in rapidly evolving pathogens. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 5095-5100,
659 doi:10.1073/pnas.96.9.5095 (1999).
660 32. LeClair, J. & Wahl, L. The impact of population bottlenecks on microbial adaptation. *J. Stat. Physics*
661 **172**, 114–125, doi:10.1007/s10955-017-1924-6 (2018).
662 33. Izutsu, M., Lake, D., Matson, Z., Dodson, J. & Lenski, R. Effects of periodic bottlenecks on the
663 dynamics of adaptive evolution in microbial populations. *bioRxiv* doi:10.1101/2021.12.29.474457
664 (2022).
665 34. Wein, T. & Dagan, T. The Effect of population bottleneck size and selective regime on genetic
666 diversity and evolvability in bacteria. *Genome Biol. Evol.* **11**, 3283-3290, doi:10.1093/gbe/evz243
667 (2019).
668 35. Garoff, L. et al. Population bottlenecks strongly influence the evolutionary trajectory to
669 fluoroquinolone resistance in *Escherichia coli*. *Mol. Biol. Evol.* **37**, 1637-1646,
670 doi:10.1093/molbev/msaa032 (2020).
671 36. Mahrt, N. et al. Bottleneck size and selection level reproducibly impact evolution of antibiotic
672 resistance. *Nat. Ecol. Evol*. **5**, 1233-1242, doi:10.1038/s41559-021-01511-2 (2021).
673 37. Windels, E. M. et al. Population bottlenecks strongly affect the evolutionary dynamics of antibiotic
674 persistence. *Mol. Biol. Evol*. **38**, 3345-3357, doi:10.1093/molbev/msab107 (2021).
675 38. Sousa, A. et al. Recurrent Reverse Evolution Maintains Polymorphism after Strong Bottlenecks in
676 Commensal Gut Bacteria. *Mol. Biol. Evol*. **34**, 2879-2892, doi:10.1093/molbev/msx221 (2017).
677 39. Handel, A. & Bennett, M. R. Surviving the bottleneck: transmission mutants and the evolution of
678 microbial populations. *Genetics* **180**, 2193-2200, doi:10.1534/genetics.108.093013 (2008).
679 40. Wilker, P. R. et al. Selection on haemagglutinin imposes a bottleneck during mammalian
680 transmission of reassortant H5N1 *Influenza* viruses. *Nat. Commun*. **4**, 2636,
681 doi:10.1038/ncomms3636 (2013).
682 41. Moncla, L. H. et al. Selective bottlenecks shape evolutionary pathways taken during mammalian
683 adaptation of a 1918-like avian *Influenza* virus. *Cell Host Microbe* **19**, 169-180,
684 doi:10.1016/j.chom.2016.01.011 (2016).

685 42. Sobel Leonard, A. et al. Deep sequencing of *Influenza* a virus from a human challenge study reveals
686 a selective bottleneck and only limited intrahost genetic diversification. *J. Virol.* **90**, 11247-11258,
687 doi:10.1128/JVI.01657-16 (2016).
688 43. Huang, S. W. et al. A selective bottleneck shapes the evolutionary mutant spectra of Enterovirus
689 A71 during viral dissemination in humans. *J. Virol.* **91**, doi:10.1128/JVI.01062-17 (2017).
690 44. Libby, E. & Rainey, P. B. Exclusion rules, bottlenecks and the evolution of stochastic phenotype
691 switching. *Proc. Biol. Sci.* **278**, 3574-3583, doi:10.1098/rspb.2011.0146 (2011).
692 45. De Ste Croix, M. et al. Selective and non-selective bottlenecks as drivers of the evolution of
693 hypermutable bacterial loci. *Mol. Microbiol*. 113, 672-681, doi:10.1111/mmi.14453 (2020).
694 46. Schmutzer, M. & Wagner, A. Gene expression noise can promote the fixation of beneficial
695 mutations in fluctuating environments. *PLoS Comput. Biol.* **16**, e1007727,
696 doi:10.1371/journal.pcbi.1007727 (2020).
697 47. Gama, J. A., Abby, S. S., Vieira-Silva, S., Dionisio, F. & Rocha, E. P. Immune subversion and quorum-
698 sensing shape the variation in infectious dose among bacterial pathogens. *PLoS Pathog*. **8**,
699 e1002503, doi:10.1371/journal.ppat.1002503 (2012).
700 48. Abel, S., zur Wiesch, PA, Davis, B. & Waldor, M. Analysis of bottlenecks in experimental models of
701 infection. *PLoS Pathogen* **11**(6):e1004823doi:10.1371/journal.ppat.1004823 (2015).
702 49. Papkou & A Gokhale, C. T., A Schulenburg, H. Host-parasite coevolution: why changing population
703 size matters. *Zoology* **117**, 330-338, doi:10.1016/j.zool.2016.02.001 (2016).
704 50. Cossart, P. & Sansonetti, P. J. Bacterial invasion: the paradigms of enteroinvasive pathogens.
705 *Science* **304**, 242-248, doi:10.1126/science.1090124 (2004).
706 51 Visick, K. L., Stabb, E. V. & Ruby, E. G. A lasting symbiosis: how *Vibrio fischeri* finds a squid partner
707 and persists within its natural host. *Nat. Rev. Microbiol.* **19**, 654-665, doi:10.1038/s41579-021-
708 00557-0 (2021).
709 52. Rowe, H. M. et al. Bacterial factors required for transmission of *Streptococcus pneumoniae* in
710 mammalian hosts. Cell Host Microbe **25**, 884-891.e886, doi:10.1016/j.chom.2019.04.012 (2019).
711 53. Bright, M. & Bulgheresi, S. A complex journey: transmission of microbial symbionts. *Nat. Rev.*
712 *Microbiol.* **8**, 218-230, doi:10.1038/nrmicro2262 (2010).
713 54. Robinson, C. D. et al. Experimental bacterial adaptation to the zebrafish gut reveals a primary role
714 for immigration. *PLoS Biol*. **16**, e2006893, doi:10.1371/journal.pbio.2006893 (2018).
715 55. Bacigalupe, R., Tormo-Mas, M., Penadés, J. R. & Fitzgerald, J. R. A multihost bacterial pathogen
716 overcomes continuous population bottlenecks to adapt to new host species. *Sci. Adv.* **5**, eaax0063,
717 doi:10.1126/sciadv.aax0063 (2019).

718 56. Batstone, R. T., O'Brien, A. M., Harrison, T. L. & Frederickson, M. E. Experimental evolution makes
719 microbes more cooperative with their local host genotype. *Science* **370**, 476-478,
720 doi:10.1126/science.abb7222 (2020).
721 57. Pobigaylo, N., Szymczak, S., Nattkemper, T. W. & Becker, A. Identification of genes relevant to
722 symbiosis and competitiveness in *Sinorhizobium meliloti* using signature-tagged mutants. *Mol.*
723 *Plant Microbe Interact.* **21**, 219-231, doi:10.1094/MPMI-21-2-0219 (2008).
724 58. Flores-Tinoco, C. E., Christen, M. & Christen, B. Co-catabolism of arginine and succinate drives
725 symbiotic nitrogen fixation. *Mol. Syst. Biol.* 2020 **16**, e9419. doi: 10.15252/msb.20199419. (2019)
726 59. Guan, S. H. et al. Experimental evolution of nodule intracellular infection in legume symbionts.
727 *ISME J*. **7**, 1367-1377, doi:10.1038/ismej.2013.24 (2013).
728 60. Xiao, T. T. et al. Fate map of *Medicago truncatula* root nodules. *Development* **141**, 3517-3528,
729 doi:10.1242/dev.110775 (2014).
730 61. Moling, S. et al. Nod factor receptors form heteromeric complexes and are essential for
731 intracellular infection in medicago nodules. *Plant Cell* **26**, 4188-4199, doi:10.1105/tpc.114.129502
732 (2014).
733 62. Cerri, M. R. et al. The symbiosis-related ERN transcription factors act in concert to coordinate
734 rhizobial host root infection. *Plant Physiol*. **171**, 1037-1054, doi:10.1104/pp.16.00230 (2016).
735 63. Liu, J. *et al.* NIN is essential for development of symbiosomes, suppression of defence and
736 premature senescence in *Medicago truncatula* nodules. *New Phytol.* **230**, 290-303,
737 doi:10.1111/nph.17215 (2021).
738 64. Capoen, W., Goormachtig, S., De Rycke, R., Schroeyers, K. & Holsters, M. SrSymRK, a plant
739 receptor essential for symbiosome formation. *Proc Natl Acad Sci U S A* **102**, 10369-10374,
740 doi:10.1073/pnas.0504250102 (2005).
741 65. Mendoza-Suárez, M., Andersen, S. U., Poole, P. S. & Sánchez-Cañizares, C. Competition, nodule
742 occupancy, and persistence of inoculant strains: Key Factors in the in the rhizobium-legume
743 symbioses. *Front. Plant Sci.* **12**, 690567, doi:10.3389/fpls.2021.690567 (2021).
744 66. Bovin, S. & Lepetit, M. Partner preference in the legume-rhizobia symbiosis and impact on legume
745 inoculation strategies. *Advances in Botanical Research*, **94**, 323-348,
746 doi:10.1016/bs.abr.2019.09.016 (2020).
747 67. Younginger, B. S. & Friesen, M. L. Connecting signals and benefits through partner choice in plant-
748 microbe interactions. *FEMS Microbiol. Lett*. **366**, doi:10.1093/femsle/fnz217 (2019).
749 68. Van den Bergh, B., Swings, T., Fauvart, M. & Michiels, J. Experimental design, population
750 dynamics, and diversity in microbial experimental evolution. *Microbiol. Mol. Biol. Rev*. **82**,
751 doi:10.1128/MMBR.00008-18 (2018).

752 69. Couce, A. & Tenaillon, O. A. The rule of declining adaptability in microbial evolution experiments.
753 *Front. Genet*. **6**, 99, doi:10.3389/fgene.2015.00099 (2015).
754 70. Maddamsetti, R., Lenski, R. E. & Barrick, J. E. Adaptation, clonal interference, and frequency-
755 dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics* **200**,
756 619-631, doi:10.1534/genetics.115.176677 (2015).
757 71. Neher, R. Genetic draft, selective interference, and population genetics of rapid adaptation. *Ann*.
758 *Rev. Ecol. Evol. Syst*. **44**, 195-215, doi:10.1146/annurev-ecolsys-110512-135920 (2013).
759 72. Katsnelson, M. I., Wolf, Y. I. & Koonin, E. V. On the feasibility of saltational evolution. *Proc. Natl.*
760 *Acad. Sci. U. S. A.* **116**, 21068-21075, doi:10.1073/pnas.1909031116 (2019).
761 73. Beringer. J. E. R factor transfer in *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **84**, 188-198, doi:
762 10.1099/00221287-84-1-188 (1974)
763 74. Boucher, C. A., Barberis, P. A., Trigalet, A. P. & Demery, D. A. Transposon mutagenesis of
764 *Pseudomonas solanacearum* isolation of Tn5-induced avirulent mutants. *J. Gen. Microbiol.* **131**,
765 2449-2457 (1985).
766 75. Plener, L., Manfredi, P., Valls, M. & Genin, S. PrhG, a transcriptional regulator responding to
767 growth conditions, is involved in the control of the type III secretion system regulon in *Ralstonia*
768 *solanacearum*. *J. Bacteriol*. **192**, 1011-1019, doi:10.1128/jb.01189-09 (2010).
769 76. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J.*
770 *Bacteriol.* **62**, 293-300, doi:10.1128/jb.62.3.293-300.1951 (1951).
771 77. Deatherage, D. E. & Barrick, J. E. Identification of mutations in laboratory-evolved microbes from
772 next-generation sequencing data using breseq. *Methods Mol. Biol.* **1151**, 165-188,
773 doi:10.1007/978-1-4939-0554-6_12 (2014).
774 78. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol*. **29**, 24-26,
775 doi:10.1038/nbt.1754 (2011).
776 79. Vallenet, D. et al. MicroScope: an integrated platform for the annotation and exploration of
777 microbial gene functions through genomic, pangenomic and metabolic comparative analysis.
778 *Nucleic Acids Res* **48**, D579-D589, doi:10.1093/nar/gkz926 (2020).
779 80. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for
780 Statistical (2019).
781 81. Noble, R. ggmuller: Create Muller plots of evolutionary dynamics. R package version 0.5.3.
782 doi:10.5281/zenodo.591304 (2019).
783 82. Dalia, A. B., McDonough, E. & Camilli, A. Multiplex genome editing by natural transformation.
784 *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8937-8942, doi:10.1073/pnas.1406478111 (2014).
785 83. Chen, W. M. et al. Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in
786 nature. *J. Bacteriol.* **185**, 7266-7272, doi:10.1128/jb.185.24.7266-7272.2003 (2003).

787 84. Fahraeus, G. The infection of Clover root hairs by nodule bacteria studied by a simple glass slide
788 technique. *J. Gen. Microbiol*. **16**, 374-& (1957).
789 85. Jensen, H. L. Nitrogen fixation in leguminous plants. I. General characters of root nodule bacteria
790 isolated from species of *Medicago* and *Trifolium* in Australia. *Proc. Int. Soc. N.S.W*. **66**, 68–108
791 (1942).
792
793 **Figure legend**
794 **Fig. 1 | Evolution of the symbiotic properties of *Ralstonia* clones along evolution cycles.**
795 **a**, Left: overview of the evolution cycles showing the symbiotic steps determining bacterial fitness
796 (nodulation competitiveness and within-host proliferation) and human intervention (nodule harvest,
797 grinding and re-inoculation). Nodule colors reflect bacterial genotypes: blue represents the ancestor,
798 and red represents a mutant with increased fitness. Right: schematic representation of the main
799 phenotypic measurements performed in this work, where the symbiotic phenotypes of one evolved
800 strain can be compared to those of a reference strain *(e.g*. the reference symbiont *C. taiwanensis*).
801 Competitive indexes for *in planta* fitness are calculated as the ratio of evolved *vs*. reference clones in
802 nodule bacterial populations normalized by the ratio of strains in the inoculum. Competitive indexes
803 for nodulation competitiveness are calculated as the ratio of nodules formed by each strain normalized
804 by the inoculum ratio. Within-host proliferation is measured in independent single inoculations of each
805 strain. **b-d**, Relative *in planta* fitness (**b**), nodulation competitiveness (**c**) and within-host proliferation
806 (**d**) of nodulating ancestors (cycle 0) and evolved clones isolated from cycles 16 and 35 were compared
807 to *C. taiwanensis*. Values correspond to means ± standard deviations. Data were obtained from at least
808 three independent experiments. For each experiment, nodules were harvested from 10 plants (**b**), 20
809 plants (**c**) and 6 plants (**d**). The sample size (n) is equal to n=3 (**b**,**c**) or comprised between n=15-18 (**d**).
810 Raw data are available in Supplementary Table 2. cfu: colony-forming units.
811
812 **Figure 2. Dynamics of molecular evolution**
813 Allele frequency trajectories of mutations that attained a frequency of 30% in at least one population
814 of the B, F, G, K and M lineages (**a**,**b**,**c**,**d**,**e**). Mutations with similar trajectories were clustered in cohorts
815 and represented by different colors**.** For simplicity, mutations travelling alone were also called cohorts.
816
817 **Figure 3. Symbiotic phenotypes of reconstructed mutants and evolved clones from lineages B and G**.
818 **a**,**d**, *In planta* relative fitness of evolved clones carrying reconstructed mutations from fixed mutational
819 cohorts. Evolved clones in which mutations were reconstructed are indicated on the left of the graph.
820 Horizontal lines separate the different mutational cohorts. Cohort number and the number of tested

821 mutations from the cohort on the total number of mutations present in the cohort are indicated in
822 brackets. Competitive indexes (CI) were calculated as the ratio of the mutant strain on the isogenic
823 parental strain in bacterial nodule populations normalized by the inoculum ratio. GFP/mCherry
824 correspond to control co-inoculation experiments of strains derived from the same evolved clone
825 labelled with different fluorophore either GFP or mCherry. **b**,**e**, Nodulation competitiveness effect of
826 adaptive mutants. CI were calculated as the ratio of the number of nodules formed by the mutant
827 strain on the number of nodules formed by the isogenic parental strain normalized by the inoculum
828 ratio. **a,b**,**d**,**e**, vertical segments correspond to mean values of CI. Red segments indicate significantly
829 beneficial mutations, gray segments indicate neutral mutations and blue segments indicate
830 significantly deleterious mutations (P<0.05, t-test). **c**,**f**, Within-host proliferation (number of bacteria
831 per nodule) of adaptive mutants and the corresponding isogenic parental clones. Rectangles span the
832 first quartiles to the third quartiles, bold segments inside the rectangle show the median, unfiled circles
833 represent outliers, whiskers above and below each box show the minimum and the maximum in the
834 absence of suspected outlying data. Red fillings indicate the mutations that significantly improved
835 within-host proliferation compared to the parental evolved clone (P<0.05, Wilcoxon test). **a**,**b**,**c**,**d**,**e**,**f**,
836 Data were obtained from 3 to 5 independent experiments. For each experiment, nodules were
837 harvested from 10 plants (**a**,**d**), 20 plants (**b**,**e**) and 6 plants (**c**,**f**). The sample size (n) is comprised
838 between n=3-5 (**a**,**d**), n=3-4 (**b**,**e**) and n=15-24 (**c**,**f**). Raw data are available in Supplementary Table 9.
839
840 **Figure 4. Relative strength of selection for nodulation competitiveness and proliferation**
841 **a,** Median fitness trajectories of 100 simulated populations evolving under different sizes of nodulation
842 bottleneck (10 to 3000 nodules). Points indicate fitness values at cycles 0, 10, 20, 30, 40 and 50. Dotted
843 grey lines represent iso-fitness lines (defined as the product of nodulation competitiveness and
844 proliferation) used to delineate 3 fitness domains: Low (L; fitness < 10^{-5}), Mid (M; 10^{-5} < fitness < 10^{-3}),
845 and High (H; fitness > 10^{-3}). The black dotted line represents the diagonal, along which population
846 would improve both phenotypic traits equally well. **b,** Fold effect (nodulation competitiveness effect
847 divided by proliferation effect) of mutations that reached a frequency of at least 30% in simulated
848 populations. Color shading indicate the fitness domain (Low, Mid or High; shown in panel **a**) in which
849 each group of mutations arose. Note that fold effects of mutations are indicated even when median
850 fitness trajectories do not reach the High fitness domain (*e.g.* 10 and 30 nodule lines in **a**) because
851 some individual replicate simulations (shown in Supplementary Figures) do reach this fitness domain.
852 Figures **c** and **d** are built following the same principles than **a** and **b**, but explore the effect of differences
853 in initial fitnesses. Note that the dashed diagonal (in **b**) or horizontal (in **d**) lines represent the expected
854 evolutionary trajectories when initial fitnesses are unequal, and upon the null hypothesis that
855 adaptation would proceed proportionally to the differences between the ancestral and the theoretical

optimum values for each phenotypic component. Figures **e** and **f** are as **a** and **b** but explore a situation where the nodulation bottleneck occurs after bacterial clonal proliferation.

Tables

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

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.

Extended data

891 **Extended Data Fig. 1**. **Experimental evolution of *Ralstonia solanacearum* GMI1000 pRalta through**
892 **serial cycles of plant (*Mimosa pudica*) inoculation-isolation of nodule bacteria.**
893 The 5 bacterial lineages, B, F, G, K, and M lines were independently derived from the spontaneous
894 nodulating mutants either CBM212, or CBM349, or CBM356. At each evolution cycle, bacterial nodule
895 populations were isolated either 21 days (black lines) or 42 days (blue lines) after plant inoculation and
896 partly directly re-inoculated to new plants to initiate the next cycle. At each cycle, one individual clone
897 representative of the bacterial population was isolated and stored. Sequenced populations and clones
898 are indicated. The number of nodules harvested at each evolution cycle is reported in Supplementary
899 Table 1.
900
901 **Extended Data Fig. 2. Evolution of population composition over time in lineages B and G.**
902 Muller plots representing the relative frequency of each genotype in lineages B (a) and G (b) for each
903 sequenced population (uneven cycles). Each color represent a different genotype, and its relative
904 frequency is shown as the vertical area at each sequenced time point. The genotypes containing
905 cohorts for which adaptive mutations were identified experimentally are indicated. For each cohort,
906 non-adaptive or deleterious mutations are shown in black, adaptive mutations are in red.
907
908 **Extended Data Fig. 3. Survival of reconstructed mutants from lineages B and G in the Jensen medium**
909 **and rhizosphere**.
910 Survival of evolved clones carrying reconstructed adaptive mutations from fixed mutational cohorts in
911 the Jensen medium (**a**) and in rhizosphere (**b**). Competitive indexes (CI) were calculated as the ratio of
912 the mutant strain on the isogenic parental strain in Jensen medium populations normalized by the
913 inoculum ratio. Red points correspond to mean values of CI. Rectangles span the first quartiles to the
914 third quartiles, bold segments inside the rectangle show the median. Red stars beside strain names
915 indicate the mutations that significantly improved survival in the rhizosphere compared to the parental
916 evolved clone ($P<0.05$, Wilcoxon test with Benjamini-Hochberg correction). Data were obtained from
917 at least 3-5 independent experiments for each comparison and the sample size (n) is comprised
918 between n=9-14. Raw data are available in Supplementary Table 10.
919
920 **Extended Data Table 1.** Summary of detected mutations in the five lineages
921 **Extended Data Table 2.** Analysis of genetic parallelism in the five lineages.
922 **Extended Data Table 3.** Global analysis of mutational cohorts
923 **Extended Data Table 4.** Neutral and deleterious reconstructed mutations
924 **Extended Data Table 5.** Frequencies of usage of codons modified by adaptive synonymous mutations
925

926

927 **Supplementary information**

928 **Supplementary Table 1.** Number of nodules harvested at each evolution cycle.

929 **Supplementary Table 2**. Raw data obtained for the phenotypic characterization of *Ralstonia* evolved

930 clones compared to *C. taiwanensis.*

931 **Supplementary Table 3.** Coverage depth for whole-population sequencing

932 **Supplementary Table 4**. List of mutations detected in evolved populations of the B lineage.

933 **Supplementary Table 5**. List of mutations detected in evolved populations of the F lineage.

934 **Supplementary Table 6**. List of mutations detected in evolved populations of the G lineage.

935 **Supplementary Table 7**. List of mutations detected in evolved populations of the K lineage.

936 **Supplementary Table 8**. List of mutations detected in evolved populations of the M lineage.

937 **Supplementary Table 9**. Raw data obtained for the *in planta* phenotypic characterization of

938 reconstructed mutants compared to their respective isogenic parental strain.

939 **Supplementary Table 10.** Raw data obtained for the *ex planta* phenotypic characterization of

940 reconstructed mutants compared to their respective isogenic parental strain.

941 **Supplementary Table 11.** Strains and plasmids used in this study.

942 **Supplementary Table 12**. List of mutations detected in evolved clones of the B lineage.

943 **Supplementary Table 13**. List of mutations detected in evolved clones of the F lineage.

944 **Supplementary Table 14**. List of mutations detected in evolved clones of the G lineage.

945 **Supplementary Table 15**. List of mutations detected in evolved clones of the K lineage.

946 **Supplementary Table 16**. List of mutations detected in evolved clones of the M lineage.

947 **Supplementary Table 17.** Oligonucleotides used in this study.

948 **Supplementary Table 18.** Parameter values used in computer simulations

949

950 **Supplementary Information. Model framework and supplementary simulation results**

951

952 **Correspondence and requests for materials** should be addressed to P.R. or D.C.

953

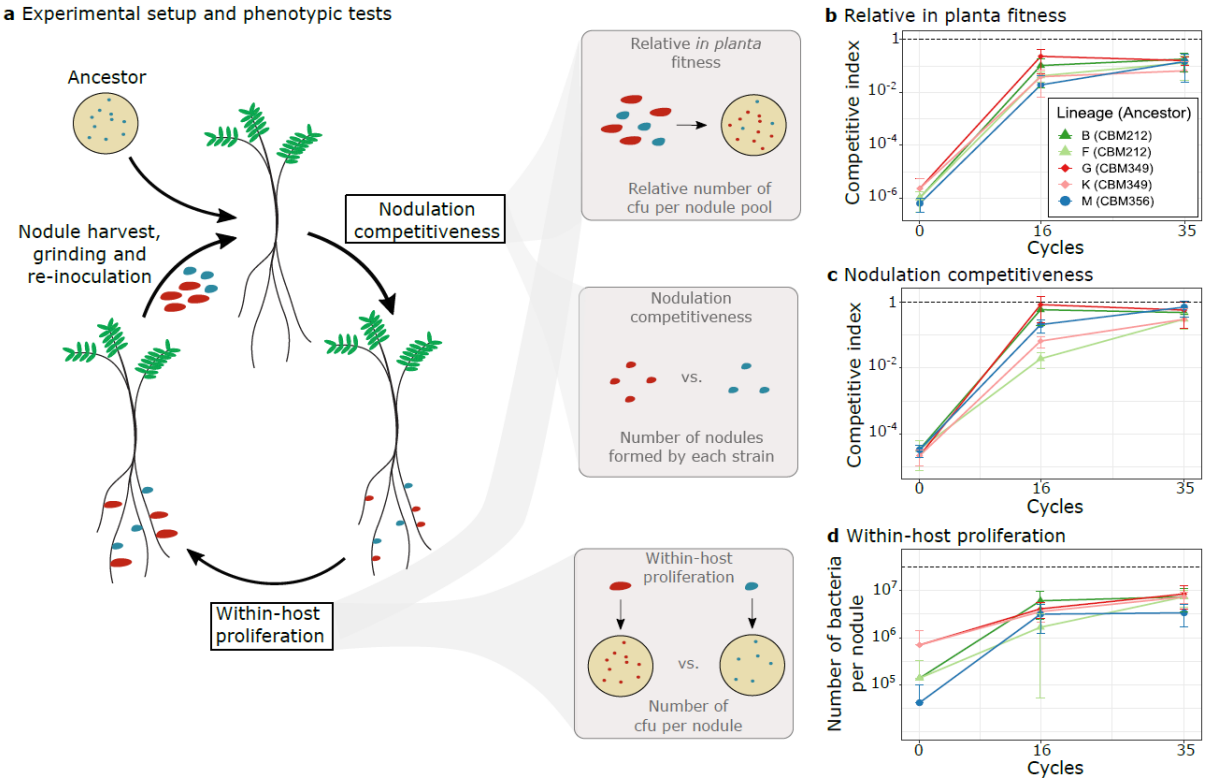


Fig. 1 | Evolution of the symbiotic properties of *Ralstonia* clones along evolution cycles.
a, Left: overview of the evolution cycles showing the symbiotic steps determining bacterial fitness (nodulation competitiveness and within-host proliferation) and human intervention (nodule harvest, grinding and re-inoculation). Nodule colors reflect bacterial genotypes: blue represents the ancestor, and red represents a mutant with increased fitness. Right: schematic representation of the main phenotypic measurements performed in this work, where the symbiotic phenotypes of one evolved strain can be compared to those of a reference strain *(e.g*. the reference symbiont *C. taiwanensis*). Competitive indexes for *in planta* fitness are calculated as the ratio of evolved *vs*. reference clones in nodule bacterial populations normalized by the ratio of strains in the inoculum. Competitive indexes for nodulation competitiveness are calculated as the ratio of nodules formed by each strain normalized by the inoculum ratio. Within-host proliferation is measured in independent single inoculations of each strain. **b-d**, Relative *in planta* fitness (**b**), nodulation competitiveness (**c**) and within-host proliferation (**d**) of nodulating ancestors (cycle 0) and evolved clones isolated from cycles 16 and 35 were compared to *C. taiwanensis*. Values correspond to means ± standard deviations. Data were obtained from at least three independent experiments. For each experiment, nodules were harvested from 10 plants (**b**), 20 plants (**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 Supplementary Table 2. cfu: colony-forming units.

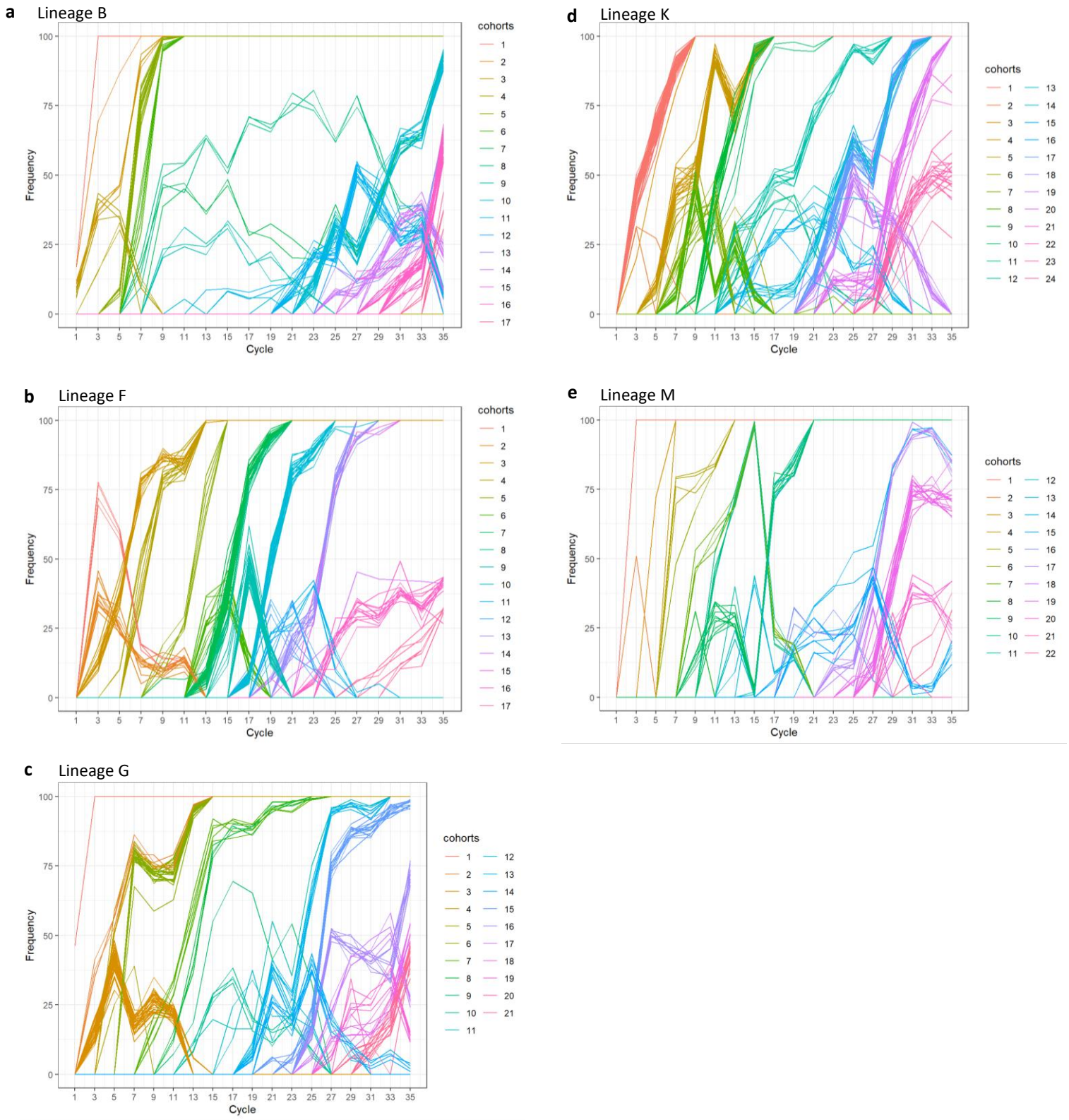


Fig. 2 | Dynamics of molecular evolution. Allele frequency trajectories of mutations that attained a frequency of 30% in at least one population of the B, F, G, K and M lineages (**a**,**b**,**c**,**d**,**e**). Mutations with similar trajectories were clustered in cohorts and represented by different colors. For simplicity, mutations travelling alone were also called cohorts.

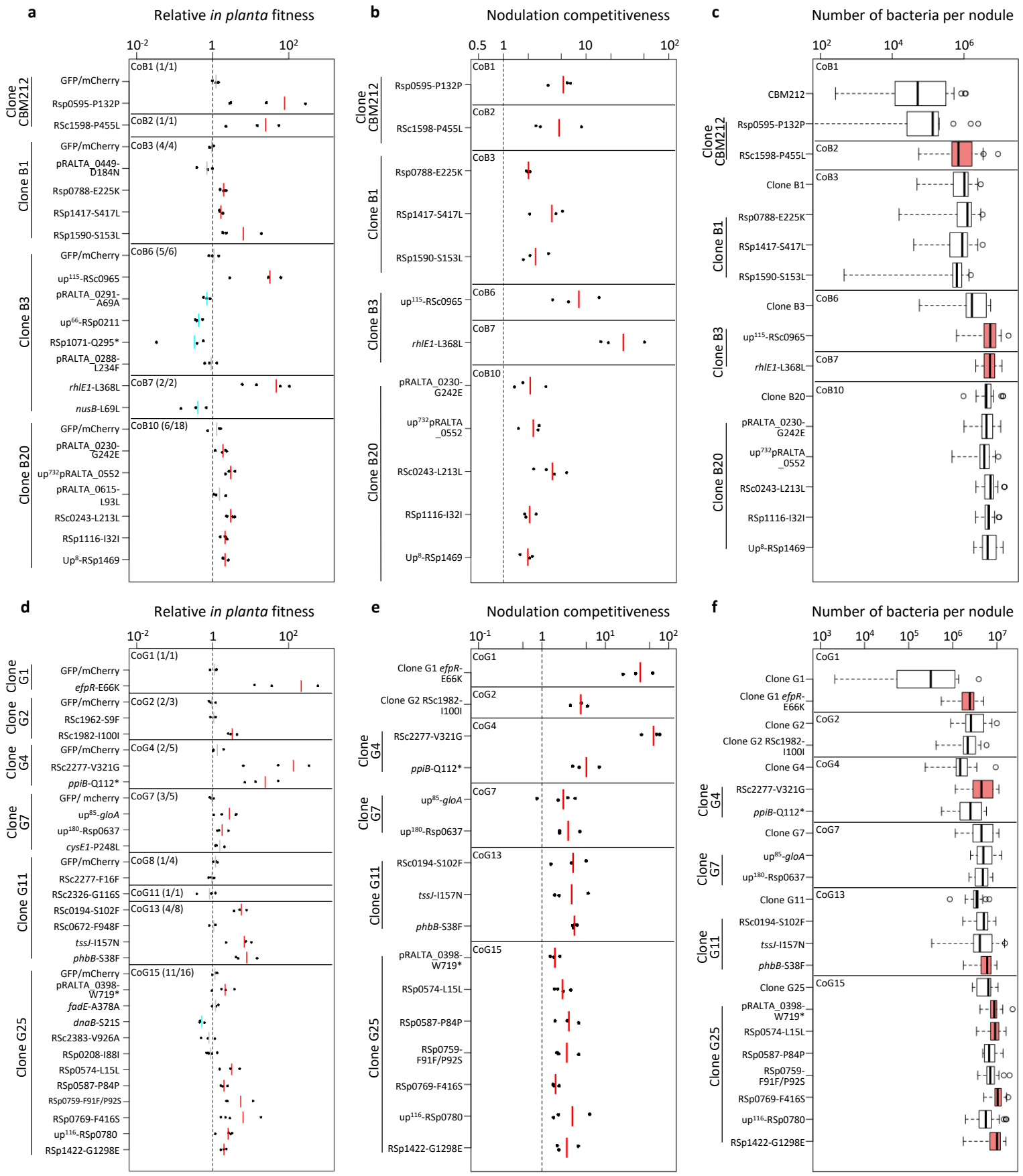

a
Relative in planta fitness
10-2
1
102
Clone CBM212
Clone B1
Clone B3
Clone B20
GFP/mCherry
Rsp0595-P132P
RSc1598-P455L
CoB1 (1/1)
CoB2 (1/1)
CoB3 (4/4)
pRALTA_0449-D184N
Rsp0788-E225K
RSp1417-S417L
RSp1590-S153L
CoB6 (5/6)
up115-RSc0965
pRALTA_0291-A69A
up66-RSp0211
RSp1071-Q295*
pRALTA_0288-L234F
CoB7 (2/2)
rhlE1-L368L
nusB-L69L
CoB10 (6/18)
pRALTA_0230-G242E
up732pRALTA_0552
pRALTA_0615-L93L
RSc0243-L213L
RSp1116-I32I
Up8-RSp1469
b
Nodulation competitiveness
0.5
1
10
102
c
Number of bacteria per nodule
102
104
106
CBM212
Clone B1
Clone B20
d
Relative in planta fitness
Clone G1
Clone G2
Clone G4
Clone G7
Clone G11
Clone G25
efpR-E66K
RSc1962-S9F
RSc1982-I100I
RSc2277-V321G
ppiB-Q112*
GFP/ mcherry
up85-gloA
up180-Rsp0637
cysE1-P248L
RSc2277-F16F
RSc2326-G116S
RSc0194-S102F
RSc0672-F948F
tssJ-I157N
phbB-S38F
fadE-A378A
dnaB-S21S
RSc2383-V926A
RSp0208-I88I
RSp0574-L15L
RSp0587-P84P
RSp0759-F91F/P92S
RSp0769-F416S
up116-RSp0780
RSp1422-G1298E
CoG1 (1/1)
CoG2 (2/3)
CoG4 (2/5)
CoG7 (3/5)
CoG8 (1/4)
CoG11 (1/1)
CoG13 (4/8)
CoG15 (11/16)
e
Nodulation competitiveness
10-1
1
101
102
Clone G1 efpR-E66K
Clone G2 RSc1982-I100I
f
Number of bacteria per nodule
103
104
105
106
107

Fig. 3| Symbiotic phenotypes of reconstructed mutants and evolved clones from lineages B and G.
a,**d**, *In planta* relative fitness of evolved clones carrying reconstructed mutations from fixed mutational cohorts. Evolved clones in which mutations were reconstructed are indicated on the left of the graph. Horizontal 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. **b**,**e**, Nodulation competitiveness effect of adaptive mutants. 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. **a,b**,**d**,**e**, vertical 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). **c**,**f**, Distribution of within-host proliferation values (number of bacteria per nodule) of adaptive mutants and the corresponding isogenic parental clones. Rectangles span the first quartiles to the third quartiles, bold segments inside the rectangle show the median, unfiled circles represent outliers, whiskers above and below each box show the minimum and the maximum in the absence of suspected outlying data. Red stars indicate the mutations that significantly improved infectivity compared to the parental evolved clone ($P<0.05$, Wilcoxon test). **a**,**b**,**c**,**d**,**e**,**f**, Data were obtained from 3 to 5 independent experiments. For each experiment, nodules were harvested from 10 plants (**a**,**d**), 20 plants (**b**,**e**) and 6 plants (**c**,**f**). The sample size (n) is comprised between n=3-5 (**a**,**d**), n=3-4 (**b**,**e**) and n=15-24 (**c**,**f**). Raw data are available in Supplementary Table 9.

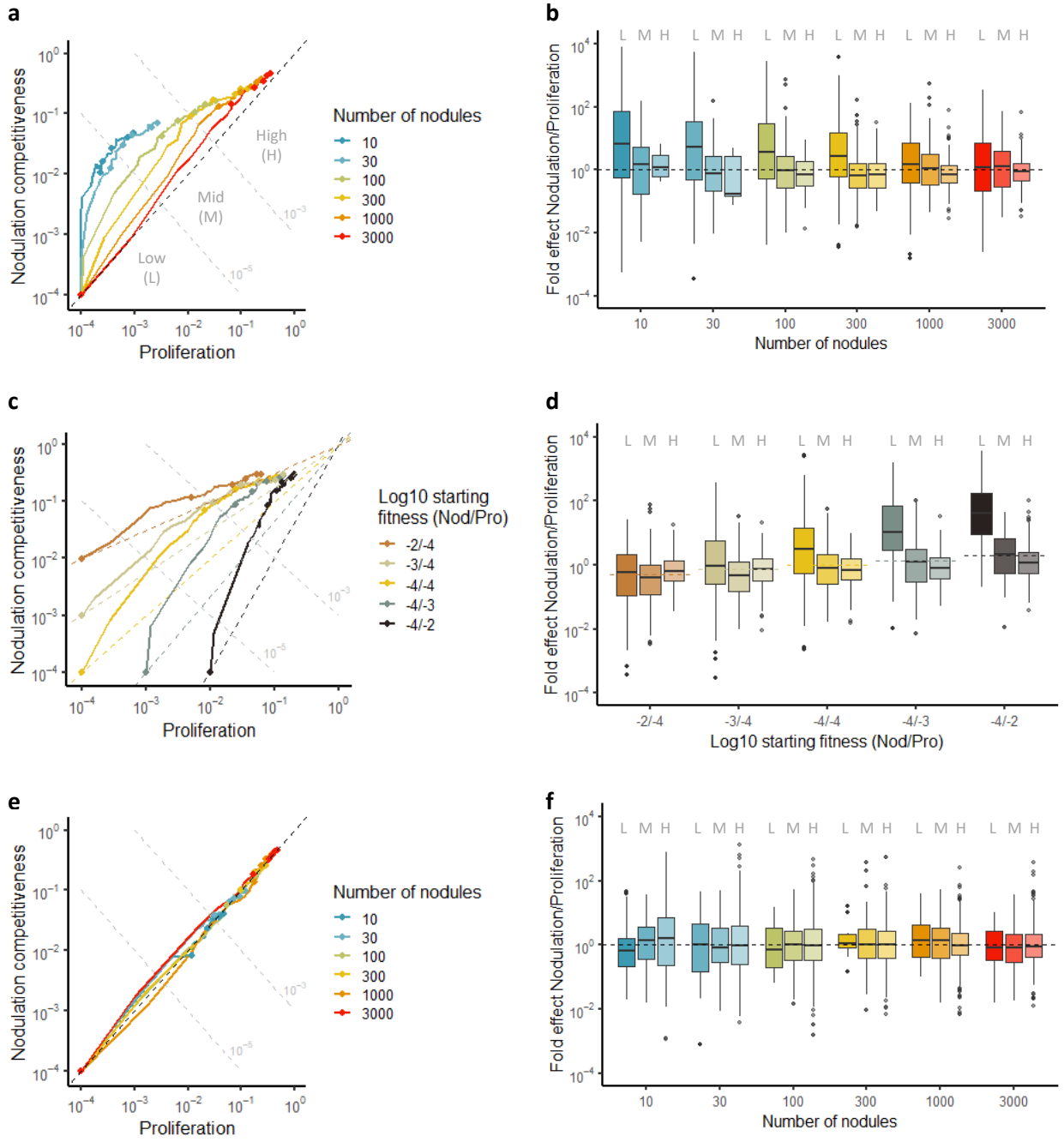


Fig. 4 | Relative strength of selection for nodulation competitiveness and proliferation
a,c,e, Median fitness trajectories of 100 simulated populations evolving under different sizes of nodulation bottleneck (10 to 3000 nodules). Points indicate fitness values at cycles 0, 10, 20, 30, 40 and 50. Dotted grey lines represent iso-fitness lines (defined as the product of nodulation competitiveness and proliferation) used to delineate 3 fitness 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 diagonal, along which population would improve both phenotypic traits equally well. **b,d,f**, Fold effect (nodulation competitiveness effect divided by proliferation effect) of mutations that reached a frequency of at least 30% in simulated populations. Color shading indicate the fitness domain (Low, Mid or High) in which each group of mutations arose. Note that fold effects of mutations are indicated even when median fitness trajectories do not reach the High fitness domain (*e.g.* 10 and 30 nodule lines in **a**) because some individual replicate simulations (shown in Supplementary Figures) do reach this fitness domain. The effect of different parameters on adaptive trajectories was tested: the size of nodulation bottleneck (**a, b**), the initial fitness values of the ancestor (**c, d**), and the size of nodulation bottleneck under a scenario where the nodulation bottleneck occurs after bacterial clonal proliferation (**e, f**).

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	Hypothetical protein	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 L33F	Syn Nonsyn	Putative low specificity L-threonine aldolase	CoB10	**2.1**	1.2	1.2	**2.1**	0.9	Nod+	0.84	1
	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 P92S	Syn Nonsyn	Putative type VI secretion-associated protein	CoG15	**5.4**	0.7	0.9	**2.5**	1.4	Nod+	1.00	1
	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*, 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.